BAYESIAN ADAPTIVE SAMPLING FOR DISCRETE DESIGN ALTERNATIVES IN CONCEPTUAL DESIGN

A Thesis
Presented to
The Academic Faculty

by

José Eugenio Valenzuela del Río

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Aerospace Engineering

Georgia Institute of Technology
December 2013

BAYESIAN ADAPTIVE SAMPLING FOR DISCRETE DESIGN ALTERNATIVES IN CONCEPTUAL DESIGN

Approved by:

Professor Dimitri Mavris, Advisor
School of Aerospace Engineering
Georgia Institute of Technology

Professor Daniel Schrage
School of Aerospace Engineering
Georgia Institute of Technology

Professor Brian German
School of Aerospace Engineering
Georgia Institute of Technology

Professor Sankaran Mahadevan
School of Engineering
Vanderbilt University

Professor J. V. R. Prasad
School of Aerospace Engineering
Georgia Institute of Technology

Date Approved: 21 August 2013

A mis padres.

ACKNOWLEDGEMENTS

I want first to thank my advisor, Dr. Mavris, for his wise advice and support that allow me to do research, write this thesis, and be more prepared for the professional world. I also thank Dr. German, Dr. Prasad, Dr. Schrage, and Dr. Mahadevan for their valuable feedback and recommendations provided as committee members.

I also express my gratitude to Dr. Chung Lee, Dr. Kyle Collins, Dr. Miguel Walter, and Dr. Jonathan Murphy for their advice and assistance on my research. Also to my colleagues Rajiv Shenoy, Sylvester Ashok, and Alek Gavrilovski.

My appreciation to "Caja Madrid" and "La Caixa" fellowships that have fully funded me for the past four years.

I would like to thank all the people that have added a personal dimension to my time in Atlanta.

Finally, a big thanks to my father, my mother, my sister, my brother, my niece, and my nephew for their unconditional love and support that make life much easier.

¡Muchas gracias!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

GLOSSARY

AABO Adaptive Approximation-Based Optimization, p. 13.

ANN Artificial Neural Networks, p. 8.

BB Branch and Bound, p. 70.

CFD Computational Fluid Dynamic, p. 4.

CSD Computational Structural Dynamics, p. 54.

DoE Design of Experiments, p. 4.

ECMF Evolutionary Concept Multi-Fidelity Surrogate, p. 33.

ECP Extended Cutting Plane, p. 72.

EGO Efficient Global Optimization, p. 14.

EMOA Evolutionary Multi-Objective Algorithm, p. 17.

ES Evolution Strategies, p. 76.

ExI Expected Improvement, p. 14.

FFT Fast Fourier Transform, p. 144.

GA Genetic Algorithm, p. 18.

GBD Generalized Bender Decomposition, p. 72.

GRFM Gaussian Random Field Metamodel, p. 8.

IP Integer Programming, p. 72.

k-**NN** k Nearest Neighbors, p. 41.

MIC Mixed-Integer-Categorical Surrogate, p. 33.

MICGA Mixed-Integer-Categorical Genetic Algorithm, p. 34.

MIGA Mixed-Integer Genetic Algorithm, p. 97.

MILP Mixed-Integer Linear Problem, p. 71.

MINLP Mixed-Integer Non-Linear Problem, p. 71.

MIO Mixed-Integer Optimization, p. 15.

MIP Mixed-Integer Programming, p. 71.

MLE Maximum Likelihood Estimation, p. 50.

M&S Modeling and Simulation, p. 45.

NLP Non-Convex Programming, p. 72.

OA Outer Approximation, p. 72.

PoI Probability of Improvement, p. 14.

rms Root Mean Squared, p. 109.

SVM Support Vector Machine, p. 8.

VDM Value Difference Metric, p. 40.

Nomenclature

Greek

α	Slope of the logarithmic fit of the rms of the surrogate error vs $N_{tr.set}$
Δ	Increase of a variable
δ	Difference between fidelities
ϵ	Error between a stochastic process and its mean μ
λ	Kriging regression constant
λ_i	Lagrange multiplier for the constraint i
μ	Mean of a stochastic process
μ^j	Mean of the mutation distribution $\mathbf{p}^j$ for the parental category j
ρ	Ratio between fidelities
σ	Standard deviation of the Kriging stochastic process error
θ	Kriging hyper-parameter
θ_1	Linear twist of the inner part of the main rotor blade
θ_2	Linear twist of the outer part of the main rotor blade
$\mathbf{\Psi}$	Kriging correlation matrix
ψ	Kriging correlation vector between the training set points and design x
Ω	Rotor angular velocity

Roman

AR	Aspect ratio of an aerodynamic surface
b	Span of an aerodynamic surface
$\mathbf{C}$	Multi-fidelity approach correlation matrix
$\mathbf{c}$	Multi-fidelity approach correlation vector between the training set points and design x
c	Rotor chord
C_d	Airfoil drag coefficient curve
C_l	Airfoil lift curve
C_P	Rotor-craft power coefficient
C_T	Rotor-craft thrust coefficient
C_W	Rotor-craft weight coefficient
$\mathbf{d}$	Difference between the expensive observations and the cheap surrogate predictions
D	Rotor diameter
d	Distance function
$D1$	Design space of the previous concept
$D2$	Design space of the new concept
$d_{j,k}$	Nominal distance between the categorical members j and k
$E\left[I\right]$	Expected improvement
$error$	Surrogate standardized validation error

$excl D2$ Exclusive design space of the new concept

F Pareto Front

f Objective function

FT Fast Fourier transform of a function

f_{x1} Discrete fast Fourier transform frequency in the x_1 direction

$\mathbb{H}$ Hyper-volume

$\mathbf{I}$ Identity matrix

I Improvement with respect to the current best sample point

Ih One dimensional Hamming distance function

J Number of categorical properties to define the nominal intrinsic distance

k Dimensionality of the design space

L Likelihood of the observed data

m Number of objective functions

n Size of the training set

$N_{1,tr.set}$ Number of training points for the old concept surrogate

$N_{2,tr.set}$ Number of training points for the new concept surrogate

$N_{tr.set}$ Number of training points of a surrogate

N_b Number of rotor blades

n_d Number of members of a categorical variable

N_{eng} Number of engines

$N_{sampl,x1}$ Number of samples in the x_1 direction

$\mathbf{p}$ Probability mutation distribution

P Set of points in the objective space

PDF Probability distribution function

PF Real Pareto front

$P\left[I\right]$ Probability of improvement

p_k^j Probability that the mutation child gets the k-th category given the parent takes the j-th category

$prop$ Categorical property to define the nominal intrinsic distance

q Number of members in the Pareto set

$\mathbb{R}$ Real numbers

$\mathbf{R}$ Kriging interpolation correlation matrix

R Rotor radius

$r_{\text{cut-off}}$ Radial position of the blade cut-off

Re Reynolds Number

res Surrogate standardized validation residual

r_{tw} Radial position of the twist change

$\hat{s}^2$ Kriging mean squared error

S Feasible objective space

shp Shaft horsepower

V Volume of a design or objective space

Vel Rotor-craft velocity

Vol Volume function

$\mathbf{w}$ Weighting vector

W Weight of a system part

W_0 Rotor-craft gross weight

$\mathbf{X}$ Training set points

$\mathbf{x}$ Design point

$\mathbf{y}$ Objective values of the training set

Y Noisy realization of the deterministic model $y(\mathbf{x})$

y Objective function value

$y^{\star}_{ch}$ Characteristic change of the objective function in its domain

y_{min} Current best sample point

y_{opt} Optimum of the objective function

$\hat{y}$ Kriging prediction

$\mathbb{Z}$ Integer numbers

Z Gaussian process

Subscript and others

$\hat{}$ Estimation of optimal value

c Cheap data of the multi-fidelity approach

$cont$	Continuous variables
$Conv$	Conventional anti-torque device
d	Additive factor of the multi-fidelity approach
e	Expensive data of the multi-fidelity approach
Fen	Fenestron anti-torque device
for	Forward flight
hov	Hover
hs	Horizontal tail
int	Integer variables
nom	Nominal or categorical variables
opt	Optimal design conditions
$quant$	Discrete-quantitative variables
r	Regression predictor
ri	Reinterpolation predictor
tr	Tail rotor
vt	Vertical tail

SUMMARY

The number of technology alternatives has lately grown to satisfy the increasingly demanding goals in modern engineering. These technology alternatives are handled in the design process as either concepts or categorical design inputs. Additionally, designers desire to bring into early design more and more accurate, but also computationally burdensome, simulation tools to obtain better performing initial designs that are more valuable in subsequent design stages. It constrains the computational budget to optimize the design space. These two factors unveil the need of a conceptual design methodology to use more efficiently sophisticated tools for engineering problems with several concept solutions and categorical design choices. Enhanced initial designs and discrete alternative selection are pursued.

Advances in computational speed and the development of Bayesian adaptive sampling techniques have enabled the industry to move from the use of look-up tables and simplified models to complex physics-based tools in conceptual design. These techniques focus computational resources on promising design areas. Nevertheless, the vast majority of the work has been done on problems with continuous spaces, whereas concepts and categories are treated independently. However, observations show that engineering objectives experience similar topographical trends across many engineering alternatives.

In order to address these challenges, two meta-models are developed. The first one borrows the Hamming distance and function space norms from machine learning and functional analysis, respectively. These distances allow defining categorical metrics that are used to build an unique probabilistic surrogate whose domain includes, not

only continuous and integer variables, but also categorical ones. The second meta-model is based on a multi-fidelity approach that enhances a concept prediction with previous concept observations. These methodologies leverage similar trends seen from observations and make a better use of sample points increasing the quality of the output in the discrete alternative selection and initial designs for a given analysis budget. An extension of stochastic mixed-integer optimization techniques to include the categorical dimension is developed by adding appropriate generation, mutation, and crossover operators. The resulted stochastic algorithm is employed to adaptively sample mixed-integer-categorical design spaces.

The proposed surrogates are compared against traditional independent methods for a set of canonical problems and a physics-based rotor-craft model on a screened design space. Next, adaptive sampling algorithms on the developed surrogates are applied to the same problems. These tests provide evidence of the merit of the proposed methodologies. Finally, a multi-objective rotor-craft design application is performed in a large domain space.

This thesis provides several novel academic contributions. The first contribution is the development of new efficient surrogates for systems with categorical design choices. Secondly, an adaptive sampling algorithm is proposed for systems with mixed-integer-categorical design spaces. Finally, previously sampled concepts can be brought to construct efficient surrogates of novel concepts. With engineering judgment, design community could apply these contributions to discrete alternative selection and initial design assessment when similar topographical trends are observed across different categories and/or concepts. Also, it could be crucial to overcome the current cost of carrying a set of concepts and wider design spaces in the categorical dimension forward into preliminary design.

CHAPTER I

INTRODUCTION

1.1 Motivation

In recent times, the aerospace industry has searched for efficient designs and optimization techniques that satisfy the more and more demanding system requirements. This search emphasized on an increase in performance and reduction in cost. Much of the work done focuses on continuous techniques. However, real aerospace design problems normally engage not only continuous variables but also discrete ones, such as materials, number of blades on a compressor disk, beam cross-sections, standardized bolt diameters, commercially available plate thicknesses, and manufacturing processes, among others. Also, different design configurations are developed to meet the increasing system requirements and trade-offs in the form of several concepts. Typical concept examples are helicopter rotor configurations, airplane tail configurations, gear-box configurations, and landing gear types, among others.

In the conceptual design phase, better discrete alternative selection and high performing region bounds are pursued after the sizing of each alternative. There exists a heavy computational load in current engineering problems because of the difficulty of the phenomenon per se, iterations between different disciplines or the designers' desire of multi-objective optimization. Traditionally, cheap empirical design tools have been used in conceptual design; however, they may wrongly select the best concept and its high-performing design space region. Additionally, these tools cannot accurately deal with new concepts and extreme geometries in the early design phase which is the precise time of the process to study new concepts and extreme geometries. Other option are simple models; however, they do not model complex physics, therefore,

they provide poor concept choice and a poor identification of the high-performing regions. Thus, the design community tends to employ more and more faithful models with the drawback of more time-consuming evaluations: designers have to trade off between cost and quality of the model.

Discrete alternatives are handled as either different concepts[1] or categorical discrete variables[2]. Discrete variable and concept selection are normally settled in early design stages because their changes are large by nature so that the impact of these choices in the design is important. Thus, they need to be correctly defined in early phases. These key decisions fix a high percentage of the life cycle cost. Variable changes in later stages of the design process could have a great enhancement in the design performance but also in the whole design concept, making changes prohibitive in many cases. Engineers prefer being aware of these desirable changes earlier in the process where the impact of the change is lower; thus, more and more faithful codes are used in conceptual design. As opposed to discrete variables, continuous changes are sometimes small, allowing engineers to accept the change even in downstream stages of the design process. Regarding discrete alternatives, there is little space for major changes in these crucial decisions, such as concept and category selection, in late stages of design. Therefore, this need of faithful codes in early design is even more crucial when concepts have to be selected and discrete design parameters are present.

Many engineering efforts attempt to bring computationally intense codes into early design. These efforts focus on reducing two limitations stemming from the high computational burden of the codes for concept evaluation and the multi-modality landscapes of typical objectives. The first limitation is the time taken to evaluate design

[1]The author means by concepts the possible solutions for a given need. Concepts normally have different design space and sometimes have some design variables in common.

[2]A detailed classification of discrete variables can be found in Section 2.1.

concepts, whereas the second one is the number of evaluations needed for the optimization in conceptual design. These efforts have been done only in the continuous world; currently, each discrete alternative is modeled and optimized independently.

In order to reduce the concept evaluation time, engineers employ cheap surrogate models for each concept when the tool to model the concept is computationally intense. These surrogate models are based on training points. The number of training points increases dramatically with the dimensionality of the concept design space for a given accuracy; it is called the curse of dimensionality. Thus, the successful utilization of surrogates is restricted to a certain dimensionality of the design space. Design spaces with high dimensionality require a large number of function calls to build an accurate surrogate, which can be impractical when using high-fidelity tools. Thus, more efficient manners to treat surrogate models are needed in these high dimensionality cases. Engineers have developed independent surrogates for each discrete alternative, so the design space they deal with is continuous. Less often surrogates for concepts with discrete-quantitative variables have been built.

The second limitation is set because of the multi-modal nature of typical engineering objective functions. Local optimizers are efficient on smooth uni-modal objective functions, but their results are not satisfactory when applying to functions with multiple local optima. Also, some of the engineering objective functions are non-smooth, which represents another challenge for local solvers. Additionally, normally there is no prior knowledge of the objective function landscape for each concept. Engineers employ global optimization methods to find optimal performance in the design space of each concept. As is well known, the global optimization methods lack convergence speed because they need many more function evaluations than traditional local optimizers. Again, these techniques are mainly applied to concepts with continuous variables; thus, it is desirable their extension to concepts with continuous, discrete-quantitative, and discrete-categorical variables. These design spaces are known as

mixed-integer-categorical ones.

In the case of modestly expensive functions, engineers opt for adaptive sampling (also called on-line or infill sampling) over traditional design of experiments (DoE). Adaptive sampling features two possible focuses or a combination of both: the global accuracy of the model to assure global search (exploration), and the accuracy of the model in the region of the optimum (exploitation). It allows bounding a region of high performance that reduces the design time in later stages of the process.

Regarding global model accuracy, sampling in low performance design regions misuses the resources, especially if design evaluations are costly. Nevertheless, these evaluations in low performance regions are necessary to brand the regions as low performing. A combination of both focuses, exploration and exploitation, is appropriate in early design where a compromise solution between global trends search, interaction between design space variables, and search for an optimum is desired. As a drawback of the adaptive sampling, there exists a loss of precision in the optimum due to the global exploration; however, it does not represent a big issue in early design. The extension of these adaptive sampling techniques to mixed-integer-categorical problems would allow more efficient tools for scenarios with categorical variables.

Currently, high-fidelity codes like Computational Fluid Dynamic (CFD) are mainly used in later design stages and rarely employed in early stages. Neither the great increase of processor speed that looks to approach the limits of silicon, nor the blossom of parallel machine, allows designers to extensively employ high-fidelity codes in early design, even for single multi-objective design.

Also, some authors have pointed out that there are not only computational problems but also some profound problems in the use of high-fidelity codes due to uncertainties inherent to the early design phase. In this phase many variables are simplified or just not considered in the first analysis. These variables will be addressed in later phases. Also, some variables come into the play in later stages once more details are

known about the design.

As a result, analysis from these high-fidelity codes must be judged, specially in these early design stages where uncertainties are higher. So, seeking for an accurate numerical value of the optimum in the conceptual design stage is a waste of resources, because there still exist some parameters left out from model reduction and uncertainty due to unknowns in the design that are settled in the detailed design phase, see Young et al. [208]. Therefore, using mid-fidelity codes in early design makes more sense than high-fidelity ones.

Additional computational burden appears when dealing with not only high-fidelity codes but also multi-objective and multi-disciplinary problems. These engineering scenarios prompt the need of building surrogates and infill techniques as efficient as possible. Pareto sets are desired in multi-objective optimization. Since a set of design solutions is sought, as opposed to a single solution, more optimization effort is usually required. Also, even coverage in the set of solutions is wished, which is neither simple nor computationally cheap.

Sometimes engineers need to use multi-disciplinary tools to model complex engineering problems where several disciplines are involved. Modeling these difficult engineering problems require iterative processes that normally result in computationally intense codes, even when simple physics-based models are utilized for each discipline. In this case, discrete changes in late design stages are even more prohibitive due to the interaction between disciplines. Therefore, it is critical to use mid-fidelity tools to reliably choose a concept and successfully find regions of high performance across the mixed-integer-categorical design space of the concept.

An example of design that involves several disciplines is rotor-craft design, where interactions between aerodynamic, structural dynamics, and control occur. Additionally, the rotor is a complex-physics problem, therefore, models used should represent

the physics with at least some fidelity. Helicopter design include not only continuous variables but also discrete ones (number of blades, type of airfoils, articulated or hingeless blades, ...). Also, a choice of concepts is involved (rotor configuration, anti-torque device, ...)

Nowadays, the techniques previously explained (surrogates and adaptive sampling techniques) enable designers to use computationally intense models in the preliminary and detailed design phases, where design spaces are small and continuous, and concept and alternative selection is already made. Nevertheless, these techniques are not ready for conceptual design stages where there are several alternatives to explore, increasing the volume of the design space; in this case, the currently available surrogates and adaptive sampling techniques provide poor accuracy for a given intense function call budget or need too many function calls for the desired accuracy.

The main goal of the present work is to build surrogates and adaptive sampling algorithms for a more efficient use of computationally expensive tools in scenarios with discrete design alternative choices in early design. The work focuses on the part of conceptual design after each alternative is sized, where it is pursued the exploration of the alternative design space, the selection of the best concept and category, and the search for design regions with high performance. An accurate numeric value of the global optimum is not a priority in conceptual design. Knowing the best concepts and bounding their regions of high performance allow the designer to narrow the search in later design phases; it reduces the likelihood of major changes in preliminary and detailed design, which is especially undesirable for discrete variables and concepts.

1.2 Meta-Modeling in Conceptual Design. State of the Art

Conceptual design is an iterative process where design requirements, trade-offs, and analysis are used to guide and evaluate configuration arrangements. Conceptual design outcomes are discrete alternative selection and initial designs. Typical conceptual

design spaces are large, so simple models have been used historically in this stage to select concepts and initial designs. However, more and more attention is turning towards physics-based tools to choose accurately the concept and initial designs. Two kinds of discrete alternative selection techniques are utilized: qualitative and quantitative.

Qualitative techniques typically rely on prior knowledge and designer experience. They use simple mathematics. Examples of qualitative methods are group preference and voting, such as feasibility screening [186], Pugh decision matrices [147], pairwise comparison charts [48], and analytical hierarchy process [211]. These methods ignore the potential for each concept to be a parameterization of its own design space, thus, Pareto fronts are not captured. Also, they cannot properly manage revolutionary concepts because there is no experience or historical trend about them as stated by Choi et al. [28]. These setbacks motivate more rigorous methods of selection such as quantitative ones.

Quantitative methods appear because of the desire of the designers for a more effective utilization of modeling and simulation in the conceptual design stage, see [72, 194]. They have a more solid mathematical basis and are based on quantitative analysis after the sizing of each alternative. Comparison between concepts is replaced by scalar quantities in the form of objectives that are evaluated with computer models. They are especially beneficial because more reliable tools could establish that a concept must be ruled out over another one. Due to their computational expense, trade-offs between development time and accuracy of the model must be done as Sobieszczanski-Sobieski and Haftka [177] point out. Among these quantitative methods one can find: morphological matrix [216], sequential multi-objective optimization, and simultaneous multi-objective optimization [7]. The design community's focus is turning to these quantitative methods [163]. Therefore, efficient concept evaluation is a key factor in early design when quantitative discrete alternative selection techniques

are utilized.

For many real world problems, a single simulation can take some time to complete: from minutes to hours, or even days. Routine engineering tasks, such as concept/category selection, design optimization, design space exploration, and sensitivity analysis, could become too costly. Surrogate modeling appears to alleviate this computational expense. They consist in constructing cheap approximation models to the objective of a concept that intend to approximate the behavior of the real function. The most common surrogate models are polynomial response surfaces, Gaussian random field meta-model (GRFM), support vector machines (SVM), and artificial neural networks (ANN). They have been widely used, as it can be seen in the literature. For more detail see [171, 94, 170, 104].

Surrogates models are widely used together with DoE techniques. These techniques sample the design space in order to get observations that are necessary for the surrogate construction. DoE is a sampling method that chooses the sampling plan a priori[3]. Once the DoE provides the sampling plan, the expensive function is evaluated at the sampling plan points and the surrogate model is fitted. DoE puts the same amount of effort in all areas of the design space exhausting sampling capabilities in low performance regions. Another limitation is that surrogate models based on DoE sampling plans suffer curse of dimensionality: the required number of sample points for a given accuracy of the meta-model increases exponentially with the number of design variables.

Regarding discrete-quantitative variables, surrogates have been utilized in mixed-integer problems, especially in composite structure design. Rikards et al. [153] employ response surfaces in the latter area. Jansson et al. [92] utilize a GRFM , also known as

[3]DoE is also called an offline sampling technique.

Gaussian processes or Kriging[4], combined with DoE sampling in the same field. However, as far as the author knows Kriging models of a concept integrating categorical and discrete-quantitative design spaces have not been built yet.

Currently, objective functions dependent on continuous, integer and categorical variables are treated independently for each value of the categorical parameter, i.e., a surrogate is fitted for each categorical member where the surrogate domain is made of the remaining design variables (continuous and discrete-quantitative). An example of this approach in the aerospace industry is provided by Keane [104]. Boukouvala et al. [22] show a similar example in the pharmaceutical industry using polynomial response surfaces and Gaussian random field meta-models as surrogates, where a surrogate is fitted for each category. A technique to include all the possible kinds of variables, including categorical ones, in the same surrogate would enable wider and more efficient (more accuracy or fewer intense function evaluations) design searches and concept/category selection.

Engineers first optimize each non-numeric category one at a time. Then, once all the optimization problems for all the non-numeric categories are solved, the best optimum across all the categories is selected by comparison. An instance of this optimization approach is found by Keane [104]. This current method is clearly poor for cases where design spaces along the categorical variables are too big to perform a complete search (too many surrogates should be fitted). In these situations, only some subset of the categories that is believed to be the best by experience is examined. So, the accuracy of the obtained optima is constrained by the accuracy of the engineers' guess.

Likewise, different concepts are treated independently from a surrogate point of view. Therefore, after a concept evolves to a new one to try to meet some requirements or trade-offs, the previous concept observations are used no longer. In conceptual

[4]Currently, Kriging methods are wide. They don't necessarily assume Gaussian fields.

design, concepts continuously evolve to new ones. Designers compare optimal values and trends for the different concepts. Then, the concept with the most suitable characteristics for the given requirements is chosen.

It is not rare that observed trends of objective functions for several categories within the same concept are similar, so some resemblance between behaviors could be taken advantage of. An example of this situation is the rotor figure of merit vs the thrust coefficient for different blade airfoils.

Similarly, the trends of an objective function for different concepts along common continuous and discrete-quantitative design parameters are usually similar. In conceptual design, concepts continuously evolve to new ones, thus, it is common that the topographical behavior of immediate concepts are very much alike. An example of this behavior is the net thrust of several engines versus airspeed, shown in Figure 1.

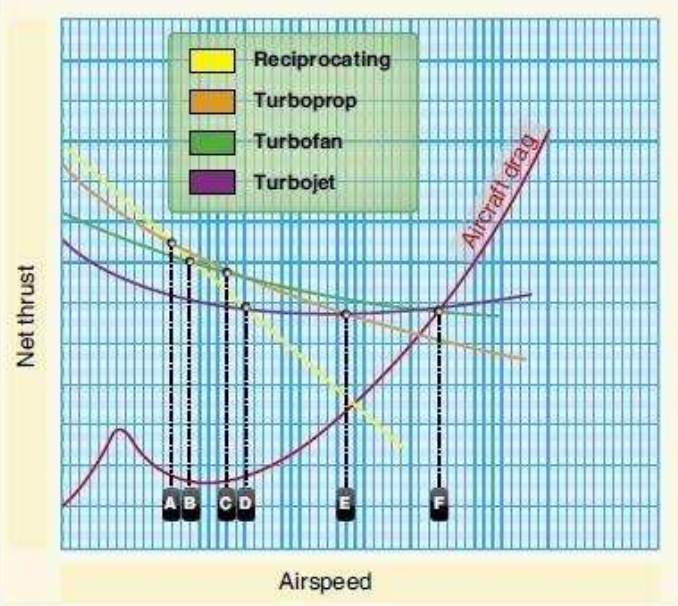


Figure 1: Net Thrust vs Airspeed. Several Engines

Hence, advantage of this resemblance could be taken by designers. When sizing engineering systems and estimating cost and weight, this similarity between several concepts have been exploited in analogy-based techniques (see [169, 131, 112]) and regression-based techniques (see [146, 185, 150, 97, 102]). However, none advantage

of these similar trends has been taken of when predicting engineering objectives with medium and high-fidelity models.

Wilson and Martinez [201] handle nominal input[5] attributes in instance-based learning techniques by making use of distance functions. Wilson and Martinez review some of the more important distance functions: Euclidean, Manhattan, Minkowsky, Camberra, Chebychev, Quadratic, and Mahalanobis, among others. In order to include the nominal inputs, the Hamming distance first introduced by Hamming [78], is used. However, Wilson and Martinez's work is in the field of statistical classification techniques within machine learning far from engineering design. Li, Eggermont et al. [126] make use of the Hamming distance to build their heterogeneous distance with the purpose of developing evolutionary strategies for domain spaces with all kinds of design variables.

Designers employ multi-fidelity techniques to build surrogates when models of several fidelities are available for the concept to explore. A possible instance of this situation is when a Reynolds Averaged Navier Stokes (RANS) model of a airfoil and a Euler equations-based model of it are at the designers' disposal. Alexander et al. [5] use this example to test their first-order multi-fidelity algorithm. A large number of observations of the lower fidelity model and fewer observations of the higher one can be combined to augment the accuracy of the surrogate. A similar application for structure design is tested by Alexander, Lewis et al. [4]. In this case the low-fidelity model is a finite element model evaluated over a coarse mesh.

In some instances, the variable fidelity models are defined in different design spaces. Robinson, Eldred et al. [155] present space mapping and corrected space mapping methods for variable fidelity models of the same concept with different design spaces. However, this method is developed only for trust-region model-management,

[5]Categorical variables are also called nominal, non-numeric and symbolic variables.

and it requires a large training set for the high-fidelity model to get accurate mappings. Also, an overhead calculations in each iteration is needed.

Regarding Bayesian multi-fidelity approaches, Journel and Huijbregts [101] present an algorithm that combines two variable fidelity Gaussian models. El-Beltagy [51] applies one-fidelity Gaussian approaches to the augmented training data-set that includes the low-fidelity approximations to a satellite beam problem (two fidelity models are available). Hevesi et al. [86] apply it to predict the average annual precipitation values using a sparse set of precipitation measurements from the site of interest; it also uses information from additional elevations to improve the estimation accuracy. Kennedy and O'Hagan [109] use an approach in which the variable fidelity codes are modeled as Gaussian processes. Forrester, Bressloff et al. [63] propose partially converged high-fidelity simulations as the low-fidelity model in a multi-level scenario. Later, Forrester, Sóbester, and Keane [64] present the geo-statistical method of Kriging to the multi-fidelity world in detail. Han, Zimmermann et al. [79] propose a new co-Kriging method to study aerodynamic coefficients and drag polars. All the work found in the literature about Bayesian multi-fidelity approaches deal with the same concept or engineering configuration but with different levels of accuracy which stem from model reduction, physics simplification, or variable discretization size.

1.3 Adaptive Sampling. State of the Art

The even distribution of DoE sampling plans wastes resources by placing samples in poor performing regions of the design space. Also, large domains and the high cost of function evaluations do not always make surrogates/DoE methods accurate for optimization purposes. Adaptive sampling techniques represent an alternative to the stiff nature of surrogate/DoE combination.

A very common adaptive technique builds surrogates on successive reductions of the design space. These surrogates are then optimized. This technique is called

adaptive approximation-based optimization (AABO). The surrogate optimization of a concept on a wide design area permits to reduce the optimization area for the next step, where a new surrogate is fitted on a smaller portion of design space. In this technique the adaptive sampling is driven by performance and not by model uncertainty. As a result, its performance is poor in problems where there are several local optima with similar performance. It is especially targeted to reduce computational expense in problems with large design domains. It would exploit a high performance region, but other potential high performance regions would not be explored because there exists little information about them. Recently Hao and Ying [80] implement this adaptive approximation-based optimization method; however, categorical design parameters are not included in their study. Also, Davis and Ierapetritou [37] propose a hybrid methodology for mixed-integer nonlinear programs, where response surface methods are locally applied at the most promising Kriging solutions.

More sophisticated adaptive procedures are called on-line, infill, or adaptive sampling. The idea is to enhance the accuracy of the surrogate model using further function calls, also known as infill or update points. Two focuses are possible: the global accuracy of the model to assure global search (exploration), and the accuracy of the model in the region of the optimum (exploitation). In early design, it is intended to combine both exploration and exploitation. The idea is to sample low performance regions enough to brand them as such and to densely sample high performing regions of the concept design space.

Traditionally, discrete alternatives are adaptively sampled in an independent and sequential way. Lately, two competing concepts are adaptively sampled simultaneously by Rousis [158]; however, he still uses independent surrogates, where there is no cross-use or reuse of computationally expensive observations. It results in the need of heavy sampling for each concept to start with accurate surrogates.

In order to carry adaptive sampling out, predictive distributions of the response

are needed to guide the sampling. Therefore, Bayesian surrogate models like Kriging must be used to approximate the objective function of a concept. Sampling points are chosen according to an optimal probabilistic infill criterion that combines both exploration and exploitation with certain weights.

Many infill criteria are found in the literature: Cox and John [33] use the statistical lower bound, Kushner [120] first proposes probability of improvement (PoI), Mockus [136] first introduces the expected improvement (ExI), Jones and Welch [100] work with conditional lower bound, later Jones proposes the goal seeking infill criterion [98], Sóbester et al. [176] present a parallel infill criterion where several infill points are chosen, later Sóbester et al. [175] work on weighted expected improvement criterion... These techniques are normally employed to overcome the budget limitation when simulations require long run times. The next sample point is obtained by taking into account not only the value of the predicted function but also the uncertainty due to the reduction of the function evaluation cost.

Among Bayesian adaptive sampling techniques, expected improvement is widely used. Jones et al. [99] build Efficient Global Optimization (EGO) algorithms that combine Kriging with ExI for continuous variables. In his work, Jones points out some monotonicity properties of the ExI landscapes. Up to date none adaptive sampling algorithm is done on mixed-integer-categorical problems.

Bayesian adaptive sampling techniques require the optimization of the infill criterion over the design space of the concept for finding the next sampling point. As a result, optimization algorithms for continuous and discrete variables are required to adaptively sample mixed-integer-categorical design spaces. Among the optimization methods developed for continuous and discrete variables, the first and most straight forward approach is to treat all discrete variables as continuous ones[6]; and then,

[6]If categorical discrete variable are present in the problem, a map is required to convert categorical variables into discrete-quantitative ones

variables are truncated once a continuous optimization technique provides the design point.

Li, Eggermont et al. [125] state the main drawback for this approach: the continuous step size may reduce to a value that is too small to generate any improvement in the discrete-quantitative variable treated as continuous. If the discrete design parameter is a ordinal or nominal value, the result is even worse, the implicit continuous neighborhood and distance assumption could cause to converge to an artificial local optimum.

A well-developed field called Mixed-Integer Optimization (MIO) intends to solve this need for optimization problems, where both continuous and discrete-quantitative variables are present. MIO is capable to deal with non-numeric variables by mapping them into integer ones. Other option is to use a set of switches to convert the categories into binary variables that represent the rejection or selection of a category. However, this conversion of categorical variables into binary variables increases the dimensionality of the design space.

Duran and Grossmann [46] build an efficient outer approximation algorithm to solve the problem of synthesis in chemical engineering, which involves binary and continuous variables. Boukouvala et al. [21] also employ mixed-integer optimization techniques for studying blending processes. In the last case, binary variables are associated with units in the superstructure. These variables imply the existence or absence of the system unit and differentiate between the alternatives [21]. Berman and Ashrafi [14] also use binary variables to select the modules in order to optimize software system structures. In heat exchanger synthesis, pump network synthesis, and trim loss minimization same procedure is taken by Floudas et al. [61], Westerlund et al. [198], and Harjunkoski [81], respectively. Davis and Ierapetritou [37] propose a methodology for mixed-integer nonlinear programs based on a branch-and-bound method that uses Kriging and response surfaces of the objective at each node of the

binary tree.

Designers have worked on global optimization techniques for mixed-integer-categorical problems. He and Prempain [82] use particle swarm methods to optimize design spaces with continuous, integer, and discrete-quantitative variables but not nominal variables. Wang and Yin [195] develop a particle swarm optimizer based on ranking selection, where the ranking selection includes categorical variables.

Also, random searches that require mutation operators have been utilized to optimize design spaces with categorical variables by some authors. Bäck [8] proposes to use a uniform distribution for the categorical variable. Other authors also opt for the same mutation distribution for their random searches [57, 127]. Cao and Wu [26] develop a mixed-variable evolutionary algorithm with different mutation operators for each type of variables. Li, Eggermont et al. [126] build an heterogeneous distance to enable the mutation in evolutionary strategies for mixed-integer problems.

1.4 Multi-Objective Optimization. State of the Art

Typical engineering system requirements are made of several objectives. If objective preferences were specified a priori only a single design point would be selected. However, designers poorly understand the influences of objective preferences in early design, so they would like to know how they affect the final decision before choosing a single design [174]; so, designers rely on Pareto fronts. Once the Pareto set is obtained, weighting between desired goals is decided to get to a trade-off solution.

Constructing Pareto fronts is difficult. Firstly, a set of solutions instead of a single solution is wanted. It means that a large optimization effort is required. Secondly, a wide and continual set is pursued in the objective space. Interpolating between design parameters and goals is not usually efficient because of the highly nonlinear relations among design parameters and goals. Therefore, obtaining an even coverage of the Pareto front is difficult, especially if relatively long time computer runs are

employed for each objective. These two reasons shape the current algorithms to construct Pareto sets.

There exist two common ways to obtain the Pareto front for a concept. In the first one, weighting functions to combine the objectives into a single objective problem are chosen; in this case standard single-objective techniques can be brought. Then, the weighting is varied to get different single objective solutions that form the Pareto set approximation. This approach is slow, one Pareto optimal point per weighting. Another drawback is that it is not straight-forward to choose the weighting to get to different parts of the Pareto front.

The second methods are population-based schemes or evolutionary multi-objective algorithms (EMOA). A set of designs is evolved according to dominance and spread criterion towards the final Pareto set. Since population based-schemes deal with a group of candidate solutions, it seems natural to use them in multi-objective optimization problems to find the Pareto optimal solutions simultaneously. For more detail, see [40, 31]. In this case no weighting function is needed. Instances of population-based methods are NSGA-II developed by Deb [41], and SMS-EMOA developed by Beume et al. [16], among others.

Adaptive sampling in multi-objective problems has been treated by several authors. Mainly, the work done is on Bayesian adaptive sampling, concretely on the statistical improvement criterion. Keane [106] extends the single-objective PoI and ExI to the multi-objective world, the latter one based on the euclidean distance to the nearest point of the Pareto front. Liu et al. [128] define a multi-objective ExI with a weighted-sum over single-objective ExIs. However, Wagner et al. [192] define necessary conditions that Keane's work [106] and Liu's work [128] do not satisfy.

The hyper-volume or Lebesgue measure[7] has been an increasingly popular measure to compare the quality of two Pareto set approximations. This Lebesgue measure

[7]Hyper-volume or Lebesgue measure is also known as S-metric

maps a set of points to a scalar, and it is first introduced as a Pareto set quality indicator by Zitzler [213]. It measures the size or volume of the objective space that is dominated by a set of points. A nice feature of the Lebesgue measure is that it captures the spread along the objective space and the accuracy of the set compared to the optimal one. Zitzler, Thiele et al. [215] show that the hyper-volume has more convenient properties than other possible metrics. Its main problem is that the direct calculation of the Lebesgue measure is computationally expensive: its complexity is exponential in the number of objectives. Other limitation is the requirement of objective normalization to assure that the gains in all the objectives have the same importance.

Several authors have developed algorithms to exactly calculate the Lebesgue measure [58, 200, 204, 199]. However, for large number of objectives, Monte Carlo integration is still preferred due to the exponential increase of the computational expense of the direct algorithms. Additionally, a large number of points in the Pareto approximating set increases the computational cost of the direct calculation. The main drawback of the Monte Carlo approximation is its lack of precision. Nevertheless, While, Hingston et al. [200] carry out a survey in the literature that reveals that the majority of authors who deal with multiple objectives study four or less objectives, and none studied problems with more than 12 objectives. Thus, direct hyper-volume calculation could be preferable for multi-objective problems with a few objectives.

Hyper-volume indicators are extensively employed together with EMOA, such as genetic algorithms (GA) and evolutionary strategies, to identify Pareto sets for computationally cheap objective functions or deterministic surrogates. Zitzler [213] first proposes the hyper-volume indicator to compare the approximate sets that are output by different EMOA; Knowles and Corne [115] compare the hyper-volume measure to other metrics employed as outcomes of EMOA; Huband, Hingston et al. [89] suggest a hyper-volume-based measure for resolving ties during the selection process of an

evolutionary strategy; Beume et al. [16] first include the maximization of the dominated hyper-volume as the primal selection operator of the genetic algorithm; Igel, Hansen and Roth [90] employ the hyper-volume contribution as a sorting criterion for their evolutionary strategy; Zitzler, Brockhoff et al. [214] employ the weighted version of the hyper-volume to guide the search to desired points such as Pareto extreme points; and Bader [9] trades off the accuracy of the hyper-volume indicator with the computational resources in his evolutionary algorithm.

However, the interest herein is in Bayesian adaptive sampling of computationally burdensome functions instead of the optimization of the cheap objective functions or deterministic surrogates. Specifically, extensions of single-objective adaptive sampling criterion to multidimensional spaces. Lately, the increment of dominated hyper-volume of a set is a possible choice for the scalar to carry out this extension, as suggested by Emmerich and Giannakoglou's work [55]. Ponweiser [145] does it in the form of Statistical Lower Bound. Emmerich and Giannakoglou [55] propose independently the same multi-objective extension of the ExI criterion as Ponweiser. The proposed ExI satisfies the necessary conditions set by Wagner [192]. Nevertheless, the quadrature for the computation of the ExI measure is expensive: Monte Carlo integration is employed by Emmerich and Giannakoglou [55] with the corresponding low accuracy. Piecewise numerical integration over hyper-rectangles is a costly alternative. Lately, Emmerich [54] develops a direct computation procedure for the ExI quadrature for two-objective problems that provides a more precise and quicker way to calculate the multi-objective ExI than the numerical integration. Herein, the use of the multi-objective ExI on problems with discrete design inputs is intended.

1.5 Rotor-craft Design

Conceptual design of rotor-craft includes the selection of concepts and categorical variables, such as anti-torque device configurations, blade airfoil sections, materials,

engines, and cross-sections of structural elements, among others [123, 68, 151]. Alternative selection in helicopter design is normally determined in early design stages and are usually not changed in later stages. It is because their changes are large by nature so that the impact of these choices in the design process is significant, making these changes prohibitive in latter stages. Once a blade airfoil section is selected, it is aerodynamically optimized in later design stages; however, the possible airfoils are limited to some similar in nature to the first selected. Engine choice changes in later stages of the rotor-craft design process could greatly enhance the design performance, but it also affects the entire design, resulting in excessive changing cost in the majority of the cases.

Rotor-craft designers are aware of the importance of alternative choices. However, the complexity and difficulty of the rotor physics, the multidisciplinary nature of the rotor-craft, and the desire to include multi-objective optimization prevent design team members from employing accurate tools for alternative selection. As in other engineering fields, rotor-craft conceptual design teams are often left with dubiously reliable traditional tools, such as empirical and/or simple models, that cannot model complex physics and may dismiss choices that high-fidelity tools can brand as optimal. Rotor-craft engineers prefer being aware of major design changes in earlier stages where the impact of the change is lower. Therefore, it is intended a more efficient use of computationally intense codes early in the design process to enhance the selection of alternatives; it would make the rotor-craft design process better.

The two approaches to alternative selection, qualitative and quantitative ones, have been used in rotor-craft design. In the qualitative approach, decisions are made by design engineers. These choices are affected by experience and personal preference, they may "feel right" but there is no quantitative supporting evidence. Walsh et al. [193] choose the airfoils by avoiding exceeding the maximum section lift coefficients on the retreating side or exceeding the section drag divergence Mach number on the

advancing side. In the quantitative approach, the exploration of different alternatives is done one by one and with sizing tools. Coupling numerical optimization tools to sizing codes is difficult [165] because of the low-fidelity of sizing codes, and the complexity of the simultaneous handling of discrete, integer and continuous variables, typical in rotor-craft design scenarios. Once each alternative has been explored, the selection is made by comparison. The use of low-fidelity codes makes results of quantitative approaches unreliable.

Within the quantitative methods, Crossley et al. [35] attempt to better address categories in helicopter early design. They realize that binary-coded genetic algorithms can properly handle discrete variables; GA searches simultaneously all the candidate airfoils as opposed to traditional independent searches in each category. Crossley et. al [34] later extend the same approach to select not only the airfoil section but also the rotor-craft engine in the frame of conceptual design; they minimize the gross weight. Wells et al. [196] use a similar approach in a more advanced design phase to search for acoustically efficient rotor blades. Berry et al. [15] choose among several candidate airfoils using a simple predictive tool coupled with a genetic algorithm. However, these methods employ simple analysis tools that fail to capture the complex physics by which categories could be accurately selected for initial designs.

Surrogates have been used in rotor-craft design to reduce the evaluation time. Saijal et al. [162] use neural networks and polynomial meta-models to reduce helicopter vibration. Sun et al. [180] fit response surfaces to design airfoils aerodynamically.

Another popular family of meta-models is approximation-based stochastic processes [160]. These surrogates have been successfully employed in several rotor-craft-related fields. One application of these meta-models predicts unsteady aerodynamic responses when flow non-linearities are present [74, 75]. Many researchers apply this stochastic techniques to rotor design [119, 19, 73, 188, 118, 91]. Vu et al. [191] apply stochastic processes in rotor-craft design to minimize the rotor-craft power in hover.

It is easy to find in the rotatory system literature examples of objective functions with similar behavior for several blade airfoil sections, blade materials and configurations [123, 151, 83, 29, 52]. The giromill performance experiences similar trends for several airfoils, as shown in Figure 2 [52]. In rotor-craft design, again independent meta-models are fitted for each alternative, see [209, 180, 140]. Therefore, current approaches in rotor-craft design do not take advantage of similar trends across alternatives.

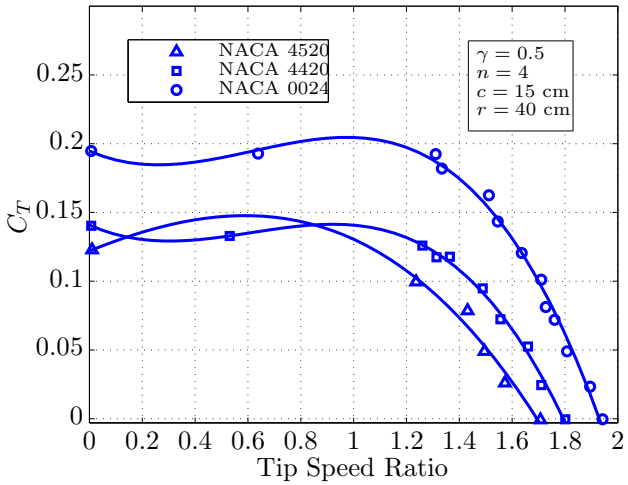


Figure 2: Giromill C_T vs Tip Speed Ratio for Several Airfoils [52]

Recently, optimization techniques have been extended to cope with surrogate modeling in rotor-craft design. The most straight-forward approach is to directly optimize the surrogate objective function without updating the surrogate during the process [180, 141, 95, 162, 156]. These surrogate-based optimization techniques have been applied to blade material and blade ply configuration selection. Murugan and Ganguli [141] optimize each material on the composite blade via response surfaces, and then, they compare the best design of each material to select the optimum one. Murugan et al. [140] optimize several composite ply configurations using surrogates and choose the best performing one by comparison. Yuan and Friedmann [209] optimize independently stiff-in and soft-in blades and compare the results. Sun et al. [180]

proceed similarly to airfoil comparison when optimizing their aerodynamic curves.

The described investigations on categorical comparison with surrogates based optimization have been done in the context of aeroelastic stability enhancement, vibration suppression, and airfoil aerodynamic optimization that fall typically within preliminary design stages. These techniques that include high-fidelity models via surrogates can bring tremendous value if applied to early design stages where the choice between categories is made. Nevertheless, the large conceptual design spaces due in part to the presence of multiple alternatives hinders the application of this technique to conceptual design.

In order to enhance the efficiency of optimization on meta-models, EGO algorithms on stochastic process surrogates attempt to focus most of the expensive function observations in good performing areas. Rotor-craft engineers have applied EGO algorithms to rotor vibration and noise, and performance optimization [119, 73, 118, 91]. EGO algorithms can explore large design spaces, typical in early design stages; however, they have not been applied to alternative selection in rotor-craft design. EGO algorithms are adequate candidates to implement conceptual design tools for the more efficient use of computationally intense models with the purpose of improving the selection of categorical variables and initial designs in rotor-craft design.

1.6 Proposed Approach

The goal of this research is to build conceptual design tools for the more efficient use of computationally intense codes in scenarios with discrete design alternative choices in early design. It would lead to enhancements in the quality of the initial designs and concept/category selection. The idea is to increase the tools (surrogate and adaptive sampling) efficiency by getting more value of the learning on each discrete design alternative. In order to achieve this goal, five paths are taken:

1. Meta-models that cross-use computationally expensive observations across categories.

2. Adaptive sampling on meta-models that cross-use observations across categories.

3. Several types of mutation operators for the mixed-integer-categorical genetic algorithm are proposed.

4. Meta-models that re-use computationally expensive observations from previously explored concepts.

5. Adaptive sampling on meta-models that re-use observations from previously explored concepts.

Meta-models that cross-use computationally expensive observations across categories. It has been previously mentioned that using intense computational tools in early design stages constrains the function call budget. Thus, attention is taken on designing accurate surrogates capable of leveraging the similarities across categories, producing more accurate approximation for a given function call budget. In current methods, the training set for each category does not interact with that of other. Thus, these current independent surrogates for each category are not effective for cases where objective function trends resemble, which happens usually in engineering scenarios. Including the non-numeric design variables in a unique surrogate would allow cross-using information of dependence along design variables across category members. It could result in surrogates that are more efficient than traditional independent ones for each category. "Efficiency" means fewer observations are needed for a given accuracy or more accuracy for the same number of expensive observations.

The first option to build the unique surrogate that includes not only continuous

and integer variables but also nominal ones is the Hamming distance. Sometimes, non-numeric inputs could intrinsically have a distance provided by a measurable property of the given category. This intrinsic nominal distance can be brought as the second option to build the unique surrogate. When the nominal distance functions (Hamming distance and the intrinsic nominal distance) are defined, categorical variables can be included in the surrogate domain. The efficiency of these unique surrogates is measured against the current independent surrogates. Also, the influence of the size of the training set is studied. Gaussian surrogates are chosen since adaptive sampling is intended. If this unique surrogate is successfully built, the available resources will be better employed when dealing with computationally burdensome tools.

Adaptive sampling on meta-models that cross-use observations across categories. The first step to build tools for the more efficient use of computationally intense codes is to develop meta-models that cross-use observations across categories. The second step to achieve the research goal is to extend EGO algorithms to handle the meta-models developed in the first step. The adaptive sampling algorithm chosen in this work applies the expected improvement infill criterion on Kriging surrogates. It allows designers to better explore the design space and exploit local optima in potentially promising regions for restrictive function call budget.

However, the surrogate models that leverage similar trends across categories have mixed-integer-categorical domains; thus, the infill criterion must be optimized in these domains. An optimization algorithm capable to optimize design spaces with not only continuous and integer variables but also categorical ones is developed. This genetic algorithm is constructed out of a mixed-integer GA by conveniently including new generation, mutation, and crossover operators for the non-numeric input parameters. This genetic algorithm, if successful, allows designers to develop more efficient conceptual design tools for intense multi-modal objective functions where there exists a

choice of discrete alternatives.

Several types of mutation operators for the mixed-integer-categorical genetic algorithm are proposed. The identification of the possible speed-up switches for the optimization algorithm is pursued. The speed of this algorithm may increase with certain mutation operators. Several categorical mutation distributions could be obtained based on the two nominal distance functions: Hamming and intrinsic distance. Therefore, it makes sense to carry out an study to see which categorical mutation operators are more effective for some test problems. Also, as it was previously mentioned, expected improvement landscapes have identical properties regardless the problem. Due to the similar ExI characteristics (independent of the objective function), it is reasonable to extrapolate the results of the case study to other problems different from the tested ones.

Meta-models that re-use computationally expensive observations from previously explored concepts. Attention is on designing accurate surrogates capable of leveraging similarities from previously explored concepts. Currently, surrogates are fitted to each concept without reusing observations from previous concepts; however, previous concepts usually experience similar trends to the new ones. The current approach is a strategy that provides poor surrogate accuracy if the function call budget is limited (usual situation while using computationally expensive tools). A Gaussian multi-fidelity approach is developed to reuse previous concept observations when fitting a surrogate for a new concept. The efficiency of these multi-fidelity surrogates is measured against the current independent surrogates. Also, the influence of the size of the new and previous concept training sets is studied. If the resulting multi-fidelity surrogate of the new concept is successful, the limited function call budget is more conveniently exploited (more surrogate accuracy) when using computationally expensive models.

Adaptive sampling on meta-models that re-use observations from previously explored concepts. Once the meta-models that reuse computationally expensive observations from previously sampled concepts are implemented, the next step is to apply EGO algorithms on these meta-models. Again, the expected improvement is the infill criterion chosen to adaptively sample the Gaussian multi-fidelity surrogate.

The domain space of the surrogate that leverages similar trends from previously sampled concepts could have continuous, integer and/or categorical design variables. The most demanding case is when the surrogate domain is mixed-integer-categorical; however, the adaptive sampling on this type of domain was already solved in the second item of the above enumeration. Therefore, no technical challenges are expected when adaptively sampling meta-models that re-use observations if concepts are explored one at a time.

1.6.1 Research Questions

Once the problem to tackle is defined, research questions help to guide the investigations herein. It is important to remember the goal of this research: build conceptual design tools for the more efficient use of computationally intense codes with the purpose of improving the quality of initial design and concept/category selection.

1.6.1.1 First Question

Is it possible to build efficient surrogates for design scenarios where there exists a design categorical choice with similar trends?

a) Is it possible to build surrogates that cross-use computationally expensive observations across categorical choices with similar trends?

b) Which nominal distances allow building these efficient surrogates?

c) Does the MIC surrogate outperform independent surrogate modeling for each category? How does the relative efficiency of MIC surrogates with respect to

the independent ones (state-of-the-art) depend on the training set size?

1.6.1.2 Second Question

How could adaptive sampling approaches be efficiently extended to a choice of categories that experiments similar trends?

a) Can the MICGA extended from MIGA drive the ExI search while doing adaptive sampling on MIC surrogates?

b) Is the adaptive sampling on MIC surrogates more efficient than simultaneous adaptive sampling on independent surrogates (state-of-the-art) in some range of training set sizes?

1.6.1.3 Third Question

Is it possible to build efficient surrogates for design scenarios where there exist computationally expensive observations from a previous concept with similar trends?

a) Is it possible to build surrogates that re-use computationally expensive observations from previous concepts with similar trends?

b) Do ECMF surrogates outperform mono-fidelity surrogates? How does the relative efficiency of ECMF surrogates with respect to the mono-fidelity ones (state-of-the-art) depend on the new concept training set size?

c) What is the influence of the old concept training set size in the performance of ECMF surrogates?

1.7 Practical Applications

First of all, it is worth mentioning that computationally intense codes are not required to better interpret canonical test results or further support the research hypothesis or predictions. Also, the use of high-fidelity codes in conceptual design has philosophical

problems as mentioned in Section 1.1 due to model reduction and unknown unknowns. However, tests on computationally burdensome models would support the practicality of the engineering motivation.

In order to find a proper practical application of this research, it is recommended to keep in mind its goal: develop tools for the more efficient use of computationally intense tools early in the design process with the purpose of enhancing the initial design and concept/category selection. Thus, the target is on engineering problems where discrete variables and different configurations are possible. Tools with long evaluation time are the ideal aim of this work, whereas heuristic models in the form of look-up tables or really simple models must be ruled out for application purposes.

Rotor-craft design fulfills these two requirements: presence of design alternatives and long concept evaluation times. Firstly, rotor and anti-torque configurations, number of blades, types of airfoil, types of articulation, gear box arrangement, and manufacturing processes, among others, represent several concepts and discrete variables. Also, some manufacturing constraints could convert some firstly thought continuous variables into discrete ones, as it occurs with commercially available standardized manufactured pieces. Secondly, the multidisciplinary nature of the design, where there is interaction between aerodynamic, structure mechanics, control, and aeroacoutics, among other fields, makes function evaluations intense, even when not too complex physics are used.

Additionally, rotor-craft design is usually intended to be multi-objective. Typical objectives are fuel consumption and noise emissions at several flight conditions, and system cost, among others. So, multi-objective optimization requires more evaluations than single-objective scenarios; thus, even mid-fidelity tools become less manageable in rotor-craft multi-objective design.

Therefore, the use of high-fidelity tools, such as CFDs, for the multi-objective optimization of multi-disciplinary objectives would be an impractical task, so it is

not considered in this work. Another reason that hinders the use of CFD codes is the limited accessibility to rotatory CFD methods. The only available CFD code for rotatory wing is a full Navier-Stokes solvers, whose running time makes it impractical in conceptual design. The last but not least reason is the large amount of time necessary to integrate CFD related set-ups into the tools to develop, especially when the available time for the research is limited as in this case.

The finding of prohibitive high cost changes in late design stages (specially for discrete variables) and the increase in computational efficiency have prompted helicopter companies to be more and more interested in getting better sizing and predictions tools in early design stages. This could shorten the design process and enhance the product quality. These tools provide more accurate initial designs that, taken to later design stages, reduce the likelihood of last-minute large design changes with their corresponding high impact on the cost. Due to all these reasons, helicopter design is a suitable application for the method developed in this thesis.

Regarding the choice of objective functions, interest is in power consumption and noise reduction. The software PSU-WOPWOP done by Dr. Brentner is available to the author. It implements the Ffowcs Williams Hawkings (FWH) analogy [25]. It is well-known that in order to get reliable acoustics results, a high-fidelity aeroelastic model is needed to properly catch the blade pressures and displacements. However, CFD analysis were ruled out in favor of mid-fidelity models. Hence, the acoustic results would not be accurate. Thus, the acoustic objective is not pursued in this work. Only the optimization of the rotor-craft power consumption at several flight conditions is pursued.

The software used to model the rotor-craft is FLIGHTLAB[1]. It is a rotor-craft simulation tool capable of multidisciplinary support with selective fidelity modeling options. FLIGHTLAB is not a conceptual design software per se, but other conceptual design software, such as RCAS, is not available to the author. Conceptual

design software, such as NDARC[97] and CIRADS[38], are available. They are used for sizing, and parametric selection of alternatives and initial designs. NDARC and CIRADS include very simple models where there are only a few design variables. For instance, NDARC or CIRADS do not permit detailed aerodynamic curves. More design variables and complex physics models can be added through interfaces with other software. However, these complex software packages have distribution restrictions. Thus, FLIGHTLAB is chosen as the modeling tool; it allows more complex physics and more design variables in the parametric selection of alternatives and initial designs.

The first practical test carried out is the multi-objective optimization of the UH60A power consumption at hover and cruise speed. For the given weight the six degree of freedom rotor-craft must be trimmed, i.e., iterations between disciplines are necessary till agreement is reached. So, even the use of mid-fidelity codes for aerodynamics and structural dynamics result in computationally long function evaluations. The categorical variable in this work is the main rotor airfoil blade. The UH60A power consumption is used to test the efficient surrogates that leverage similar trends across categories and do multi-objective adaptive sampling on them.

A new "evolutionary incremental concept" from the conventional UH60A is tested: UH60A with a fenestron tail. In this second practical experiment, the optimization of the power consumption of the UH60A with fenestron tail is carried out at hover and cruise speed. The considered previous concept is UH60A with conventional tail. The UH60A with a fenestron tail power consumption is used to test the efficient surrogates that leverage similar trends from previous concepts with similar trends and do multi-objective adaptive sampling on them.

1.8 Thesis Organization

Following the present chapter, the remaining chapters go in more details through the research. Chapter 2 provides a literature review with a summary of the necessary background and current available theory to develop the desired techniques. It begins with a classification of types of discrete variables and the definition of order and distance. Then, distance functions are presented with focus on nominal distances used in classification techniques Later, the typical estimates utilized by engineers to size systems and evaluate objectives are reviewed. Emphasis is on the capability of reusing observations of concepts with similar trends. Then, the framework of conceptual design is presented. It is followed by an introduction of the GRFM used in this work, Kriging, to get familiar with it and its nomenclature. Kriging regression aimed to noisy functions is also explained. After it, an overview of typical multi-fidelity techniques, specifically Gaussian approaches, is presented. Again multi-fidelity regression is explained.

Later, a description of adaptive sampling techniques centered on the single-objective expected improvement infill criterion is provided. Then, a summary of the most common mixed-integer optimization techniques is presented. These techniques are classified according to convex problems and non-convex problems that, in turn, are divided by deterministic and stochastic searches. It is followed by the discussion of a multi-objective extension of the expected improvement criterion based on the increment of dominated hyper-volume. The last part of the chapter is devoted to a review of fenestron tails.

Following this literature review chapter, a detailed description of the development of the new techniques is provided in Chapter 3. Firstly, the chapter start with an introduction. Secondly, the research questions that guide the investigations, their hypothesis, and predictions are discussed. Thirdly, several nominal distance functions are proposed to enable the development of a common Gaussian surrogate that

leverages similar trends across categories.

Fourthly, the implementation of this common surrogate is explained together with the definition of the success indicators; this meta-model is called mixed integer-categorical surrogate (MIC). Then, it is presented the implementation of a new concept surrogate that leverages observations from previously sampled concepts. A multi-fidelity approach is developed to accomplish the meta-model; it is called evolutionary concept multi-fidelity (ECMF) surrogate. Then, adaptive sampling on the proposed surrogates is discussed with emphasis on the extension of mixed-integer genetic algorithms to a mixed-integer-categorical ones. Later, a study is carried out regarding the influence of the mutation operator on the performance of the mixed-integer-categorical genetic algorithm. The principle of maximum entropy is employed to come up with possible new mutation procedures. It is followed by the discussion of the choice of the fenestron tail configuration as the new "evolutionary incremental concept". Emphasis is on the assessment of the fenestron baseline values and the weight estimation. Finally, the methodology diagrams are presented.

Chapter 4 presents the FLIGHTLAB rotor-craft computational model employed in this research. Firstly, the UH60A baseline characteristics and parameters are reviewed. Then, the computational model is validated at hover and cruise speed. Finally, the noisy nature of the computational model is shown.

The MIC surrogate experiments are presented in Chapter 5. MIC surrogate efficiency is compared with that of independent surrogates for each category in terms of the success indicators. Experiments to understand the influence of the training set size and the nominal distance in the MIC performance are carried out. Results are discussed. The tests are done on a noise-free function (disturbed Branin function) and on a noisy function (UH60A hover power consumption). A screening of the general UH60A hover power consumption is performed with the purpose of testing MIC surrogate on a handy function with a few design variables.

Chapter 6 presents the mixed-integer-categorical genetic algorithm (MICGA) as the driver of EGO algorithms on MIC surrogates. Firstly, an intermediate mixed-integer GA is validated for assuring appropriate further use. Then, EGO behavior is shown when adaptively sampling the disturbed Branin function and the UH60A hover power consumption.

The ECMF surrogate experiments are presented in Chapter 7. ECMF surrogate efficiency is compared with that of surrogates where there is no reuse of observations from previous concepts in terms of the success indicators. Experiments to understand the influence of the new and old concept training set size are carried out. Results are discussed. The tests are done on a noise-free canonical function (2-dimensional Michalewicz function) and on a noisy function (UH60A with fenestron hover power consumption).

Chapter 8 demonstrates MIC adaptive sampling on a rotor-craft practical scenario: the multi-objective optimization of the UH60A power consumption at hover and cruise speed. The objectives are optimized in two design spaces: a) the one with four design variables resulted from the screening process presented in Chapter 5; and b) a design space with seven design variables. The categorical variable is again the main rotor airfoil blade. In the screened domain case, the Pareto front assessed with the adaptive sampling on the MIC surrogate is compared with the one obtained from simultaneous adaptive sampling on independent surrogates (current state-of-the-art).

The second practical conceptual design scenario is presented in Chapter 9: the multi-objective optimization of the UH60A with fenestron power consumption at hover and cruise speed. Again, two design spaces with different dimensionality are tested. Also, a screened domain is employed to compare the quality of the Pareto front obtained from adaptive sampling on ECMF surrogate and on a surrogate that does not include observations from previous concepts. The considered previous concept is UH60A with conventional tail.

The final part of the thesis, Chapter 10, summarizes the research and its contributions. The recommendations on when to use the MIC and ECMF surrogates are discussed. Finally, the future work is argued.

CHAPTER II

BACKGROUND AND THEORY REVIEW

This Chapter contains a summary of the necessary background and current available theory to develop the desired techniques. It covers the following topics: classification of types of discrete variables; nominal distances and correlation functions in the context of classification techniques; summary of typical estimates utilized by engineers; review of the conceptual design stage; brief introduction of the GRFM with focus on Kriging; multi-fidelity meta-models, specifically Gaussian approaches; description of adaptive sampling techniques centered on the single-objective expected improvement infill criterion; summary of the most common mixed-integer optimization techniques; review of a multi-objective extension of the expected improvement infill criterion based on the increment of dominated hyper-volume; and finally, a review of fenestron tails is presented.

2.1 Discrete Variables

Discrete variables are those whose possible values are only distinct points on the scale, as opposed to continuous ones that can take an uncountably infinite number of values. Discrete variables can assume finite or countably infinite values. Any set which can be put in a one-to-one correspondence with the natural numbers (or integers) so that a prescription can be given for identifying its members one at a time is called a countably infinite (or denumerably infinite) set. Discrete variables can be ordered or unordered. The distance between set values can be equal, uneven or meaningless. Herein, the distance between members of a discrete design variable is employed.

Discrete variables can be classified into several groups according to several characteristics of the variable. In the literature, several classifications are found. The

ones proposed in this work are based on the distance and order characteristics of the variable. The classification is the following:

Categorical, nominal, or non-numeric variables Individual items can only be measured in terms of whether they belong to a category or not. There is no quantifiable distance or order for the categorical variable. No arithmetic or logical operation can be performed on this data. Only qualitative classification is possible and the category name is arbitrarily assigned. A categorical variable is a generalization of the binary variable that has more than two states. Examples of such variables in sociology are marital status (unmarried, married, divorce, or widower) and gender (male or female). For aerospace design, possible non-numeric variables can be the engine, material, and airfoil choice available for the aircraft designer.

Ordinal variables The values of these variables are ordered in a meaningful way. However, the intervals between members (the distance) are unquantifiable and uneven. Consequently, arithmetic operations can not be realized; however, logical operations are possible. A typical example of ordinal variable is the socio-economic status of families (upper, upper-middle, middle, or low class). Middle class income is higher than low class income, but one can not quantify what the difference is between high and low class families. Engineering management usually faces the decision of what qualification is required for a given job: B.Sc, M.Sc. or Ph.D.. The possible qualifications represent a discrete-ordinal variable.

Quantitative variables These variables can be ordered and the interval between scale points is meaningful. Therefore, both logical and arithmetic operations are possible. Integers are a special subgroup of quantitative variables where the distance between points is the same and equal to one; however, it does not apply

for all the discrete-quantitative variables. The number of children per family is an example of discrete-quantitative variable in sociology. Also, the standardized sizes of some engineering parts, the number of blades, the number of compressor stages, and the screw diameters and pitches are discrete-quantitative variables. The whole group of discrete-quantitative is sometimes referred as integer variables; however, in reality it is just a subgroup of it. It could create confusion. For instance, mixed-integer programming and mixed-integer stochastic searches can handle not only integer variables but also discrete-quantitative ones; however, from their name one can think that these techniques just handle integer variables.

2.1.1 Definition of Order and Distance

Formal definitions of order and distance can be found in the topology literature [108]. Order is a relation $<$ on a set X such that for each pair of points β_1, and β_2 of X, one and only one of the following holds:

$$\beta_1 < \beta_2, \qquad \beta_1 = \beta_2, \qquad \beta_2 < \beta_1$$

A metric or distance function is a measure d that satisfies the following properties for all points β_1, β_2, and β_3 of X:

$$d(\beta_1, \beta_2) \geq 0 \tag{1}$$

$$d(\beta_1, \beta_2) = d(\beta_2, \beta_1) \tag{2}$$

$$\text{if } d(\beta_1, \beta_2) = 0, \text{ then } \beta_1 = \beta_2 \tag{3}$$

$$d(\beta_1, \beta_2) + d(\beta_2, \beta_3) \geq d(\beta_1, \beta_3) \tag{4}$$

2.2 Nominal Distance Functions

Distance functions for continuous and discrete-quantitative variables are frequently used in engineering design. For instance, surrogates and mutations of mixed-integer

evolutionary algorithms (ES and GA) are normally based on continuous and integer distances. The euclidean distance is the most popular continuous distance between points. It is defined as

$$d_{ecl}\left(\mathbf{x}^{(i)}, \mathbf{x}^{(l)}\right) = \sqrt{\sum_{j=1}^{k}\left(x_j^{(i)} - x_j^{(l)}\right)^2} \tag{5}$$

where $\mathbf{x}^{(i)}$ and $\mathbf{x}^{(l)}$ are two points in the k-dimensional design space.

Regarding distance functions for discrete-quantitative variables, a common choice is the Manhattan distance

$$d_{manh}\left(\mathbf{x}^{(i)}, \mathbf{x}^{(l)}\right) = \sum_{j=1}^{k}|x_j^{(i)} - x_j^{(l)}| \tag{6}$$

which is computationally cheaper than the Euclidean distance. None of the so far presented distances, Equations 5 and 6, are appropriate for categorical attributes because there is no possible order, so one can not place the categorical variable in a coordinate.

2.2.1 Classification Techniques: Nominal Distances

Classification is a machine learning problem where a new observation is identified with a class of a set of possible classes. It is done on the basis of a training set of data containing observations whose category memberships are known. The observations are analyzed in terms of a set of quantifiable features. These features could be categorical, ordinal, discrete-quantitative and/or real.

These classification techniques require the definition of nominal distances for categorical attributes. Researchers in machine learning and pattern recognition have investigated possible choices of nominal distances. The simplest nominal distance is an arbitrary mapping of the categorical variable to a subset of integer values whose size is the same as the number of categories.

A common nominal distance is the Hamming distance, also called overlap distance in machine learning, where the distances between all the categorical attributes are the same. For one categorical variable, it is defined as follows

$$d_{ham}(a,b) = \begin{cases} 0 & \text{if } a = b \\ 1 & \text{if } a \neq b \end{cases}$$

where a, and b are categories.

Also, it is common to employ the Value Difference Metric (VDM) for nominal attributes, which was developed by Stanfill an Waltz[179]. It is defined as follows:

$$d_{vdm}(a,b) = \sum_{c=1}^{C} |P_{a,c} - P_{b,c}|^2 \tag{7}$$

where d_{vdm} is the VDM distance, C is the number of possible classes, c is an index that goes through the classes, $P_{a,c}$ is the conditional probability that the output class is c given that the attribute takes the categorical value a, and a and b are values of the categorical attribute. The values of the conditional probabilities can be assessed from the training set.

Additionally, more sophisticated nominal distance functions are found in machine learning literature: Minkowsky [13], Mahalanobis [142], Camberra, Chebychev, Quadratic, Correlation, Chi-square [134], hyper-rectangle distance functions [164]. These distance functions for categorical variables have not been employed in engineering design.

Typical classification scenarios have not only nominal attributes but also discrete-quantitative, and continuous. In this cases, Aha et al. [2] develop a heterogeneous euclidean-overlap metric not only for continuous and nominal variables but also for discrete-quantitative variables. Without loss of generality, one can order the design space so that the first n_{c} dimensions host the continuous design variables, the following n_{n} dimensions contain the nominal ones, and the last n_{q} the discrete-quantitative ones, i.e.

$$\mathbf{x} = \left[x_{cont,1}, ..., x_{cont,n_{\mathrm{c}}}, x_{nom,1}, ..., x_{nom,n_{\mathrm{n}}}, x_{quant,1}, ..., x_{quant,n_{\mathrm{q}}}\right]$$

where $\mathbf{x_{cont}}$, $\mathbf{x_{nom}}$, and $\mathbf{x_{quant}}$ are the continuous, nominal and discrete-quantitative design variables, respectively. The heterogeneous euclidean-overlap metric is given by expression 8

$$\sqrt{\sum_{j=1}^{n_c}\left(d_{ord}\left(x_{cont,j}^{(i)},x_{cont,j}^{(l)}\right)\right)^2+\sum_{j=1}^{n_n}d_{ham}\left(x_{nom,j}^{(i)},(x_{nom,j}^{(j)}\right)+\sum_{j=1}^{n_q}\left(d_{ord}\left(x_{quant,j}^{(i)},x_{quant,j}^{(l)}\right)\right)^2} \tag{8}$$

where $d_{ord}\left(x^{(i)},x^{(l)}\right)=\dfrac{x^{(i)}-x^{(l)}}{\text{range}\,(x_j)}$, and $\text{range}\,(x_j)$ is the range for the variable x_j.

Also, this heterogeneous euclidean-overlap distance has been used for evolutionary strategies applied on dynamic niching, see Li et al.'s work[126]. In each population, the niche radius is obtained making use of the heterogeneous distance function.

Classification techniques have normally two stages: inference in which the training data is used to learn a probabilistic model, and decision in which the probability distribution is used to make a class assignment. The techniques that solve the two stages at the same time and simply provide a function that maps the new observation directly to decision are called discriminant. For more details about classification techniques, machine learning, and pattern recognition, see references [17, 135, 130, 18].

Common classification techniques that use distance functions are the following:

k Nearest Neighbors (k-NN) The technique consists in finding the k instances in the training dataset that are the closest to the new observation. After, these k instances vote to determine the class of the new observation [32]. In order to determine the k closest instances, distance functions are needed. Hamming distance and its improved version via attribute weighting are usually used when dealing with categorical attributes[124].

Kernel Methods The classification outcome is based on kernel functions. These kernel functions normally require a metric that measures the similarity of two vectors in the design space. Several types of kernels are used in classification

techniques. Among them, a popular one are the radial basis functions. Its main limitation is that kernel functions have to be evaluated for all the members of the training set, which can be computationally infeasible. For more details, see reference[166].

Support Vector Machine Kernel-based algorithms that have sparse solutions, so there is no need to evaluate kernel functions for all the training points. The most common approach with nominal attributes is to assume equal distance between nominal members, as explained in Tian and Deng's work[183].

In other classification instances, the boundaries for classifying a member are not sharp, i.e., the observation can belong to or not belong to the class. Fussy sets are developed for these cases. Membership functions are introduced to deal with this uncertainty of the classification. These functions are subjective as the boundaries between members; also, these functions are generalizations of the indicator function in classical sets [172]. Fuzzy sets and their corresponding membership functions are employed in the context of control theory, and medical diagnosis. For more details, see Zadeh's work [210].

Herein, interest is on design, where no classification is needed. However, the nominal distances employed in these classification techniques are useful for the purpose of this research.

2.2.2 Correlation Functions: Categorical Variables

Correlation is the degree to which two or more variables are linearly associated. In the case of a two-dimensional space, correlation coefficient quantifies the degree of correlation between the values of the two variables. While calculating the correlation coefficient, the numerical values of the variables are needed. For more details, see [178, 110].

Dependency and/or association is treated differently for categorical variables because there is no numerical value to input to the correlation coefficient. Their association is normally done by the chi-square test [137]. It tests the null hypothesis that there is no relationship between two categorical variables.

In order to account for the epistemic uncertainty of a function, Gaussian process meta-models treat deterministic responses as outcomes of a stochastic process. Then, correlation functions between the stochastic processes at different design locations are used to quantify the influence of a sample point on its neighbor. The correlation functions depend on the distance between sample points; then, the prediction is based on observations from nearby locations[202]. Radial basis functions are a common choice of correlation functions: their value is one for zero distance (very close correlation between sample points), and their value tend to zero as the distance between sample points becomes large (the sample points are not correlated)[65].

Categorical variables have not been modeled by surrogates in design. When it comes to deal with categorical variables in classification techniques, some authors have used nominal distances, such as Hamming and VDM, for classification using k-NN[201] and SVM techniques[183], respectively. These techniques use nominal distances as inputs for their kernel functions. Also, nominal distances have been used as inputs for correlation functions when doing ranking learning [30], also known as ordinal regression.

Herein, nominal distances allow bringing correlation functions based on distance to build Bayesian meta-models for engineering design when there is a categorical choice. Engineering observations of similar trends across several categories encourage the use of approximating methods that employ neighboring solutions, such as correlation functions based on distance. These correlation functions allow cross-using computationally expensive observations across categories.

2.3 Types of Estimates. Reuse of Data in Engineering

Sizing engineering systems and estimating/predicting engineering objectives are a crucial part of the design process. Engineers want estimation tools that are fast, accurate, and reliable in order to properly design systems. As explained in Section 1.2, some engineering objectives experience similar trends across different discrete alternatives. Thus, the ability of reusing other alternative observations is included and emphasized as a desired estimate quality in this research. The desired qualities in this thesis for the estimating techniques are presented in the following description:

Reuse other alternative data The ability of reusing data from similar alternatives makes the design process more efficient, especially when computationally expensive observations are brought.

High-fidelity estimate The closer the estimation is to the actual value, the more value the estimate holds.

Low computational effort Engineers want fast estimation tools to design quickly and be able to explore large design spaces.

No experience needed It is desirable that the estimating process does not rely on experts' opinions, which are usually subjective.

Deal with revolutionary designs The increasing requirements make design teams to look for completely new concepts to meet the new high standards. The estimation tool should be able to accurately predict these revolutionary designs.

Include the effect of many significant variables The addition of many significant design factors assures a more detailed parameterization of the design space. Thus, more optimal designs are possible.

Estimate typical engineering objectives Engineers want not only sizing tools but also estimates of real and practical engineering objectives.

There exist several estimation methods in engineering. The first one is called analogy-based prediction; it solves a new problem using and adapting solutions from previously solved similar problems [131]. With the help of experience, similar previous problems are chosen, relevant cost drivers are selected, and appropriate adjustments to reflect technical differences are applied. It normally produces low-fidelity estimates at low computational cost. The quality of the estimate depends on the availability of a similar previous problem. The number of significant design variables included in the estimate depends on the previous problem solution and the skills of the design team. These techniques are normally used for cost and weight estimation in software development [112] and engineering projects such as space missions [169].

The second estimating method is the regression-based prediction (also called statistics-based predictions). It consists in developing an estimation relationship based on sufficient historical data of similar systems and projects [169]. They are easy, fast, and simple to use at the expense of obtaining a low-fidelity estimate. However, some experience is needed to specify the fitting model. If the historical database is not big enough the estimate loses value. Also, the fact it is based on historical databases hinders the validity of the prediction when assessing revolutionary alternatives. Normally, a few significant design factors are included, and it is mainly used for sizing purposes. These regression-based methods have been widely used in helicopter design [146, 185, 150, 97, 102].

The last prediction approach is modeling and simulation (M&S), where normally high-fidelity tools are employed in the estimation. M&S provides virtual duplication of products and processes, and represents those products or processes in readily available and operationally valid environments [143]. It uses models, including emulators, prototypes, and simulators, to develop data as a basis for making managerial

or technical decisions. The main drawback of these predicting techniques is the high computational effort when modeling systems with high-fidelity. The labor effort to set the modeling environment is high, but there is no need of experts' subjective opinion. As many as desired design variables can be added in the simulation, and revolutionary designs can be explored with M&S techniques. Also, typical and practical engineering objectives can be output from these techniques. Multiple examples of M&S on engineering systems can be found in the literature [129, 85]. So far, it does not reuse observations from previous similar alternatives. This work focuses on bringing to M&S techniques this new capability of reusing previous observations of alternatives with similar trends.

Figure 3 summarizes the advantages and disadvantages of the these estimation techniques.

Desired Qualities	Analogy-based	Regression-based	M&S
Reuses other alternative data	Yes	Yes	No
Fidelity estimate	Low	Low	High
Computational effort	Low	Low	High
Need of experience	A lot	Some	None
Deals revolutionary designs	No	No	Yes
Many significant design variables	Depends	Few	Many
Estimates engineering objectives	No	No	Yes

Figure 3: Advantages and Disadvantages of Estimate Types

2.4 Conceptual Design

The design process is broken into 3 main phases: conceptual design, preliminary design, and detailed design. Conceptual design is an iterative process where design requirements, trade-offs, and analysis are used to guide and evaluate configuration arrangements. Basic questions of performance and configurations are answered in the conceptual design stage. In this design stage, key decisions are made, fixing up to 80% of the life cycle cost [190]. Several concepts and categorical alternatives are studied during the conceptual design phase. Figure 4 shows the typical conceptual design wheel.

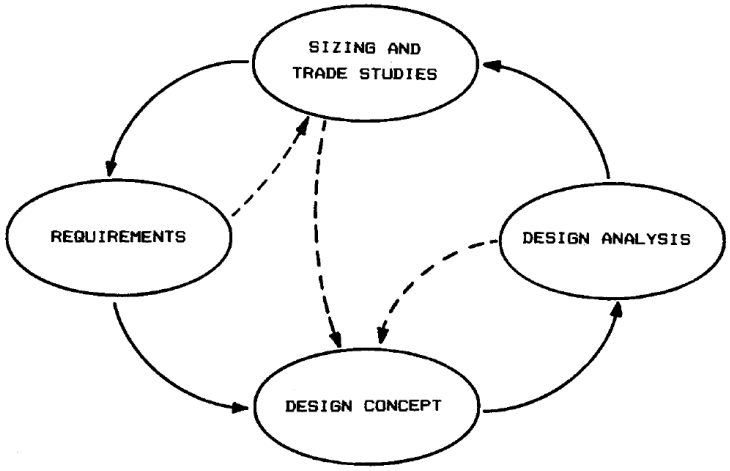


Figure 4: Conceptual Design Wheel [151]

Concepts evolve incrementally to meet requirements, achieve satisfactory trade-offs, and improve the performance of weak designs obtained from previous concepts. The process is very fluid and the configuration incorporates changes that potentially can make the design meet the requirements and produce satisfying trade-offs. In this thesis, the new concept obtained from incremental changes on the previous concept is called "evolutionary incremental concept". The new and old concept are expected to experience similar trends if these incremental changes are not too large.

The conceptual design emphasis is normally on the crucial components interactions instead of on deep details in each system configuration and geometry. Finding these interactions normally requires a lot of experience and sharpness. As the process goes on, the analysis sophistication keeps increasing. In the aircraft industry a typical conceptual design period to select the best concept is around six months. Good conceptual design reviews, specifically for aircraft design, are provided by Raymer [151], and Roskam [157].

2.5 Gaussian Meta-Models: Kriging

In science and engineering, researchers use computer simulation codes instead of expensive physical experiments in order to improve the quality and performance of engineering products. Scientists look for more and more flexible and accurate models to study a given phenomenon. Unfortunately, this tendency makes computer simulations to take a substantial computational time. One simulation may take many minutes, hours, days or even weeks, rendering design studies uneconomical.

Normally, this issue is solved by the construction of simpler approximation models that mimic and predict the expensive code output. The idea is to develop a cheaper relationship between the system inputs and outputs. If the model is properly built, these approximation models mimic the behavior of the intense simulation accurately, and at the same time they can be evaluated faster. Several approximation methods exist in the literature, each one with its pros and cons. Among the most popular methods, one finds polynomial response surfaces [24], Gaussian random field meta-models [160], support vector machines [187], and artificial neural networks [17].

An important class of surrogates are the Bayesian ones. As opposed to other surrogates, they quantify the uncertainty due to the lack of knowledge about the function to approximate. Instead of estimating unique values of the meta-model parameters, Bayesian surrogates provide, according to the data already observed, a probability

distribution that reflects the degree of certainty in the surrogate parameters, i.e., it provides a measure of confidence for its parameters. Prediction-wise, the Bayesian surrogate output is also a probabilistic distribution instead of a deterministic one. Poorly sampled areas of the design space have more uncertainty in the predictive distribution. The prediction uncertainty is necessary for the use of adaptive sampling techniques, where the confidence in the prediction can drive the sampling plan, see Section 2.7.

Figure 5 shows a predictive distribution for a given observed data. The 95% confidence interval indicates the reliability of the estimate. Sparsely sampled regions have a larger confidence interval than highly sampled regions.

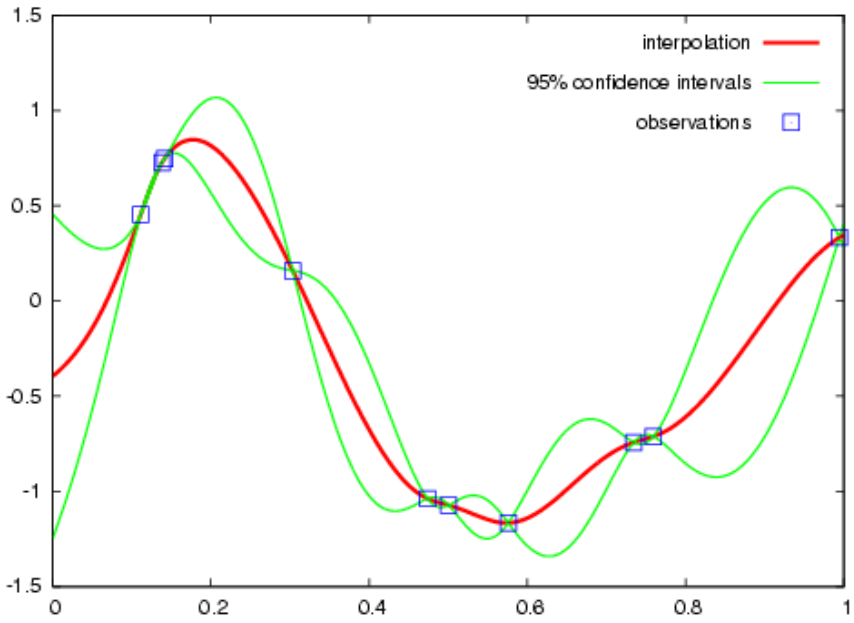


Figure 5: Bayesian Predictive Distribution

A very prolific type of Bayesian surrogate is the GRFM. GRFM assumes that the observed data is a noisy realization of a deterministic model. This realization is modeled as a normally distributed random vector [104]. However, it must be clear that GRFM builds surrogate models of deterministic computer functions, even though the technique is based on stochastic processes. As a Bayesian surrogate, its prediction

is a probabilistic distribution.

GRFM, also known as Kriging[1], has become popular because of its capability of predicting the uncertainty of the prediction while being able to cope with non-linear problems. Much of the efforts in Bayesian modeling have been devoted to this type of surrogates. Also, there exists a rich literature about it. These characteristics make Kriging really appealing for engineering design problems. Krige first developed it for mining problems [117]. Sacks et al. were the first ones, before many, to apply Kriging to computer experiments[160].

The main Kriging drawback is the curse of dimensionality as shown by Koch et al. [116], and Wang et al. [194]. As the dimension of the design space and number of sample points increase, the Kriging tuning process becomes more and more resource-demanding. The tuning process is done with the maximum likelihood estimation (MLE), that is later explained. It requires the optimization of the observed data likelihood, which is normally performed with a stochastic search. Each likelihood evaluation needs the factorization of a matrix whose size is the number of sample points. For large number of sample points, the cost of the MLE is driven by these time-consuming matrix factorizations. Also, a large dimensionality of the design space provokes that, when optimizing for tuning the surrogate, the search is in a large design volume which makes the process even more resource-demanding.

In spite of the curse of dimensionality, the widespread use of Kriging in the literature and its robustness are enough benefits to choose it as the Bayesian surrogate. In the following, a mathematical description of it is presented. There are many variations of Kriging; the one presented herein is called ordinary Kriging.

Kriging bases its prediction on a training or observed set, $\mathbf{X} = \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, ...\mathbf{x}^{(n)}\}$. Each input variable has k dimensions. The responses at the observed points are given

[1]Currently, Kriging methods are wide. They don't necessarily assume Gaussian fields. However, herein, the values of the exponents of the correlation or basis functions are 2, making the Kriging a Gaussian process.

by $\mathbf{y} = \{y^{(1)}, y^{(2)}, ...y^{(n)}\}$. Let $\mathbf{Y}$ be a noisy realization of the deterministic model $y(\mathbf{x})$

$$\mathbf{Y} = \mu\mathbf{1} + \boldsymbol{\epsilon}$$

where $\mathbf{1}$ is a k-dimensional vector of ones; and $\boldsymbol{\epsilon}$ is the error between the stochastic process, $\mathbf{Y}$, and a constant vector $\mu\mathbf{1}$. ϵ_i is normally distributed with uniform variance σ^2, i.e., $\epsilon_i = \mathcal{N}(0, \sigma^2)$ for $i = 1, .., n$. The constant value of the mean μ is a specific characteristic of the ordinary Kriging.

The components of this vector are correlated according to Kriging basis functions [65, 184] given by

$$\boldsymbol{\Psi} \doteq cor\left[Y\left(\mathbf{x}^{(i)}\right), Y\left(\mathbf{x}^{(l)}\right)\right] = \exp\left(-\sum_{j=1}^{k} \theta_j |x_j^{(i)} - x_j^{(l)}|^{p_j}\right) \tag{9}$$

Note that $\boldsymbol{\theta}$ and $\mathbf{p}$ are free parameters. For simplicity, this work assumes $p_j = 2$ for all j, making the meta-model a Gaussian one. The parameter θ_j controls how much influence the design variable x_j has on the objective function. Large values of θ_j implies low correlations for the dimension j. Thus, by comparing different θ_j for a fitted model one can tell which design attributes are more influential in the objective. For instance, $\theta_j > \theta_k$ implies that the input x_k has a lower effect in the objective function $y(\mathbf{x})$ than x_j.

Once Kriging basis functions have been introduced, one must estimate proper values of hyperparameters[2], $\boldsymbol{\theta}$, μ, and σ, that lead to an accurate surrogate model given the training data, $\mathbf{X}$ and $\mathbf{y}$. There are two main approaches to estimate the unknown parameters: likelihood-based ones and fully Bayesian ones. The second ones are ruled out in this investigation for their high computational expense and their need of assigning a prior distribution.

The likelihood-based approaches are chosen to construct the surrogate model that

[2]Hyperparameters are the unknown parameters that are used to tune the model.

interpolates the data; thus, the likelihood of the observed data is needed. The training dataset is assumed to be generated independently. The likelihood of the observed data is given by the product of their probabilities according the Gaussian distributed noisy realization,

$$L\left(\mathbf{Y}^{(1)},...\mathbf{Y}^{(n)}|\mu,\sigma,\boldsymbol{\theta}\right)=\frac{1}{\left(2\pi\sigma^{2}\right)^{n/2}}\exp\left[-\frac{\sum\left(\mathbf{Y}^{(i)}-\mu\right)^{2}}{2\sigma^{2}}\right] \tag{10}$$

making use of the correlation matrix, $\boldsymbol{\Psi}$, the likelihood takes the form of

$$L\left(\mathbf{y}|\mu,\sigma,\boldsymbol{\theta}\right)=\frac{1}{\left(2\pi\sigma^{2}\right)^{n/2}\left|\boldsymbol{\Psi}\right|^{\frac{1}{2}}}\exp\left[-\frac{\left(\mathbf{y}-\mathbf{1}\mu\right)'\boldsymbol{\Psi}^{-1}\left(\mathbf{y}-\mathbf{1}\mu\right)}{2\sigma^{2}}\right] \tag{11}$$

The unknown parameters of the model are found by seeking for the values that maximize the likelihood of the observed data, Equation 11. It is the MLE procedure. MLE can be applied to not only Kriging hyperparameters tuning [65] but also any general Bayesian inference parameter estimation as Tipping shows [184].

For simplicity's sake, the natural logarithmic function, $\ln(x)$, is brought to the analysis. It is a monotonic increasing function of x, so one can maximize the $\ln(L(.))$ instead of the $L(.)$. The natural logarithm of the training data likelihood is

$$\ln L\left(\mathbf{y}|\mu,\sigma\right)=\frac{n}{2}\ln\left(2\pi\right)-\frac{n}{2}\ln\left(\sigma^{2}\right)-\frac{1}{2}\ln\ \left|\boldsymbol{\Psi}\right|-\frac{\left(\mathbf{y}-\mathbf{1}\mu\right)'\boldsymbol{\Psi}^{-1}\left(\mathbf{y}-\mathbf{1}\mu\right)}{2\sigma^{2}} \tag{12}$$

The hyperparemeters in the discussion are $\boldsymbol{\theta}$, μ, and σ. Derivatives respect to μ and σ are relatively simple, so the maximization against these last two hyperparemeters can be analytically tackled. The optimization results in the optimal hyperparemeters given by Equations 13 and 14

$$\hat{\mu}=\frac{\mathbf{1}'\boldsymbol{\Psi}^{-1}\mathbf{y}}{\mathbf{1}'\boldsymbol{\Psi}^{-1}\mathbf{1}} \tag{13}$$

$$\hat{\sigma}^{2}=\frac{\left(\mathbf{y}-\mathbf{1}\hat{\mu}\right)'\boldsymbol{\Psi}^{-1}\left(\mathbf{y}-\mathbf{1}\hat{\mu}\right)}{n} \tag{14}$$

Note that the hat, $\hat{}$, implies estimation for optimal values. Also, it is worth realizing that the inverse of the correlation matrix shows up. It is the reason because the previously mentioned matrix factorization is needed. Substituting optimal values into the natural logarithm of the likelihood, and removing some constant terms that are useless in the optimization process results in the following natural log of the likelihood

$$\ln(L) \approx -\frac{n}{2}\ln\left(\hat{\sigma}^2\right) - \frac{1}{2}\ln\ |\boldsymbol{\Psi}| \tag{15}$$

The hyper-parameter $\boldsymbol{\theta}$ is still unknown and can further maximize the likelihood. Taking derivative of $\ln(L)$ respect to $\boldsymbol{\theta}$ is analytically tedious, see Equation 15. As alternative a numerical optimization technique like a genetic algorithm can be used to maximize the likelihood against $\boldsymbol{\theta}$.

Forrester, Sóbester, and Keane [65] point out the correlation matrix $\boldsymbol{\Psi}$ varies in a logarithmic scale along $\boldsymbol{\theta}$. Hence, the optimization is done in terms of $\ln\theta_k$ instead of θ_k. Experience on typical engineering problems sets the search bound for optimal values in the range of $\boldsymbol{\theta}$ between 10^{-3} and 10^2. The optimal estimation obtained for this last hyperparemeter is denoted as $\hat{\boldsymbol{\theta}}$.

Once the hyperparemeters of the Gaussian meta-model are tuned, one can predict the expensive function over the design space. Let us say that the prediction, $\hat{y}(\mathbf{x})$, at the point $\mathbf{x}$ is wanted. The correlation vector between the point $\mathbf{x}$ and the sample points $\mathbf{X}$ is

$$\psi = \{cor\left[Y\left(\mathbf{x}^{(1)}\right), Y(\mathbf{x})\right], ...cor\left[Y\left(\mathbf{x}^{(n)}\right), Y(\mathbf{x})\right]\}$$

where the $cor[,]$ is given by Kriging basis functions, see Equation 9.

The idea is to maximize the likelihood of the augmented data (now the new point $\mathbf{x}$, where the prediction is pursued, is included). The only free parameter is the prediction value at the new point, $\hat{y}(\mathbf{x})$. The natural log of the augmented data

likelihood is

$$\ln(L) \approx -\frac{\begin{pmatrix}\mathbf{y}-\mathbf{1}\hat{\mu}\\ \hat{y}-\hat{\mu}\end{pmatrix}^T \begin{pmatrix}\boldsymbol{\Psi} & \psi\\ \psi^T & 1\end{pmatrix}^T \begin{pmatrix}\mathbf{y}-\mathbf{1}\hat{\mu}\\ \hat{y}-\hat{\mu}\end{pmatrix}}{2\hat{\sigma}^2} \tag{16}$$

where the terms non-dependent on the only free parameter, $\hat{y}$, are not written. Once the maximization is done, the prediction $\hat{y}(\mathbf{x})$ that maximizes the likelihood of the augmented data is given by

$$\hat{y}(\mathbf{x}) = \hat{\mu} + \psi^T \boldsymbol{\Psi}^{-1}(\mathbf{y} - \hat{\mu}\mathbf{1}) \tag{17}$$

The GRFM also estimates the error or uncertainty in the model, permitting to find new sample points not only in promising regions but also in sparsely sampled regions where the uncertainty in the prediction is high. Hence, it enables the exploration of the surrogate, while the prediction $\hat{y}(\mathbf{x})$ enables the surrogate exploitation. This error is called the mean squared error, $\hat{s}^2(\mathbf{x})$. It is given by

$$\hat{s}^2(\mathbf{x}) = \hat{\sigma}^2\left[1 - \psi^T \boldsymbol{\Psi}^{-1}\psi\right] \tag{18}$$

The mean squared error is inversely related to the curvature of the augmented $\ln(L)$. A proof of Equation 18 is given by Jones [98]. Note that $\hat{s}^2(\mathbf{x})$ is zero at sample points because the Kriging proposed method is an interpolation through the sample points.

2.5.1 Kriging Regression

The Gaussian interpolating surrogate fails when dealing with noisy data. It is well known that computer simulations, such as computational structural dynamics (CSD) and CFD codes, contain numerical noise. When observations are sparse, interpolating techniques can deal with small amplitude noise; however, when the optimization technique starts to converge to an optimum, the data becomes more dense at good

performing regions, and the interpolating model may fail to mimic the intense function [66].

The regressive model is achieved by adding a regression constant, λ, to the diagonal of the Kriging correlation matrix

$$\mathbf{R} + \lambda \mathbf{I} \tag{19}$$

λ is the regression constant that resolves the problem of approximating noisy functions. This constant avoids that the prediction at an observed design point passes through the observed objective value[66]. The Kriging correlation matrix, $\mathbf{R}$, is again chosen to be based on Gaussian radial basis functions

$$\mathbf{R} = R_{i,l}\left[Y\left(\mathbf{x}^{(i)}\right), Y\left(\mathbf{x}^{(l)}\right)\right] = \exp\left(-\sum_{j=1}^{k} \theta_j |x_j^{(i)} - x_j^{(l)}|^2\right) \tag{20}$$

Note that the matrix $\mathbf{R}$ is the same as the one for the interpolating case, $\mathbf{\Psi}$, defined in Equation 9. The likelihood of the observed data in the Kriging regression is given by

$$\ln L\left(\mathbf{y}|\mu,\sigma\right) = -\frac{n}{2}\ln\left(\sigma^2\right) - \frac{1}{2}\ln\ |\mathbf{R} + \lambda\mathbf{I}| - \frac{\left(\mathbf{y} - \mathbf{1}\mu\right)'\left(\mathbf{R} + \lambda\mathbf{I}\right)^{-1}\left(\mathbf{y} - \mathbf{1}\mu\right)}{2\sigma^2} \tag{21}$$

The unknown Kriging hyper-parameters are not only the same ones as for the interpolating Kriging, $(\boldsymbol{\theta}, \mu, \sigma)$, but also λ. These hyper-parameters are again obtained from the MLE of the observed data [66]. Regression adds one extra variable to the process of searching for the MLE; it increases the tuning computational cost. The optimal hyper-parameters $\hat{\mu}$ and $\hat{\sigma}$ are obtained similarly as in Section 2.5. The assessment of the optimal hyper-parameters $\hat{\boldsymbol{\theta}}$ and $\hat{\lambda}$ is done by a GA since the process is analytically intractable.

Similarly as in the interpolating case, the maximization of the likelihood of the augmented data is used to find the regression predictor, $\hat{y}$. It is given by

$$\hat{y_r}(\mathbf{x}) = \hat{\mu_r} + \psi^T \left(\mathbf{R} + \hat{\lambda}\mathbf{I}\right)^{-1} (\mathbf{y} - \hat{\mu_r}\mathbf{1}) \tag{22}$$

where

$$\hat{\mu_r} = \frac{\mathbf{1}' \left(\mathbf{R} + \hat{\lambda}\mathbf{I}\right)^{-1} \mathbf{y}}{\mathbf{1}' \left(\mathbf{R} + \hat{\lambda}\mathbf{I}\right)^{-1} \mathbf{1}}$$

$\mathbf{1}$ is a n-dimensional vector of ones. Notice that again the optimal hyper-parameter values are denoted by a hat, $\hat{}$.

ψ is the correlation vector between the point $\mathbf{x}$ where the prediction is wanted and observed points $\mathbf{x}^{(i)}$ for $i = 1, 2, ..., n$

$$\psi = \{\exp\left(-\sum_{j=1}^{k} \theta_j |x_j^{(1)} - x_j|^2\right), \exp\left(-\sum_{j=1}^{k} \theta_j |x_j^{(2)} - x_j|^2\right), ... \exp\left(-\sum_{j=1}^{k} \theta_j |x_j^{(n)} - x_j|^2\right)\}$$

Kriging models permit not only to predict the function value in new sample points, but also to estimate the uncertainty in the prediction. The uncertainty, $\hat{s}_r^2(\mathbf{x})$ in the regression case is given by Equation 23

$$\hat{s}_r^2(\mathbf{x}) = \hat{\sigma}_r^2 \left[1 + \hat{\lambda} - \psi^T \left(\mathbf{R} + \hat{\lambda}\mathbf{I}\right)^{-1} \psi\right] \tag{23}$$

where

$$\hat{\sigma}_r^2 = \frac{(\mathbf{y} - \mathbf{1}\hat{\mu_r})' \left(\mathbf{R} + \hat{\lambda}\mathbf{I}\right)^{-1} (\mathbf{y} - \mathbf{1}\hat{\mu_r})}{n} \tag{24}$$

It is worth mentioning that the model error in the regressive case, $\hat{s}_r$, is not zero at the observed points because the value of the regression constant, $\hat{\lambda}$, differs from zero; it produces a non-zero ExI. It can lead to optimizations trapped at sample points as is explained in the Subsection 2.5.2.

2.5.2 Kriging Re-interpolation

The Kriging regression mean squared error is not zero at the observed points because the value of $\hat{\lambda}$ is different from zero, see Equation 23. Therefore, the ExI at these points could differ from zero. It may result in maximum expected improvement at

design points previously sampled. This is a plausible scenario for non-deterministic experiments, where there is no repeatability. However, it stalls the EGO algorithm for repeatable deterministic computer experiments (the ones in this work) because the results of the observations at a given design point are always the same.

Forrester et al. [66] redefine the mean squared error when Kriging regression is used on deterministic experiments. They build an interpolation of the regressive Kriging surrogate through the values predicted by the regression meta-model at the observed points. This surrogate is called re-interpolation. The re-interpolating predictor is the same as the regressive one, i.e. $\hat{y}_{ri} = \hat{y}_r \forall \mathbf{x}$ and $\hat{\mu}_{ri} = \hat{\mu}_r$.

The re-interpolation mean squared error, $\hat{s}^2_{ri}$, that Forrester et al. [66] derive is given by

$$\hat{s}^2_{ri}(\mathbf{x}) = \hat{\sigma}^2_{ri}\left[1 - \psi^T \mathbf{R}^{-1}\psi\right] \tag{25}$$

where

$$\hat{\sigma}^2_{ri} = \frac{(\mathbf{y} - \mathbf{1}\hat{\mu_r})' \left(\mathbf{R} + \hat{\lambda}\mathbf{I}\right)^{-1} \mathbf{R} \left(\mathbf{R} + \hat{\lambda}\mathbf{I}\right)^{-1} (\mathbf{y} - \mathbf{1}\hat{\mu_r})}{n} \tag{26}$$

Equation 25 guarantees that the re-interpolating mean squared error $\hat{s}_{ri}$ is zero at the sample locations, unlike what happens for the regression case, $\hat{s}_r$. Thus, the ExIs (see Equation 46) at the observed points are zero in the re-interpolating case, and consequently, EGO process does not stall in this case.

2.6 Multi-Fidelity Meta-Models: Gaussian Approach

In many engineering situations more information than a mere vector of function values is available. It is not unusual to have access to several simulation models with different fidelities. Multi-fidelity techniques merge all the objective function information of varying fidelity with the purpose of constructing a surrogate model. These techniques help to soften the effect of the curse of dimensionality while enabling

the user to include high-fidelity tools. Techniques range from local approximations to more general approaches that model objectives globally.

In the present discussion two fidelities models are used to construct the approximation without loss of generality: high-fidelity, $\mathbf{y_e}$, and low-fidelity, $\mathbf{y_c}$. The number of samples of the high and low-fidelity model are n_e and n_c, respectively. The output of the multi-fidelity analysis is a surrogate model, again denoted by $\hat{y}$. This surrogate is a corrected low-fidelity approximation that takes into account high-fidelity objective information. The low-fidelity correction is usually done by two methods: additive scale factor $\Delta(\mathbf{x})$ (the difference between fidelities is modeled) and multiplicative scale factor $\rho(\mathbf{x})$ (the ratio between fidelities is modeled). A general framework for the correction is given by

$$\mathbf{y_e}(\mathbf{x}) = \rho(\mathbf{x})\,\mathbf{y_c}(\mathbf{x}) + \Delta(\mathbf{x})$$

Typical scenarios where designers find varying fidelity models are: a) governing equations that capture different amount of physical details; b) fully converged versus partially converged runs for iterative solvers, such as CFD and CSD simulations; c) heuristic approximation models versus physics-based ones; and d) converged solutions for different discretization sizes.

Within the local approximation techniques, Haftka [77] proposes zero order scaling with both multiplicative and additive scaling. In the same work, Haftka also presents a first-order scaling approach when the first-order derivatives of the varying fidelity models are available. Eldred et al. [53] formulate an extension of Haftka's work to the second-order scaling. The applicability of the last extension is reduced to models where the second-order derivative could be efficiently assessed. Same authors [53] develop a hybrid approach that combines the multiplicative and the additive correction. These methods improve the surrogate only in the neighborhood of single points because of which these techniques are called local approximation techniques.

Global surrogate models improve the accuracy of the low-fidelity model globally

by employing a training data-set. Global approaches allow improving the accuracy of the models with respect to the ones that are built only out of high-fidelity training data-sets. The reason is that the scale factors may behave better than the objective itself, as is shown by Keane [105] and Leary et al. [122]. One can find response surfaces, and Gaussian processes, among the typical multi-fidelity approaches.

Manson et al. [132] build response surfaces with two variable-fidelity aerodynamic codes. Knill et al. [114] use Euler equations to construct response surfaces that represent a correction to the linear theory, which reduces computational burden. Vitali et al. [189] propose a correction response surface to relate the high-fidelity models with the low-fidelity ones to study the crack propagation in the skin of the stiffened panel. Keane and Pretruzzelli build a multi-level wing design environment in the conceptual design of a commercial airliner wing [107]. However, the interest herein is in the probabilistic multi-fidelity methods.

In what follows a derivation of a multi-fidelity Gaussian process approach is presented. The corrected Gaussian process could be written as

$$\mathbf{y_e}\left(\mathbf{x}\right) = Z_\rho\left(\mathbf{x}\right)\mathbf{y_c}\left(\mathbf{x}\right) + Z_d\left(\mathbf{x}\right) \tag{27}$$

where Z_ρ and Z_d are the Gaussian process multiplicative and additive scale factor, respectively. For simplicity, the multi-fidelity algorithm developed herein assumes only the additive factor, i.e., $Z_\rho = 1$ and only Z_d is to be modeled. Eldred, Giunta et al. [53] show that the additive scale factor is appropriate in a wider variety of problems. Let $\mathbf{X}_e$ and $\mathbf{X}_c$ be the expensive and the cheap sample points, respectively. The values of the low and high-fidelity stochastic processes at the training set points are $Y_c\left(\mathbf{X}_c\right)$ and $Y_e\left(\mathbf{X}_e\right)$, respectively. They are treated as noisy realizations of the deterministic models y_c and y_e as in Section 2.5. The whole training data set could be written as $\mathbf{y}\left(\mathbf{X}\right) = \left\{\mathbf{y}_c\left(\mathbf{X}_c\right), \mathbf{y}_e\left(\mathbf{X}_e\right)\right\}$

In order to build this multi-fidelity approach, cheap and expensive data must

be correlated. Kennedy and O'Hagan [109] propose an auto-regressive model; it assumes that nothing can be learned about the objective function from the low-fidelity model provided that the high-fidelity model value y_e at the point $\mathbf{x^{(i)}}$ is known. Mathematically, this condition takes the form of $cov\left[Y_e\left(\mathbf{x^{(i)}}\right), Y_c\left(\mathbf{x}\right) | Y_e\left(\mathbf{x^{(i)}}\right)\right] = 0$, $\forall \mathbf{x} \neq \mathbf{x}^{(i)}$.

Using the mentioned auto-regressive model and $Z_\rho = 1$, the expensive Gaussian process surrogate, Z_e, takes the form

$$Z_e\left(\mathbf{x}\right) = Z_c\left(\mathbf{x}\right) + Z_d\left(\mathbf{x}\right) \tag{28}$$

where Z_c is the low-fidelity Gaussian process.

The covariance matrix of the whole data is given by

$$cov\left[Y\left(\mathbf{X}\right), Y\left(\mathbf{X}\right)\right] = \begin{pmatrix} \sigma_c^2 \mathbf{\Psi}_c\left(\mathbf{X}_c, \mathbf{X}_c\right) & \sigma_c^2 \mathbf{\Psi}_c\left(\mathbf{X}_e, \mathbf{X}_c\right) \\ \sigma_c^2 \mathbf{\Psi}_c\left(\mathbf{X}_e, \mathbf{X}_c\right) & \sigma_c^2 \mathbf{\Psi}_c\left(\mathbf{X}_e, \mathbf{X}_e\right) + \sigma_d^2 \mathbf{\Psi}_d\left(\mathbf{X}_e, \mathbf{X}_e\right) \end{pmatrix} \tag{29}$$

More details about the derivation of Equation 29 are found in Forrester, Sóbester, and Keane's work [64]. For compactness' sake, the covariance matrix of the whole data, $cov\left[Y\left(\mathbf{X}\right), Y\left(\mathbf{X}\right)\right]$, is denoted as $\mathbf{C}$.

Now, there are two correlation matrices as opposed to Section 2.5 where there is only one. The free parameters are μ_c, μ_d, σ_c, σ_d $\boldsymbol{\theta}_c$, and $\boldsymbol{\theta}_d$. The low-fidelity data is assumed to be independent of the high-fidelity one. So, MLE for μ_c, and σ_c can be assessed with Equations 13, and 14, respectively. Optimal values of $\boldsymbol{\theta}_c$, according to the MLE, are obtained with the same numerical process explained in Section 2.5. In such manner the low-fidelity surrogate $\hat{y}_c$ can be built.

In order to estimate μ_d, σ_d, and $\boldsymbol{\theta}_d$, the variable $\mathbf{d}$ is defined as the difference between the expensive observations and the cheap surrogate prediction in the expensive sample points $\mathbf{X}_e$, i.e., $\mathbf{d} = \mathbf{y_e} - \hat{\mathbf{y}}_c\left(\mathbf{X}_e\right)$

Ignoring terms that are already defined, and therefore constant, such as μ_c, σ_c, and $\boldsymbol{\theta}_c$, the likelihood of the expensive data can be written as

$$-\frac{n_e}{2}\ln\left(\sigma_d^2\right) - \frac{1}{2}\ln\ \left|\boldsymbol{\Psi}_d\left(\mathbf{X}_e, \mathbf{X}_e\right)\right| - \frac{\left(\mathbf{d}-\mathbf{1}\mu_d\right)'\boldsymbol{\Psi}_d\left(\mathbf{X}_e, \mathbf{X}_e\right)^{-1}\left(\mathbf{d}-\mathbf{1}\mu_d\right)}{2\sigma_d^2} \tag{30}$$

The MLEs for μ_d, σ_d are obtained by taking the partial derivatives of Equation 30 respect to μ_d, and σ_d.

$$\hat{\mu}_d = \frac{\mathbf{1}'\boldsymbol{\Psi}_d\left(\mathbf{X}_e, \mathbf{X}_e\right)^{-1}\mathbf{d}}{\mathbf{1}'\boldsymbol{\Psi}_d\left(\mathbf{X}_e, \mathbf{X}_e\right)^{-1}\mathbf{1}} \tag{31}$$

$$\hat{\sigma}_d^2 = \frac{\left(\mathbf{d}-\mathbf{1}\mu_d\right)'\boldsymbol{\Psi}_d\left(\mathbf{X}_e, \mathbf{X}_e\right)^{-1}\left(\mathbf{d}-\mathbf{1}\mu_d\right)}{n_e} \tag{32}$$

Taking the partial derivative of the likelihood respect to the hyper-parameter $\boldsymbol{\theta}_d$ is analytically tedious. As in Section 2.5, a GA is used to maximize the likelihood against $\boldsymbol{\theta}_d$. The optimal estimation obtained for this last hyper-parameter is denoted as $\hat{\boldsymbol{\theta}}_d$.

The multi-fidelity surrogate prediction and mean squared error at a point $\mathbf{x}$ can be assessed once the Gaussian process hyper-parameters are tuned. Forrester, Sóbester, and Keane [64] provide a detailed derivation via augmented training data-set.

$$\hat{y}_e\left(\mathbf{x}\right) = \hat{\mu} + \mathbf{c}^T\mathbf{C}^{-1}\left(\mathbf{y}-\hat{\mu}\mathbf{1}\right) \tag{33}$$

$$\hat{s_e}^2\left(\mathbf{x}\right) = \hat{\sigma}_c^2 + \hat{\sigma}_d^2 - \mathbf{c}^T\mathbf{C}^{-1}\mathbf{c} \tag{34}$$

where $\hat{\mu} = \dfrac{\mathbf{1}'\mathbf{C}^{-1}\mathbf{y}}{\mathbf{1}'\mathbf{C}^{-1}\mathbf{1}}$ and

$$\mathbf{c} = \begin{pmatrix} \hat{\sigma}_c^2\psi_c\left(\mathbf{X}_c, \mathbf{x}\right) \\ \hat{\sigma}_c^2\psi_c\left(\mathbf{X}_e, \mathbf{x}\right) + \hat{\sigma}_d^2\psi_d\left(\mathbf{X}_e, \mathbf{x}\right) \end{pmatrix}$$

2.6.1 Multi-Fidelity Regression

As explained in Subsection 2.5.1, the multi-fidelity interpolation method fails when dealing with noisy data from computer simulations, such as CSD and CFD codes [64].

The regressive model is achieved by adding a regression constant for the expensive model, λ_e, to the diagonal of the high-fidelity block that contains the correlation matrix $\mathbf{\Psi}_d(\mathbf{X}_e, \mathbf{X}_e)$. Forrester et al. [64] propose also to add a regression constant, λ_c, to the diagonal of the low-fidelity block that contains the correlation matrix $\mathbf{\Psi}_c(\mathbf{X}_c, \mathbf{X}_c)$. However, in the present work sparse sampling is expected on the low-fidelity model, so there is no need of the regression constant λ_c. For this case the covariance matrix takes the form

$$cov\left[Y(\mathbf{X}), Y(\mathbf{X})\right] = \begin{pmatrix} \sigma_c^2 \mathbf{\Psi}_c(\mathbf{X}_c, \mathbf{X}_c) & \sigma_c^2 \mathbf{\Psi}_c(\mathbf{X}_e, \mathbf{X}_c) \\ \sigma_c^2 \mathbf{\Psi}_c(\mathbf{X}_e, \mathbf{X}_c) & \sigma_c^2 \mathbf{\Psi}_c(\mathbf{X}_e, \mathbf{X}_e) + \sigma_d^2 \left(\mathbf{\Psi}_d(\mathbf{X}_e, \mathbf{X}_e) + \lambda_e \mathbf{I}_{(n_e, n_e)}\right) \end{pmatrix} \tag{35}$$

The regression constant, λ_e, resolves the problem of approximating noisy functions. Notice that the covariance matrix could be written as $\mathbf{C} + \boldsymbol{\lambda}$. Again, Gaussian radial basis functions are employed.

The unknown hyper-parameters of the multi-fidelity approach are not only the same ones as for the interpolating case, $(\mu_c, \sigma_c, \boldsymbol{\theta}_c, \mu_d, \sigma_d, \boldsymbol{\theta}_d)$, but also λ_e. These hyper-parameters are again obtained from the MLE of the observed data [66]. λ_e adds one extra variable to the MLE process; it increases the tuning computational cost.

MLE for μ_c, and σ_c can be assessed with Equations 13, and 14, respectively. MLE for $\boldsymbol{\theta}_c$ are obtained with the same numerical process as the one explained in Section 2.5. In such manner the low-fidelity surrogate, $\hat{y}_c$, can be built.

The estimations for μ_d and for σ_d are assessed similarly as in Section 2.6. The assessment of the optimal hyper-parameters $\hat{\boldsymbol{\theta}}_d$ and $\hat{\lambda}_e$ is done by a GA since the

process is analytically intractable.

The resulting predictor, $\hat{y}_{e,r}$, is given by

$$\hat{y}_{e,r}(\mathbf{x}) = \hat{\mu_r} + \mathbf{c}^T \left(\mathbf{C} + \hat{\boldsymbol{\lambda}}\right)^{-1} (\mathbf{y} - \hat{\mu}\mathbf{1}) \tag{36}$$

where

$$\hat{\mu}_r = \frac{\mathbf{1}' \left(\mathbf{C} + \hat{\boldsymbol{\lambda}}\right)^{-1} \mathbf{y}}{\mathbf{1}' \left(\mathbf{C} + \hat{\boldsymbol{\lambda}}\right)^{-1} \mathbf{1}}$$

$\mathbf{1}$ is a $n_e + n_c$-dimensional vector of ones. Notice that again the optimal hyper-parameter values are denoted by a hat, $\hat{}$.

c is the correlation vector between the point $\mathbf{x}$ where the prediction is pursued and the observed points $\mathbf{X}_c$ and $\mathbf{X}_e$

$$\mathbf{c} = \begin{pmatrix} \hat{\sigma}_c^2 \psi_c(\mathbf{X}_c, \mathbf{x}) \\ \hat{\sigma}_c^2 \psi_c(\mathbf{X}_e, \mathbf{x}) + \hat{\sigma}_d^2 \psi_d(\mathbf{X}_e, \mathbf{x}) \end{pmatrix}$$

where the $\psi[,]$ is given by Gaussian radial basis functions, see Equation 9.

The multi-fidelity approach permits not only to predict the function value in new sample points, but also to estimate the uncertainty in the prediction $\hat{s}_{e,r}^2(\mathbf{x})$. In the regressive case it is given by Equation 37

$$\hat{s}_{e,r}^2(\mathbf{x}) = \hat{\sigma}_{c,r}^2 + \hat{\sigma}_{d,r}^2 \left(1 + \hat{\lambda}_e\right) - \mathbf{c}^T \left(\mathbf{C} + \hat{\boldsymbol{\lambda}}\right)^{-1} \mathbf{c} \tag{37}$$

where

$$\hat{\sigma}_{c,r}^2 = \frac{(\mathbf{y}_c - \mathbf{1}\hat{\mu_{c,r}})' \boldsymbol{\Psi}_c(\mathbf{X}_c, \mathbf{X}_c)^{-1} (\mathbf{y}_c - \mathbf{1}\hat{\mu_{c,r}})}{n_c}$$

$$\hat{\sigma}_{d,r}^2 = \frac{(\mathbf{d} - \mathbf{1}\hat{\mu_{d,r}})' \left(\boldsymbol{\Psi}_d(\mathbf{X}_e, \mathbf{X}_e) + \hat{\lambda}_e \mathbf{I}_{(n_e,n_e)}\right)^{-1} (\mathbf{d} - \mathbf{1}\hat{\mu_{d,r}})}{n_e}$$

$$\hat{\mu}_{c,r} = \frac{\mathbf{1}'\boldsymbol{\Psi}_c(\mathbf{X}_c, \mathbf{X}_c)^{-1} \mathbf{y}_c}{\mathbf{1}'\boldsymbol{\Psi}_c(\mathbf{X}_c, \mathbf{X}_c)^{-1} \mathbf{1}}$$

$$\hat{\mu}_{d,r} = \frac{\mathbf{1}' \left(\mathbf{\Psi}_d \left(\mathbf{X}_e, \mathbf{X}_e \right) + \hat{\lambda_e} \mathbf{I}_{(n_e, n_e)} \right)^{-1} \mathbf{d}}{\mathbf{1}' \left(\mathbf{\Psi}_d \left(\mathbf{X}_e, \mathbf{X}_e \right) + \hat{\lambda_e} \mathbf{I}_{(n_e, n_e)} \right)^{-1} \mathbf{1}}$$

It is worth mentioning that the model error for the regressive case $\hat{s}_{e,r}$ is not zero at the observed points because the value of the regression constant $\hat{\lambda}_e$ differs from zero. It can produce a non-zero value of the ExI at the observed points, which can lead to optimizations trapped at sample points as explained in the Subsection 2.5.2.

2.6.2 Multi-Fidelity Re-interpolation

The multi-fidelity regression mean squared error, Equation 37, is not zero at the observed points because the value of $\hat{\lambda}_e$ is different from zero. It may result in maximum ExI at design points previously sampled, which would stall the EGO algorithm on deterministic experiments.

Forrester et al. [64] redefine the mean squared error when regressive multi-fidelity approaches are used on deterministic experiments. They build an interpolation of the regressive multi-fidelity surrogate through the values predicted by the regression meta-model at the observed points. It is called re-interpolation. The re-interpolating predictor is the same as the regressive one, i.e., $\hat{y}_{e,ri} = \hat{y}_{e,r} \forall \mathbf{x}$ and $\hat{\mu}_{ri} = \hat{\mu}_r$.

The re-interpolation mean squared error, $\hat{s}^2_{e,ri}$, that Forrester et al. [64] derive is given by

$$\hat{s}^2_{e,ri} \left(\mathbf{x} \right) = \hat{\sigma}^2_{c,ri} + \hat{\sigma}^2_{d,ri} - \mathbf{c}^T \mathbf{C}^{-1} \mathbf{c} \tag{38}$$

where $\hat{\sigma}_{c,ri} = \hat{\sigma}_{c,r}$ because $\lambda_c = 0$ and

$$\hat{\sigma}_{d,ri} = \left(\mathbf{d}_{ri} - \mathbf{1} \hat{\mu}_{d,r} \right) \mathbf{\Psi}_d \left(\mathbf{X}_e, \mathbf{X}_e \right)^{-1} \left(\mathbf{d}_{ri} - \mathbf{1} \hat{\mu}_{d,r} \right) / n_e \tag{39}$$

$$\mathbf{d}_{ri} = \mathbf{1} \hat{\mu_r} + \left\{ \mathbf{c} \left(\mathbf{x_e^{(1)}} \right), ..., \mathbf{c} \left(\mathbf{x_e^{(n_e)}} \right) \right\}^T \left(\mathbf{C} + \hat{\boldsymbol{\lambda}} \right)^{-1} \left(\mathbf{y} - \mathbf{1} \hat{\mu_r} \right) - \hat{\mathbf{y}}_c \left(\boldsymbol{X}_e \right) \tag{40}$$

It is important to realize that in Equation 38 $\mathbf{c}$ and $\mathbf{C}$ depend on $\hat{\sigma}_{c,ri}$ and $\hat{\sigma}_{d,ri}$.

Equation 38 guarantees that the re-interpolating mean squared error $\hat{s}_{e,ri}$ is zero at the sample locations unlike what happens for the regression case $\hat{s}_r$. Thus, the ExIs (see Equation 46) at the observed points are zero in the re-interpolating case, and then, the EGO process does not stall.

2.7 Adaptive Sampling

In Section 2.5 it is pointed out the engineers' desire of using computationally burdensome objective functions in design. A possible solution to these impractical simulations is the construction of simpler and cheaper approximation models of the objective function called surrogate models.

Suppose that due to long function evaluation times, there is a limit on the number of function calls that can be done; where should one sample the intense function to make the most out of the limited function call budget? A possible option is to choose sample points iteratively, in places where the information gained in previous steps is maximized according to a specific criterion. The reason to select this iterative process is that the function is unknown.

No proactive strategy can be relied on when the intense function is unknown. Therefore, traditional DoE space filling techniques do not provide efficient outcomes for limited function call budgets. When a surrogate is available, the only available data are the current function values on the points already sampled and the surrogate prediction built out of them. Thus, it is crucial to use a reactive strategy that is based on previously observed data (as opposed to the blind strategies of conventional DoE) and therefore cleverly samples the function in new designs. This reactive strategy is carried out by an infill criterion. This reactive procedure is usually called adaptive, on-line, or infill sampling.

The infill criterion of an adaptive sampling technique is based on the following

two aims or a weighted combination of them: 1) accurate optimal value or 2) enhance the global accuracy of the model. The first aim is called exploitation, whereas the second is called exploration. Exploitation focuses on the good performing areas using the surrogate prediction, whereas exploration focuses on areas with high uncertainty where there exists a lack of sampling. Therefore, Bayesian surrogate models, such as GRFMs, are brought to provide a predictive distribution to the adaptive sampling algorithm.

When a surrogate is available, new data points[3] are evaluated with the purpose of updating the surrogate and gaining information about the unknown objective. It is done by the optimization of the infill criterion according to the desired balance of the two previous aims. An infill criterion is a function that measures how interesting a design point is. While offline DoE sampling methods use the same sample density in regions of both low and high performance, adaptive sampling distributes the samples according to exploitation, exploration or a combination of both to make the most out of the limited objective function call budget.

The simplest infill criteria fully focus on either of the two possible goals: exploitation or exploration. Other more advanced criteria combine both. Infill criterion examples are statistical lower bound, probability of improvement, expected improvement, goal seeking, and conditional lower bound. Each of these methods has a different balance between exploitation and exploration. Since the present research is interested in early design phases, a good balance between the aims is pursued. The expected improvement (ExI) is a well-known infill criterion that can effectively solve this trade-off; it has been popularized by Jones [99].

ExI is chosen in this work to drive the adaptive sampling algorithm. This criterion has been extensively employed in conceptual design. Also, there exists a wide literature about it that makes it a really attractive criterion to extend adaptive sampling

[3]also known as infill or update points.

techniques to domains with categorical and discrete-quantitative design variables. Subsection 2.7.1 explains ExI in detail.

The adaptive sampling process is iterative. First, a warm-up sampling plan is needed to first initialize the Bayesian surrogate. Normally space filling techniques serve for this purpose. Then, the first infill point (or update) is obtained from the infill criterion optimization, and the surrogate model is updated with the first infill point. Again, the infill criterion searches the new surrogate for a new infill point. This iterative process is continued until a specified convergence criterion is satisfied.

2.7.1 Towards the Expected Improvement Infill Criterion

The expected improvement has been previously branded as an adaptive sampling criterion. It uses the predictive distribution of the probabilistic meta-model to find a new point to sample according to a balance between exploration and exploitation.

The idea of ExI was first introduced by Mockus et al. [136]. ExI compares the current best sample point, y_{min}, with the surrogate prediction over the design space. Note that objective minimization is supposed in the next derivation without loss of generality.

The Bayesian surrogate brands the function's value, $y(\mathbf{x})$, as a random variable, $\mathbf{Y}(\mathbf{x})$. The probability distribution of $\mathbf{Y}(\mathbf{x})$ could be seen in Figure 6 (even though the function $y(\mathbf{x})$ is deterministic, the surrogate output is a probability distribution). If a Gaussian surrogate is used, $\mathbf{Y}(\mathbf{x})$ is a normally distributed random variable with a prediction $\hat{y}(\mathbf{x})$ and a mean squared error $\hat{s}(\mathbf{x})$ around the prediction. The improvement with respect to the current best sample point of a design concept is a random variable which is defined as

$$I(\mathbf{x}) = max(y_{min} - Y(\mathbf{x}), 0) \tag{41}$$

If Kriging is the chosen Bayesian surrogate, the probability of the random variable $Y(\mathbf{x})$ is given by a Gaussian distribution $\frac{1}{\hat{s}(\mathbf{x})\sqrt{2\pi}} \exp\left(-\frac{[Y(\mathbf{x})-\hat{y}(\mathbf{x})]^2}{2\hat{s}^2(\mathbf{x})}\right)$. The probability of improvement, being the probability of $I > 0$, is given by

$$P[I(\mathbf{x})] = \frac{1}{\hat{s}(\mathbf{x})\sqrt{2\pi}} \int_{-\infty}^{y_{min}} \exp\left(-\frac{[Y(\mathbf{x})-\hat{y}(\mathbf{x})]^2}{2\hat{s}^2(\mathbf{x})}\right) dY \tag{42}$$

Doing a change of variables from $Y(\mathbf{x})$ to $I(\mathbf{x})$, carrying out the integration, and making use of the error function $erf(.)$, the probability of improvement could be written as

$$P[I(\mathbf{x})] = \frac{1}{2}\left[1 + erf\left(\frac{y_{min}-\hat{y}(\mathbf{x})}{\hat{s}(\mathbf{x})\sqrt{2}}\right)\right] \tag{43}$$

The probability of improvement just quantifies the likelihood of an improvement with respect to the current best sample solution y_{min}. The amount of improvement is not taken into account in the probability of improvement criterion. Nevertheless, it is obvious that points in the bottom tail of the predictive distribution, far below from y_{min}, represent a large amount of improvements, whereas the probability distribution close to y_{min} does not bring large improvements. However, the tail points have low values of probability density. In order to consider these different amounts of improvement according to their probability, a weighted sum could be employed to get the expected value of the improvement; it is called expected improvement. The first moment of area of the improvement is used for this purpose, which is the red area of the probability distribution in Figure 6.

$$E[I(\mathbf{x})] = E[max(y_{min} - Y(\mathbf{x}), 0)]$$

which particularized for the Kriging predictor, $\mathbb{N}(\hat{y}(\mathbf{x}), \hat{s}(\mathbf{x}))$, takes the form of

$$E[I(\mathbf{x})] = \frac{1}{\hat{s}(\mathbf{x})\sqrt{2\pi}} \int_{-\infty}^{y_{min}} (y_{min} - Y(\mathbf{x})) \exp\left(-\frac{[\hat{y}(\mathbf{x})-Y(\mathbf{x})]^2}{2\hat{s}^2(\mathbf{x})}\right) dY(\mathbf{x}) \tag{44}$$

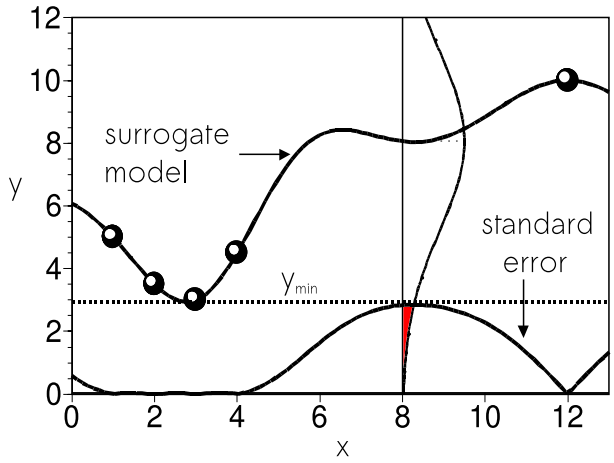


Figure 6: Expected Improvement

If one changes the variables according to $I(\mathbf{x}) = y_{min} - Y(\mathbf{x})$, Equation 44 transforms to

$$E[I(\mathbf{x})] = \frac{1}{\hat{s}(\mathbf{x})\sqrt{2\pi}} \int_0^{\infty} I(\mathbf{x}) \exp\left(-\frac{[\hat{y}(\mathbf{x}) + I(\mathbf{x}) - y_{min}]^2}{2\hat{s}^2(\mathbf{x})}\right) dI(\mathbf{x}) \tag{45}$$

adding and subtracting $\frac{\hat{y}(\mathbf{x}) - y_{min}}{\hat{s}(\mathbf{x})\sqrt{2\pi}} \exp\left(-\frac{[\hat{y}(\mathbf{x}) + I(\mathbf{x}) - y_{min}]^2}{2\hat{s}^2(\mathbf{x})}\right)$ leads to two simpler immediate integrals that can be easily evaluated. So, $E[I(\mathbf{x})]$ can be written as

$$E[I] = \begin{cases} \hat{s}\phi\left(\frac{y_{min} - \hat{y}}{\hat{s}}\right) + (y_{min} - \hat{y})\,\Phi\left(\frac{y_{min} - \hat{y}}{\hat{s}}\right) & \text{if } \hat{s} > 0 \\ 0 & \text{if } \hat{s} = 0 \end{cases} \tag{46}$$

where $\phi(.)$, and $\Phi(.)$ are the standard normal probability and cumulative distribution function, respectively.

There exist two main explanations for a large value of $E[I(\mathbf{x})]$ in regions of the design space. Firstly, the points already sampled hint that the expected value at $\mathbf{x}$ is likely to be smaller than the current best sample point. Secondly, the deviation of the distribution is spread because not many samples have been taken in the design region where $\mathbf{x}$ seats. The first reason supports exploitation, whereas the second does exploration. Hence, $E[I(\mathbf{x})]$ represents a balance between the two phenomena, and the maximum value of $E[I(\mathbf{x})]$ is a solid choice to pick the next sample point. Usual

stopping criteria for the infilling iterative process are: 1) reaching certain number of updates given by the limited function budget or 2) when $E[I(\mathbf{x})]$ is less than a certain percentage of the current best sample point.

Landscapes of the expected improvement possess similar characteristics. Their values are zero at the sample points and are positive between them. The ExI landscapes are multi-modal. They usually feature several flat regions where the ExI values are close to zero because of the high concentration of sample points. These characteristics make the optimization of the expected improvement rather difficult. Specifically, gradient techniques are not normally used due to this multi-modality nature of the ExI function.

Jones [99] points out that $E[I(\mathbf{x})]$ is decreasingly and increasingly monotonic in $\hat{y}$, and $\hat{s}$, respectively. These properties allow setting an $E[I(\mathbf{x})]$ upper bound on the design space. This bound could be employed by a branch and bound (BB) method to guarantee optimality. Herein, the focus is to explore and exploit the design space; therefore, guarantee of optimality and high accuracy in the successive new sample points are not strictly necessary. Hence, the $E[I(\mathbf{x})]$ maximization is done by well-known methods like GA where a trade between precision and speed can be easily set by the user.

2.8 Mixed-Integer Optimization

Many engineering, scientific, and management applications involve both discrete and continuous decisions. Also linear and/or nonlinear relationships define the feasibility and optimality of the solutions. These kind of problems are solved with mixed-integer optimization (MIO) techniques. Possible application fields are management science, process industry, finance, optical filter design [167], portfolio selection, chemical engineering [56], machine learning [125], optimal design of gas and water distribution

networks, automobile engineering, aircraft design, and integrated circuit manufacturing, among others.

Examples of discrete-quantitative variables in mechanical engineering are number of items of one type, size of standard manufactured items, ... In aerospace engineering these integer variables can be the number of engines, helicopter blades, stages in compressor and turbine, and number of composite layers and their orientation, among others.

The general MIO problem could be laid down as follows

$$\min \quad f\left(\mathbf{x}_{cont}, \mathbf{x}_{int}\right) \tag{47}$$

subjected to:

$$\begin{aligned}
& g_j\left(\mathbf{x}_{cont}, \mathbf{x}_{int}\right) \leq 0, \qquad j = 1, ..r_i \\
& h_j\left(\mathbf{x}_{cont}, \mathbf{x}_{int}\right) = 0 \qquad j = r_i + 1, ..r_i + r_e \\
& \mathbf{x}_{cont}^{L} \leq \mathbf{x}_{cont} \leq \mathbf{x}_{cont}^{U} \\
& \mathbf{x}_{int}^{L} \leq \mathbf{x}_{int} \leq \mathbf{x}_{int}^{U} \\
& \mathbf{x}_{int} \in \mathbb{Z} \\
& \mathbf{x}_{cont} = [x_{cont,1}, x_{cont,2}, ...x_{cont,n1}] \\
& \mathbf{x}_{int} = [x_{int,1}, x_{int,2}, ...x_{int,n2}]
\end{aligned} \tag{48}$$

where $\mathbf{x}_{cont}$ and $\mathbf{x}_{int}$ represent the set of continuous and integer variables, respectively; g_j and h_j the inequality and equality constraints; $\mathbf{x}_{cont}^{L}$ and $\mathbf{x}_{cont}^{U}$ the lower and upper continuous bounds; and $\mathbf{x}_{int}^{L}$ and $\mathbf{x}_{int}^{U}$ the lower and upper integer bounds. The number of continuous and integer decisions are n_1 and n_2, respectively, whereas the number of inequalities and equalities are r_i and r_e.

MIO algorithms can be classified into two main ones: mixed-integer programming (MIP), and stochastic searches. A MIP can deal with linear problems (MILP) or nonlinear ones (MINLP). MIP can also deal with constraints by which some or all

decision variables are restricted to certain values. A popular MIP technique, called branch and bound (BB), were developed only for linear problems MILP. Later, new techniques were implemented to handle nonlinear convex problems. These techniques are outer approximation (OA), generalized bender decomposition (GBD), and extended cutting plane technique (ECP). Also, extensions of BB were developed to tackle not only convex but also non-convex problems. The convex extensions guarantee local optima convergence. However, for non-convex problems these programming techniques may provide the wrong global optimum.

MINLP are hard to solve because they have all the difficulties of both the integer programming (IP) and non-convex programming (NLP). IP problems are of combinatorial nature that provokes a great increase of the possible integer combinations as the dimension and density of the integer space, $\mathbf{y}$, grows. For example, in an IP problem with k dimensions with n_d points each dimension, the possible combinations are n_d^k. The idea is to arrange the problem such that one can get information that can be used to exclude large numbers of solutions from further consideration. This is the philosophy of the Balas method for IP problems [11], and branch and bound methods for MIP problems [36].

For non-convex and highly nonlinear problems, stochastic searches have been proposed in the literature with more successful results than the ones of programming techniques. Among the stochastic searches, one can find: particle swarm optimization, line-up competition algorithm, simulated annealing techniques, and genetic algorithms.

In this section, a brief review of the mixed-integer convex programming techniques is presented (it includes the techniques for linear problems). It is followed by the discussion of the extensions of the convex programming to non-convex scenarios. Finally, an introduction to the MIO stochastic algorithms is presented.

2.8.1 Convex Programming

The following approaches are explained: branch and bound, outer approximation, generalized benders decomposition, and extended cutting plane. The target problem is the one described by Equations 47 and 48 when $f(\cdot)$ and $g(\cdot)$ are convex. The approaches that are described in this Subsection only guarantee global optimality under convexity conditions.

2.8.1.1 Branch and Bound Algorithms

It was introduced by Balas [11] for binary linear programming problems with k variables. It sets all the variable to zero for a maximization problem; then, successively assigns the value 1 to certain variables, in such a manner that a smaller subset of all possible combinations, 2^k, is sufficient to obtain an optimal solution or conclude that no feasible solution exists. In order to do it, branching and bounding techniques are employed. Branching refers to the fact that throughout the process a tree is branched to assign possible values (0 and 1 for the binary case) to the variables. The bounding technique groups a subset of the possible decision variable combinations, assign a lower bound to this subset, and assess whether the set is impossible or not. The lower bound might permit comparing subsets to rule out some of them, reducing the number of design combination evaluations, therefore it speeds-up the process. Also, impossible sets are ruled out because none of the containing designs satisfies the restrictions, reducing again the number of evaluations.

Dakin [36] developed a branch and bound method for MILP. It ignores the integer restrictions and solves the problem as if all the variables were continuous (using LP). This relaxation provides a good lower bound for maximization problems. One of the integer variables, $x_{int,j}$, is chosen for branching the current node into two child nodes. The parental node relaxation problem has directly been solved and the value of the variable j for the optimal solution is between l and $l+1$. These values are used to

impose bounds to each of the new child node problems:

Child Problem 1. Parental node constraints plus a new bound $x_{int,j} \leq l$.

Child Problem 2. Parental node constraints plus a new bound $x_{int,j} \geq l + 1$.

The same process is repeated until the linear relaxation solution has integer values in all the integer variables and the solution is feasible, then non-improvement can be achieved by the expansion of the node. This reminds of the bound obtained by Balas for the node descendant subset. Again, comparison between feasible nodes leads to the discard of nodes and their descendant subsets. It is noticeable that solving MILP problems generally needs quite longer than its continuous counterpart, LP problems, due to the mentioned combinatorial nature of the MILP problems.

Gupta and Ravidran [76] study the computational feasibility of branch and bound methods in solving convex nonlinear programming problems. Quesada and Grossmann [148] apply the BB method to convex problems. The idea is to solve a NLP subproblem by linearizations similar as in OA method, then a BB algorithm solves the mentioned NLP subproblem at the nodes where integer feasible solutions are found. Later, Borchers and Mitchell [20] describe a branch and bound method for mixed-integer nonlinear programs with convex objective functions and constraints.

The BB method is usually attractive when NLP problems are inexpensive and only few of them are to be solved. This conditions normally occur when a few discrete variables are present in the problem.

2.8.1.2 Outer Approximation and Generalized Benders Decomposition

The MINLP problems are solved with successive solutions of related MIP problems. The methods divide the MINLP problem into a subproblem that has fixed discrete variables, converting it into a NLP problem and a linear master problem MILP, as Duran and Grossmann [47] demonstrate in their OA algorithm. Another OA solver

is presented by Fletcher and Leyffer [59]. Geoffrion [71] presents Bender's approach and generalizes it to a wider class of programming techniques in which subproblems do not need to be linear.

The difference between OA and GBD resides in the definition of the MILP master problem. The GBD master problem only considers active inequalities and the continuous bounds are disregarded, whereas OA methods use tangential hyperplanes (or linearizations) to set the feasible space in the master problems. These feasible space linearizations provide a successively shrinking space for the master problem that represents convergence to the solution.

2.8.1.3 Extended Cutting Plane

Unlike other methods previously explained in this Section (BB, OA, and GBD), ECP does not make use of NLP subproblems. It relies on an iterative solution of the master linear problem, which is a MILP problem, by successively adding linearizations; see Westerlund and Pettersson [197]. The chosen linearization is the one of the constraint that is the most disrespected at the previous solution point in the iteration process. Since the functions $f(\mathbf{x}, \mathbf{y})$ and $g(\mathbf{x}, \mathbf{y})$ are convex, the nonlinear feasible region defined by $g(\mathbf{x}, \mathbf{y})$ is outer-approximated. So, the addition of hyperplanes throughout the iterative process provides a non-decreasing bound for the minimization problems.

2.8.2 Non-Convex Programming

Some programming algorithms are implemented to solve non-convex and multi-modal MINLP problems. These global deterministic optimization methods for mixed-integer non-convex problems normally rely on a branch and bound procedure. The manner to perform the branching on the variables is the main difference across these methods. Convex relaxation is successfully employed to obtain solutions as demonstrated by Tawarmalani and Sahidinis [182], and Floudas [60]. Smith and Pantelides [173] reformulate the problem and apply convex relaxation. Then, they use an spatial branch

and bound search. Kesevan and Barton [111] propose their branch and cut method.

Other methods branch a tree on both discrete and continuous variables. Others perform the BB on the continuous variables and solve each node problem, a MINLP, by any of the methods explained previously. Other ones branch on the discrete variables and then use BB on nodes where feasibility on discrete variables is found. The existing methods are not explained in details because they are not used in this work. These deterministic non-convex programming methods are generally outperformed by stochastic or random mixed-integer searches.

2.8.3 Stochastic or Random Search

In the last two decades many random search algorithms are developed for MINLP. Among them, one can find the following ones: particle swarm optimization by Yiqing et al. [206] , line-up competition algorithm by Yan et al. [203], differential evolution by Regulwar [152], simulated annealing techniques, evolution strategies (ES) by Li et al. [125], and GA by Yokota [207] and Deep et al. [44]. Stochastic methods are more powerful than the methods introduces in Subsection 2.8.2. As opposed to programming methods, stochastic methods do not rely on gradient calculations after continuous relaxation of the integer set. Gradients are not meaningful in the discrete world. Also, the gradient information is not always useful towards global optimum in multi-modal functions. There is a enormous literature available for stochastic processes, but only GAs are employed in this research.

Genetic algorithms are stochastic algorithms that copy the principles of genetics and natural selection. Holland [88] introduced the concept of GA as a search method; De-Jong [39] was the first one in using it to solve an optimization problem. The set of candidate solutions or individuals is called population. The first population is randomly chosen and it evolves toward better solutions. In each generation, individuals are assigned a fitness and rank according to their fitness value. A new population is

formed from the current population based on the best ranked individuals and new individuals obtained by crossover and mutation. The same process is repeated for new generations until satisfactory fitness levels are achieved or a maximum number of generations is reached.

The random search philosophy of MINLP methods is similar to the one of their continuous counterparts. The fitness assignment and ranking techniques are usually the same. However, the main differences for the GAs lie on the handling of integer restriction, population mutation, and population crossover functions.

Generation techniques to initialize the first population of a mixed-integer problem are usually performed by using discrete techniques, such as full factorial techniques, or just rounding-off available continuous generation techniques, such as random Latin hypercube and space-filling Latin hypercubes[113]. Forrester, Sóbester and Keane [65] explain in detail these techniques.

Regarding mutation and crossover techniques the simplest way to deal with integer variables is to utilize continuous mutation and crossover functions pretending that the integer variables are continuous. Then, simple truncation can convert the crossover and mutation children back into the discrete world. Nevertheless, there exists a clear drawback for this truncation: the step-size could be diminished to a so small value that the necessary leap between two consecutive discrete points is not reached with the consequent catastrophic results, as Li [125] and Deep [45] point out. It is because in discrete spaces the smallest distance in l_1-norm between two points is greater than zero as opposed to what happens in continuous spaces. Hence, when the step size is smaller than the smallest distance between discrete values, search stagnation occurs. The situation is worse for categorical variables where a possible continuous relaxation assumes neighboring values that may produce convergence to a local optima instead of a global one.

Several mutation and crossover functions have been proposed in the literature for

the integer case. The principle of maximum entropy expresses maximum uncertainty with respect to everything but the given information, as Jaynes [93] shows. Rudolph [159] proposes a mutation algorithm for unbounded integer search spaces by using this principle of maximum entropy: the resulting mutation distribution obtained by Rudolph is a geometric distribution. Then, he applies the resulting distribution to GA and ES. The latter distribution is extensively employed when dealing with integer design variables. For the non-numeric variables, Li et al. [127] follow the same maximum entropy principle to get to the conclusion that a uniform probability mutation distribution is the most convenient. They apply this uniform mutative distribution in an ES.

Regarding crossover functions for integer and categorical variables, less instances are found in the literature. Laplace crossover is proposed by Deep and Thakur [45], although later truncation is needed to get integer values for the children. Some crossover techniques from continuous GA can be employed in the discrete case. The main crossover techniques could be divided into two main groups:

1) Scattered Crossover. Genes are chosen with equal probability from the parents. Thus, it combines the parental genes to form the children.

2) Intermediate or Blending Crossover. An arithmetic or weighted sum of the parental chromosomes is used to produce the children.

Note that crossover techniques can be based on not only two parents but also more than two parents.

The scattered crossover can be easily applied to all kinds of discrete variables. The intermediate technique can also be applied to discrete-quantitative variables, however intermediate crossover is not useful for non-numeric variables due their lack of order and intermediate values.

The main limitation of the GA for mixed-integer optimization is that none of

the freely available software provides open access to the generation, mutation, and crossover functions. For instance, a mixed-integer GA is available in MATLAB® 2011b, but the access to the MATLAB® files is restricted by encryption, therefore variations of the code in the mutation, generation, and crossover functions are not possible.

2.9 Multi-Objective Expected Improvement

Optimization techniques search for the best solution among all possible. However, many engineering problems require the simultaneous optimization of two or more conflicting objectives: these scenarios are called multi-objective optimization problems. Satisfactory trade-offs between these conflicting objectives have to be reached in the final design.

Typical fields where multi-objective optimization problems can be found are finance, process design, aircraft and automobile design, robust design, or wherever trade-offs between two or more conflicting objectives exist. Within the aerospace industry, typical objectives are lightweight, high-performance, low-cost, robustness,... Common instances of multi-objective optimization are the maximization of the aircraft performance while minimizing engine fuel consumption; the minimization of the weight while maximizing the strength of a particular aircraft component; and robust design of a aircraft wing where the six-sigma objective function is included.

In mathematical terms, the general multi-objective optimization problem can be written as follows:

$$\min_{\mathbf{x}_{cont},\mathbf{x}_{quant},\mathbf{x}_{nom}} \left[f_1\left(\mathbf{x}_{cont},\mathbf{x}_{quant},\mathbf{x}_{nom}\right),...,f_m\left(\mathbf{x}_{cont},\mathbf{x}_{quant},\mathbf{x}_{nom}\right)\right]^T \tag{49}$$

subjected to,

$$\mathbf{g}\left(\mathbf{x}_{cont}, \mathbf{x}_{quant}, \mathbf{x}_{nom}\right) \leq 0 \tag{50}$$

$$\mathbf{h}\left(\mathbf{x}_{cont}, \mathbf{x}_{quant}, \mathbf{x}_{nom}\right) = 0 \tag{51}$$

$$\mathbf{x}_{cont}^{L} \leq \mathbf{x}_{cont} \leq \mathbf{x}_{cont}^{U} \tag{52}$$

$$\mathbf{x}_{quant}^{L} \leq \mathbf{x}_{quant} \leq \mathbf{x}_{quant}^{U} \tag{53}$$

$$\mathbf{x}_{nom}^{L} \leq \mathbf{x}_{nom} \leq \mathbf{x}_{nom}^{U} \tag{54}$$

$$\mathbf{x}_{quant} \in \mathbb{D}_{quant} \tag{55}$$

$$\mathbf{x}_{nom} \in \mathbb{D}_{nom} \tag{56}$$

where $\mathbf{x}_{cont}$, $\mathbf{x}_{quant}$, and $\mathbf{x}_{nom}$ represent the set of continuous, discrete-quantitative, and nominal variables, respectively; $[f_1, ..f_m]$ the m-component objective vector; $\mathbf{g}$ and $\mathbf{h}$ the inequality and equality constraint vector; $\mathbf{x}_{cont}^{L}$ and $\mathbf{x}_{cont}^{U}$ the lower and upper continuous bound vector; $\mathbf{x}_{quant}^{L}$ and $\mathbf{x}_{quant}^{U}$ the lower and upper discrete-quantitative bound vector; and $\mathbf{x}_{nom}^{L}$ and $\mathbf{x}_{nom}^{U}$ the lower and upper nominal bound vector. Finally, $\mathbb{D}_{quant}$ and $\mathbb{D}_{nom}$ represent the discrete subsets at which the discrete-quantitative and nominal variables are restricted, respectively.

When dealing with more than one objective that are in conflict with each other, it is usually not possible to find a feasible solution which is optimal for all the conflicting objectives. A subset of solutions must be considered optimal in terms of trade-offs. It means that it is not possible to improve one of the objectives without degrading one or more of the others within this optimal family. No point of this optimal family of solutions could be said to be better than other one, unless a given importance weight of the objectives is known a priori. This optimal family is called Pareto optimal set in honor to the economist Pareto [144] who introduced this concept within the field of welfare economics. Later, it has been extended to engineering and social sciences.

Before defining formally a Pareto set, the concept of domination is presented. Let us assume that two points $\mathbf{y}$ and $\mathbf{z}$ are in the solution space, i.e., $\mathbf{y}, \mathbf{z} \in \mathbb{R}^m$. The

point $\mathbf{y}$ strictly dominates $\mathbf{z}$ if and only if $\forall i \in 1, .., m : y_i \leq z_i$ and $\mathbf{y} \neq \mathbf{z}$. A usual shorthand notation of strict domination is $\mathbf{y} \prec \mathbf{z}$. If a point $\mathbf{y}$ strictly dominates or equals other point $\mathbf{z}$, then the following notation is used $\mathbf{y} \preceq \mathbf{z}$. A finite set of points $P \subset \mathbb{R}^m$ dominates or equal a point $\mathbf{z}$, in mathematical terms $P \preceq \mathbf{z}$, if and only if $\exists \mathbf{y} \in P : \mathbf{y} \preceq \mathbf{z}$. A set of points P is a non-dominated set if and only if all the points in the set P are mutually non-dominant.

A feasible point $\mathbf{y}$ is Pareto optimal or Pareto efficient if and only if there is no other feasible point that dominates it

$$\mathbf{y} \text{ is a Pareto-optimal } \nexists \mathbf{z} \in S : \mathbf{z} \preceq \mathbf{y} \tag{57}$$

where S is the feasible objective space. In other words, for a Pareto optimal point there exists no feasible solution point which would decrease some objectives without provoking an increase in at least one of the other objectives. Generally, this concept provides a set of points instead of a single solution. The Pareto front or Pareto set is the set of points that are Pareto efficient, see Equation 57; consequently, it is a non-dominated front. Thus, a Pareto front includes designs which are so optimized that, in order to improve one goal of any of its members, its performance in at least one of the other goals diminishes. A Pareto front does not assume any specific relative weighting between opposing goals. However, once the Pareto front is obtained engineers can make a better compromise decision by weighting the conflicting goals.

Many multi-objective algorithms are developed to assess Pareto fronts, but in the research herein, interest is in obtaining Pareto fronts when a precise evaluation of the design is not available, but a prediction with a measure of the uncertainty is. The prediction takes the form of a duple made of m-dimensional vectors: the mean $\boldsymbol{\mu}$ and the standard deviation $\boldsymbol{\sigma}$ of the prediction vector.

Typical quality indicators for finding Pareto fronts are the epsilon and hyper-volume indicators. Deb, Mohan, and Mishra [42] select the next generation based on

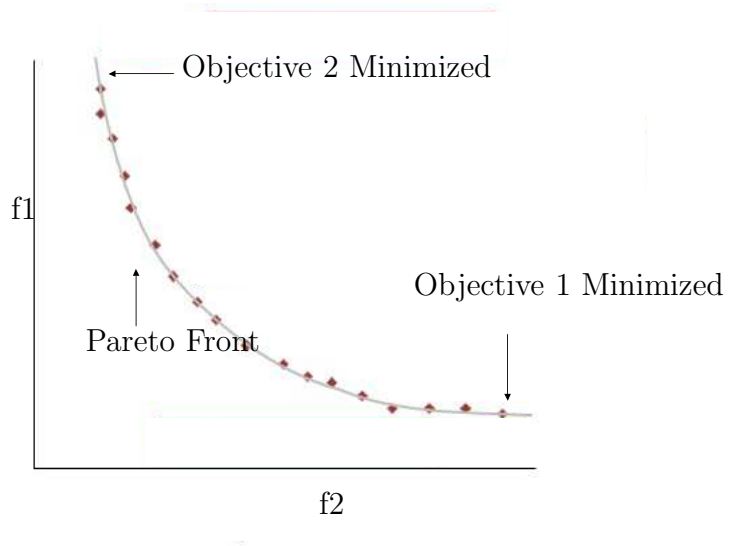


Figure 7: Pareto Front Example

the concept of epsilon-dominance. Friedrich, Horoba et al. [67] compare the efficiency of the hyper-volume indicator with that of the epsilon based EMOA.

Emmerich [55] proposes the quality indicator hyper-volume, introduced by Zitzler [213] for comparing Pareto outcomes from different EMOA, as an improvement measure for extending the single-objective improvement to the multi-objective space. The difficulty of the extension resides in the fact that the best found solution is not clearly defined in multi-objective problems. However, the increment of dominated hyper-volume is a reduced scalar for the Pareto set with appropriate features as is explained in Subsection 2.9.2.

Within this Bayesian scope, a survey in the literature shows that the two most common criteria for multi-objective adaptive sampling are the statistical improvement criterion proposed by Keane [106] and the hyper-volume based ExI by Emmerich [55].

The multi-objective adaptive sampling criterion used in this research is based on the increment of dominated hyper-volume or Lebesgue measure, i.e., the one proposed by Emmerich [55]. A study done by Wagner et al. [192] shows that this method based on the Lebesgue measure satisfies the proposed necessary conditions, which justifies its use.

In Subsection 2.9.1 a formal definition of the hyper-volume of a set of points is explained. Also, Fleisher's approach, that exactly calculates the Lebesgue measure, is presented. Later, in Subsection 2.9.2 the hyper-volume increment of a Pareto set due to a new point is postulated as the improvement scalar needed to extend the ExI criterion from uni-dimensional objective spaces to multi-dimensional ones.

2.9.1 Hyper-Volume Definition and its Calculation

Let $Vol\,(U)$ be the m-dimensional volume of a space $U \in \mathbb{R}^m$. Then, the Lebesgue measure or hyper-volume of a set of solutions P, $\mathbb{H}\,(P)$, quantifies the size of the goal space which is dominated by P, it could be written as

$$\mathbb{H}\,(P) = Vol\left(\mathbf{y} \in \mathbb{R}^m : P \preceq \mathbf{y} \preceq \mathbf{y}^{ref}\right) \tag{58}$$

where $\mathbf{y}^{ref}$ is a reference point to bound the infinite domain dominated by the set of points P.

If F is the set of non-dominated solutions of P, i.e., the Pareto front of P, then $\mathbb{H}\,(P) = \mathbb{H}\,(F)$

A common and efficient way to calculate the hyper-volume of a set of points P is presented by Fleischer [58]. The algorithm he proposes trims off hypercubes that are dominated by P and adds their hyper-volume to the Lebesgue measure. The original set of points is stored in a list denoted by *List*. In each step, a point of *List* is taken and removed. Then, bound values, $\mathbf{b}$, for this removed point are obtained based on the remaining points in *List*. The contribution of the removed point is added to the partial Lebesgue sum. Later, "'spawned points" are generated with the help of $\mathbf{b}$. A filter compares the remaining points in *List* with the "'spawned points"; then, the non-dominated "'spawned points" are included in *List*. A more detailed discussion is presented in Fleisher's work [58].

2.9.2 Expected Improvement Based on Dominated Hyper-Volume

Similarly to the single-objective case, a statistical distribution $[\boldsymbol{\mu}, \boldsymbol{\sigma}]$ in the solution space for each design space point, $\mathbf{x}$, is given by the GRFM. The main challenge is to generalize the single-objective best solution concept to the multi-objective Pareto front idea. An improvement scalar similar to the one defined in Equation 41 (see Subsection 2.7.1) is searched for in order to generalize from one to multiple objectives. The increment of the dominated hyper-volume of the Pareto front has been proved as an appropriate scalar function for this purpose because of some of its characteristics.

Fleisher [58] proves that the dominated hyper-volume measure of a set is maximum if the set is on the true Pareto front. Another attractive features is its behavior when a new point, $\mathbf{y}_{new}$, is added to the Pareto front F. The hyper-volume of the augmented set $\mathbb{H}(F \cup \mathbf{y}_{new})$ increases if and only if no point in F dominates $\mathbf{y}_{new}$. Hence, the augmented set, $F \cup \mathbf{y}_{new}$, could be branded as a better approximation to the Pareto set than F.

As a drawback, normalization of the objective space is required to assure that equal gains in objectives are equally reflected by the dominated hyper-volume measure. Finding crude bounds for the objective space is not normally hard, so a rude normalization should not be really difficult to perform. However, the bias of the improvement towards some objectives increases when only rough objective bounds are available. Therefore, in these cases some care should be taken of when interpreting the value of the improvement.

The improvement due to a new solution point $\mathbf{y}(\mathbf{x})$ is given by Equation 59

$$I(\mathbf{y}(\mathbf{x})) = \begin{cases} \mathbb{H}(F \cup \mathbf{y}(\mathbf{x})) - \mathbb{H}(F) & \text{if no point of } F \text{ dominates } \mathbf{y}(\mathbf{x}) \\ 0 & \text{otherwise} \end{cases} \tag{59}$$

Once the improvement for multiple-objectives is defined, the expectation of improvement has to be obtained in a similar way as is done in Subsection 2.7.1. The

main difference is that now the integration is on a multi-dimensional space $\mathbb{R}^m$, where m is the dimension of the objective space. The multi-objective $E\left[I\left(\mathbf{x}\right)\right]$ can be written as

$$E\left[I\left(\mathbf{x}\right)\right] = \int_V I\left(\mathbf{y}\right) PDF\left(\mathbf{y}\right) d\mathbf{y} \tag{60}$$

where V represents the solution space. The integral is over the objective space, which is assumed continuous in this research regardless the nature of the design space.

The integrand in Equation 60 is intended to be computed with a multi-objective meta-model $[\boldsymbol{\mu}, \boldsymbol{\sigma}]$. Monte Carlo integration of Equation 60 could be done, but it is time consuming and not accurate. Emmerich [54] proposes a direct computation for Equation 60 for the case of two objectives; it is explained in this Subsection with a notation similar to the one utilized by Emmerich.

The idea to direct integrate Equation 60 is to divide the integration region into boxes, where piecewise integration can be directly done, converting the integral into a sum over the boxes. Let us assume that the current Pareto front consists in q points $\mathbf{y}_1, ..., \mathbf{y}_q$.

A sorted list of all the i-components of the Pareto front points is $b_i^{(1)}, ..., b_i^{(q)}$. The coordinates $\mathbf{b}_i$ for $i = 1, ..., m$ are called the grid coordinates. The coordinates $\mathbf{b}_i$ are augmented with $b_i^{(0)} = -\infty$, $b_i^{(q+1)} = y_i^{ref}$, and $b_i^{(q+2)} = \infty$ for $i = 1, ..., m$.

The grid cells are named $C\left(i_1, i_2, ..., i_m\right)$. Grid cells are defined by their upper and lower bound vectors: $\mathbf{l}\left(i_1, i_2, ..., i_m\right)$ and $\mathbf{u}\left(i_1, i_2, ..., i_m\right)$

$$\mathbf{u}\left(i_1, i_2, ..., i_m\right) = \left(b_1^{(i_1+1)}, b_2^{(i_2+1)}, ..., b_m^{(i_m+1)}\right)^T \qquad \mathbf{l}\left(i_1, i_2, ..., i_m\right) = \left(b_1^{(i_1)}, b_2^{(i_2)}, ..., b_m^{(i_m)}\right)^T$$

Another auxiliary variable required to directly integrate the expected improvement in multi-objective problems is the vector $\mathbf{v}\left(i_1, i_2, ..., i_m\right) \in \mathbb{R}^m$. The jth-component of $\mathbf{v}\left(i_1, i_2, ..., i_m\right)$ is the jth-component of the intersection point between the Pareto

surface and a j coordinate line[4] that passes through the lower bound of the given cell, $\mathbf{l}(i_1, i_2, ..., i_m)$. $\mathbf{v}(i_1, i_2, ..., i_m)$ is called the cell reference point vector.

Cells are classified into active and inactive cells. The former ones are those where the contribution to the integral is different from zero, whereas the latter are those for which the contribution is null. If a cell meets one of the following criteria, the cell is branded as inactive

1) Cell lower bound is dominated or equal to points in the Pareto front F, i.e. $F \preceq \mathbf{l}(i_1, i_2, ..., i_m)$.

2) Cell upper bound does not dominate the reference point, i.e., $\mathbf{u}(i_1, i_2, ..., i_m) \npreceq \mathbf{y}^{ref}$.

Hence, active cells, denoted by $\mathbb{C}^+$, are those whose upper corner dominates the reference point and whose lower bound dominates at least one point of the Pareto front F.

The expected improvement could be written in terms of a sum throughout all the active cells,

$$E[I(\mathbf{x})] = \sum_{C(i_1,i_2,...,i_m)\in\mathbb{C}^+} \delta(i_1, i_2, ..., i_m) \tag{61}$$

where the active cell contribution $\delta(i_1, i_2, ..., i_m)$ is given by

$$\delta(i_1, i_2, ..., i_m) = \int_{\mathbf{y}\in C(i_1,i_2,...,i_m)} I(\mathbf{y})\, PDF(\mathbf{y})\, d\mathbf{y} \tag{62}$$

Once the integration of Equation 60 has been divided, active cell contributions are discussed. For clarity and compactness' sake, cell indexes are omitted. The improvement in a cell, $I(\mathbf{y})$, can be divided into the contribution of the hyper-volume of the non-dominated part in the hypercube $[\mathbf{u}, \mathbf{v}]$, denoted by L^+ (see Figure 8), and

[4]In this report, a j coordinate line refers to a line parallel to the jth coordinate axes

the "boundary contribution", denoted by B^+ (see Figure 8), given by $[\mathbf{y}, \mathbf{v}] - [\mathbf{u}, \mathbf{v}]$ for the two-objective case. For more than two objectives, the "boundary contribution" has a more complex form. Then, in the two-objective case, the $I(\mathbf{y})$ takes the form

$I(\mathbf{y}) = Vol([\mathbf{y}, \mathbf{v}] - [\mathbf{u}, \mathbf{v}] + L^+)$

The reference point for the increment of the dominated hyper-volume of each cell is given by the vector $\mathbf{v}$. The improvement expression can be simplified by realizing that $S^- = [\mathbf{u}, \mathbf{v}] - L^+$, where S^- is the fraction of dominated hyper-volume inside the hypercube $[\mathbf{u}, \mathbf{v}]$, see Figure 8. Then, $I(\mathbf{y})$ can be written as,

$$I(\mathbf{y}) = Vol\left([\mathbf{y}, \mathbf{v}] - S^-\right) \tag{63}$$

Note that correction terms must be added for more than two dimensions.

Therefore, in order to calculate cell contribution to $I(\mathbf{y})$, two volumes should be obtained,

1) $Vol([\mathbf{y}, \mathbf{v}])$ which is $\mathbf{y}$ dependent.

2) $Vol(S^-)$ which is independent for all $\mathbf{y}$ within the same cell.

In order to calculate the volume S^- for the two-objective case, it is enough to obtain the Pareto points that are dominated by the upper point of the cell, i.e., $\forall \mathbf{y} \in F : \mathbf{u} \preceq \mathbf{y}$. For higher dimensions, the assessment of S^- is more difficult.

By plugging 63 into 62 one gets two terms for the cell contribution $\delta(i_1, i_2, ..., i_m) = \delta_1(i_1, i_2, ..., i_m) + \delta_2(i_1, i_2, ..., i_m)$

$$\delta_1(i_1, i_2, ..., i_m) = \int_{\mathbf{y} \in C(i_1, i_2, ..., i_m)} \prod_{i=1}^{m} (v_i - y_i) \cdot PDF(\mathbf{y})\, dy_1 dy_2 \tag{64}$$

$$\delta_2(i_1, i_2, ..., i_m) = Vol\left(S^-\right) \int_{\mathbf{y} \in C(i_1, i_2, ..., i_m)} \prod_{i=1}^{m} PDF(\mathbf{y})\, dy_1 dy_2 \tag{65}$$

Equations 64 and 65 could be analytically integrated as follows

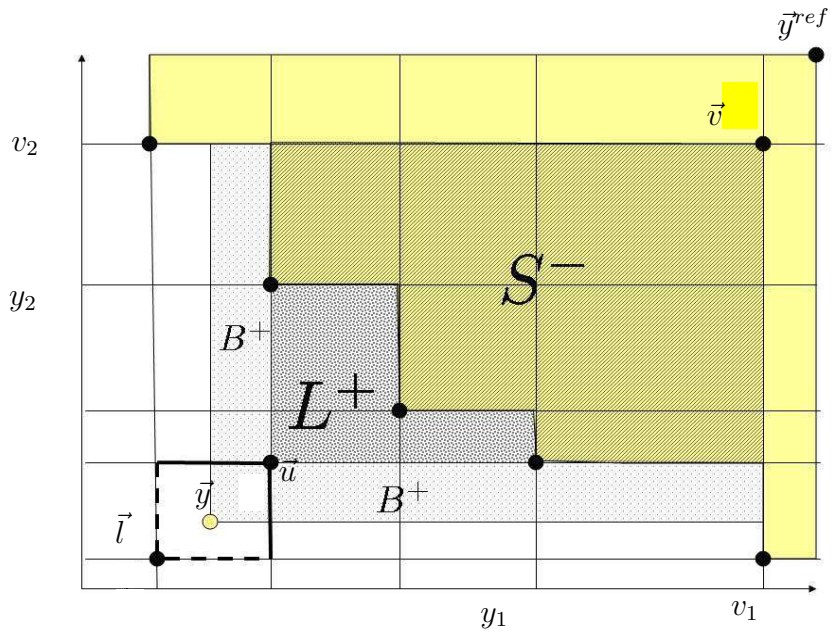


Figure 8: Expected Improvement Integration Area for Two-Objective Case

$$\delta_1(i_1, i_2, ..., i_m) = Vol\left(S^-\right) \prod_{i=1}^{m} \left(\Phi\left(\frac{u_i - \mu_i}{\sigma_i}\right) - \Phi\left(\frac{l_i - \mu_i}{\sigma_i}\right) \right) \tag{66}$$

$$\delta_2(i_1, i_2, ..., i_m) = \prod_{i=1}^{m} \left(\mathbf{\Psi}(v_i, u_i, \mu_i, \sigma_i) - \mathbf{\Psi}(v_i, l_i, \mu_i, \sigma_i) \right) \tag{67}$$

where $\mathbf{\Psi}(a, b, \mu, \sigma) = \sigma\phi\left(\frac{b-\mu}{\sigma}\right) + (a - \mu)\,\Phi\left(\frac{b-\mu}{\sigma}\right)$, and $\phi(.)$ and $\Phi(.)$ are the standard normal probability and cumulative distribution function, respectively. It is necessary to remember that in this Subsection just the two-objective extension of expected improvement has been presented. Correction terms must be included for more than two objectives.

2.10 Fenestron Tail Rotors

Fenestron tail rotors[5] have been frequently used as anti-torque devices. They are normally used in light and medium size helicopters. The main advantage of fenestron tails is the lower power requirements than conventional open tail rotors to obtain a given thrust. The ducting increases the anti-torque device efficiency by reducing blade tip losses and adding some extra thrust. In order to obtain this efficiency enhancement, flow separation has to be avoided due to a proper inlet lip design. Additionally, the fin protects the anti-torque device from the main rotor wake, reducing the interference effect of the conventional rotor-craft [138]. It also reduces the fin interference with the tail rotor [154].

Other significant advantage is the reduction of noise, which makes military rotor-crafts more difficult to detect and civil rotor-crafts less disturbing for heliport communities. Edwards et al. [50] obtain a dramatic improvement in the sound quality: the total helicopter noise is reduced up to 6 dBA during hover and forward flight with reductions in tail rotor harmonics of 5 to 20 dBA. The helicopter noise spectrum is extended over higher ranges of frequencies as the number of blades increases. Unfortunately, humans are more sensitive to these high frequencies. Nevertheless, this high part of the spectrum is better absorbed by the atmosphere.

Other pros are the protection of the anti-torque device by the fin, which reduces the risk of any object impact and leads to safer ground approaching and landing [138]. Mouille [139] argues that more than 20% of helicopters crashes are due to a fail or impact on tail rotors, so the increase in reliability and the improvement in security provided by the fenestron rotors are important.

The main fenestron rotor drawback is the increase in anti-torque device weight. The structural requirements of the hub are larger than for a conventional tail rotor

[5] Also called shrouded tail rotor or fan-in-fin.

configuration. Since centrifugal forces are inversely proportional to the rotor diameter for the same rotor angular speed, fenestron centrifugal forces tend to be higher than in the conventional tail case[154]; hence, more structural requirements are needed.

Normally fan-in-fin rotors have more blades than conventional tail rotors. However, fenestron blade spans are shorter than in the conventional case. The rotational speed of the ducted tail rotors is higher than that of conventional rotors with the purpose of keeping a same order of magnitude tip speed. Also, typical blade root cut-offs are located in more outboard sections for fenestron configurations than for conventional cases.

CHAPTER III

RESEARCH METHODOLOGY

3.1 Introduction

The aim of this research is on modifying available techniques and developing new ones to build approaches for the more efficient use of computationally intense tools for initial design assessment and discrete alternative selection. Using computationally high-cost tools in conceptual design represents still a challenge for not only high-fidelity codes, such as CFD and CSD, but also mid-fidelity tools.

Long evaluation times could have different origins: complete physics-based models for single-objective problems, iterative tools in multidisciplinary problems, and the increase of design evaluations featured in multi-objective problems.

Taking a glance at the current literature, surrogate and adaptive sampling techniques handle inefficiently design alternatives. These design alternatives, generated by the iterative process of conceptual design, normally experience similar trends; however, surrogate and adaptive sampling techniques do not leverage these similarities. Similar trends can be identified in scenarios of "Evolutionary Incremental Concepts", where changes from one concept to another are small, and other scenarios where designers are aware of similar behavior across alternatives by experience.

As mentioned in Chapter 1.1, an improvement in the computational efficiency of concept surrogates where there is a choice of design alternatives could be crucial to use more efficiently intense tools in conceptual design. Certain objective functions show similar trends across all categories and concepts. Raymer [151] argues that in the conceptual phase designers normally rely on existing airfoils. Then, airfoils could be considered as a non-numeric variable in early design. For instance, while designing

a helicopter rotor, different blade airfoils result in similar figure of merit versus thrust coefficient curves. Therefore, it is reasonable to think that surrogates that cross-use computationally intense observations across lifting sections can provide more accurate surrogates than fitting a surrogate for each airfoil (current state-of-the-art) for a given function call budget.

In order to build this surrogate that leverages similar trends across categories, a distance function for categorical variables has to be defined. The Hamming distance, employed to handle nominal input attributes in instance-based learning techniques, can be one of the candidates to accomplish this task.

Furthermore, some design variables that are usually treated as categorical can have some natural distance due to underlying parameterizations that are not visible in the conceptual design stage. This intrinsic distance can be used to build the target surrogate. Measurable properties of the non-numeric input relevant for the pursued objective are combined to construct an intrinsic distance function for the categorical variables. Therefore, different objective functions hint different intrinsic distance functions for the same categorical set.

This intrinsic distance enables the handling of categorical design variable as a discrete-quantitative ones, which eases the surrogate building. An intrinsic distance function can be defined on a choice of materials when studying their structural resistance. A ductile material is expected to have similar behavior than other ductile material. However, its behavior compared to a brittle one is expected to be different. Thus, the intrinsic distance between the two ductile materials should be smaller than that between a ductile and a brittle material.

Distinct solutions for the same engineering problem, especially when immediate in the iterative conceptual design phase, usually experience landscapes very much alike. Multi-fidelity techniques are brought to re-use intense observations of previously sampled concepts while building a surrogate of a new concept whose objective function

is unknown. Current methods do not re-use observations from previous concepts while fitting new concept surrogates.

Accuracy of the optima in early stages is usually traded for interactions between design parameters, and sensitivity studies, among others. The combination of exploration and exploitation provided by Bayesian adaptive sampling methods is appropriate to study these interactions and sensitivities. Much research has been done in continuous adaptive sampling; however, no focus on discrete alternatives is found in the literature.

Among the Bayesian adaptive sampling techniques, the use of Kriging surrogates together with the expected improvement criterion gives an appropriate trade-off between the exploration and exploitation. Also, Kriging is a well-documented technique as a quick glance in the literature shows. The next sample in each step is determined by the maximization of the expected improvement in the design space.

The extension of this Kriging/ExI technique to explore and exploit surrogates that leverage similar trends across categories would use more efficiently computationally intense tools in conceptual design stages. Also, optimization/sampling techniques that can deal with not only continuous but also discrete variables would increase the efficiency in the use of these intense tools. MIO provides algorithms to optimize problems that possess both continuous and integer variables. Mixed-integer optimization techniques can be adapted to include more efficiently non-numeric variables. This adaptation is achieved by developing generation, cross-over and mutation functions suited to deal with categorical variables.

Later, the development of adaptive sampling techniques for surrogates that leverage similar trends from previously sampled concepts would help in making the most out of the limited function call budget on multi-modal and intense functions.

It is worth noticing that ExI functions have some same characteristics regardless the objective functions: the value is always greater or equal than zero in the whole

space, the value is zero in design points already sampled, and monotonicity properties in the mean prediction value and the mean squared error. This similarity between ExI landscapes allows extrapolating efficient mutation functions from one problem to another. It prompts to study the efficiency of several mutation operators while optimizing the expected improvement in a mixed-integer-categorical domain.

Then, the UH60A with fenestron tail is presented as the new concept for the ECMF surrogate practical test. It represents an incremental concept, which is built out of the UH60A with conventional tail, so they are expected to experience similar trends. The fenestron baseline values and its weight estimation are discussed.

Finally, the diagrams that review and summarize the research methodologies for categories and concepts are presented.

3.2 Research Questions: Hypothesis and Predictions

Research questions guide the investigations that are carried out in this thesis. It is important to remark that the proposed algorithms herein have a practical engineering purpose: develop tools for the more efficient use of computationally intense models that enhance the initial design assessment and discrete alternative selection.

The scientific method is not ruled out when performing the investigations. Even though no knowledge of physical phenomenon is pursued, efficient methods for conceptual design are pursued. Once the problem to tackle has been defined, a set of research questions are formulated. Then, their corresponding hypothesis to be further tested (with computational experiments) are discussed. Answers for these questions need a deep understanding of the current design tools and their limitations. Finally, results of the computational experiments are predicted.

3.2.1 First Question

Is it possible to build efficient surrogates for design scenarios where there exists a design categorical choice with similar trends?

a) Is it possible to build surrogates that cross-use computationally expensive observations across categorical choices with similar trends?

b) Which nominal distances allow building these efficient surrogates?

c) Does the MIC surrogate outperform independent surrogate modeling for each category (state-of-the-art)? How does the relative efficiency of these surrogates with respect to the independent ones depend on the training set size?

3.2.1.1 Hypothesis

Nominal distances permit building a Bayesian surrogate that includes not only continuous and integer design inputs but also categorical. This surrogate is called mixed-integer-categorical (MIC) surrogate. MIC surrogates cross-use observations across all categories.

Hamming distance is a possible nominal distance that allows building MIC surrogates. For certain engineering objectives, intrinsic distance measures can be obtained from functional dependence of the objective. MIC surrogates can also be based on this intrinsic distance.

The cross-use of observations in MIC surrogates is efficient for small training set sizes where only global trends are shared across categories. This cross-use of observations is not exploited by traditional independent surrogates fitted for each category.

The size of the training set affects the surrogate accuracy. The benefit obtained from other category observations becomes a rigidity when the training set becomes large. Above a large training set size, observations from other categories can capture high frequency information, which normally goes beyond common trends, and therefore can mislead the MIC surrogate, affecting its accuracy. So, for large training sets, independent surrogates, which have available enough information, can outperform MIC surrogates.

3.2.1.2 Prediction

A proof of concept can substantiate the hypothesis.

In terms of the performance metrics, MIC surrogates outperform the traditional independent surrogates (one surrogate for each category) for the small range of training sets. However, as the training set becomes larger, traditional methods gain efficiency up to a point where its performance metric surpasses that of MIC surrogates.

A working definition of performance metrics will be presented later. "Efficiency" means fewer observations are needed for a given accuracy or more accuracy for the same number of computationally expensive observations.

The resulting MIC surrogate is appropriate to be employed in conceptual design when a design categorical choice has to be made and objective functions are computationally intense.

3.2.2 Second Question

How could adaptive sampling approaches be efficiently extended to a choice of categories that experiments similar trends?

a) Can the MICGA extended from MIGA drive the ExI search while doing adaptive sampling on MIC surrogates?

b) Is the adaptive sampling on MIC surrogates more efficient than simultaneous adaptive sampling on independent surrogates (state-of-the-art) in some range of training set sizes?

3.2.2.1 Hypothesis

A set of categories can be adaptively sampled by applying the ExI criterion on MIC surrogates. The use of more accurate surrogates results in more efficient adaptive sampling processes. Therefore, the MIC surrogate, proposed in R.Q. 1, is chosen as the surrogate for the efficient extension of the adaptive sampling techniques.

Adaptive sampling approaches on surrogates like the ones proposed in R.Q. 1 needs of the extension of mixed-integer stochastic search techniques to spaces that contain non-numeric variables. This extension is possible by modifying the generation, crossover, and mutation functions of a mixed-integer genetic algorithm (MIGA). The MIGA extension to mixed-integer-categorical design spaces is called mixed-integer-categorical genetic algorithm (MICGA). MICGA can drive the adaptive sampling algorithm on MIC surrogates via ExI infill criterion.

While finding Pareto fronts with small warm-up training sets, the adaptive sampling on MIC surrogates via MICGA is more efficient than the state-of-the-art (simultaneous adaptive sampling on independent surrogates).

3.2.2.2 Prediction

A proof of concept can substantiate the hypothesis.

When a one-objective function is adaptively sampled using a MIC surrogate, the regions of high sampling density are the globally best performing areas but not the individual best performing regions for each category. It is an evidence that the MICGA can adequately drive the expected improvement search on MIC domains.

The quality of the Pareto frontier is the performance metric for multi-objective optimization. The adaptive sampling on MIC surrogates via MICGA outperforms the simultaneous adaptive sampling on traditional surrogates in terms of the quality of the Pareto frontier. It happens for small warm-up training sets. A working definition of the quality of the Pareto frontier will be presented later.

The resulting adaptive sampling approach is appropriate to be employed in conceptual design when a design categorical choice has to be made and objective functions are computationally intense and multi-modal.

3.2.3 Third Question

Is it possible to build efficient surrogates for design scenarios where there exist computationally expensive observations from a previous concept with similar trends?

a) Is it possible to build surrogates that re-use computationally expensive observations from previous concepts with similar trends?

b) Do ECMF surrogates outperform mono-fidelity surrogates (state-of-the-art)? How does the relative efficiency of ECMF surrogates with respect to the mono-fidelity ones depend on the new concept training set size?

c) What is the influence of the old concept training set size in the performance of ECMF surrogates?

3.2.3.1 Hypothesis

A multi-fidelity approach allows reusing observations from a previously sampled concept while building a new concept surrogate, these surrogates are called Evolutionary Concept Multi-Fidelity (ECMF) surrogates.

The re-use of observations in ECMF surrogates is efficient for small new concept training set sizes because supplementary information is useful in these cases. It is assumed that the new and old concept experience similar trends. This re-use of observations is not exploited by traditional mono-fidelity surrogates.

Training sets sizes for both concepts affect the ECMF accuracy.

a) The size of the new concept training set size affects the ECMF performance. Above a large new concept training set size, observations available from the new concept provide high frequency detail. In this case observations of the previously sampled concept are less needed and can only pollute this high frequency information degrading the ECMF accuracy. It does not happen to mono-fidelity surrogates, which have no observations from the previous concept.

b) The similar trends benefit obtained from previously sampled concept observations becomes a rigidity when the old concept training set becomes large. Above a large old concept training set size, high frequency information, which normally goes beyond common trends, is captured, so it can mislead the ECMF surrogate and reduce its accuracy. Thus, for large old concept training set sizes, the performance of the ECMF surrogate saturates or even degrades as the old concept training set becomes bigger.

3.2.3.2 Prediction

A proof of concept can substantiate the hypothesis.

In terms of the performance metrics, the ECMF surrogates outperform the traditional mono-fidelity surrogates for small ranges of new and old concept training set sizes. However, as the new concept training set size increases, mono-fidelity surrogates gain efficiency with respect to ECMF ones in terms of the performance metrics. Also, for large old concept training sets, the ECMF surrogate error does not decrease as the old concept training set becomes larger. A working definition of performance metrics will be presented later.

The resulting ECMF surrogate is appropriate to be employed in conceptual design when there exist evolutionary concepts with similar trends and objective functions are computationally intense.

3.3 Nominal Distance for Cross-Using Observations

As discussed in Section 2.5, the distance between design points is the input to tune Kriging surrogates. In order to include categories in the same meta-model, it is necessary to define a categorical distance, which is also called nominal distance.

3.3.1 Requirements for Nominal Distance

The desired requirements for the nominal distance are the following:

Lack of order As explained in Section 2.1, non-numeric variables are characterized for the lack of order.

Definition to keep radial basis functions meaningful Since Kriging meta-models are employed in this work, the nominal distance should result in positive definite covariance matrices. If two categories are too close (distance-wise), these matrices may be non-positive definite. It would bring bias in the Kriging parameters, which has consequent numerical problems.

Versatility to different objectives Designers would like to have a nominal distance that is useful for many objective functions.

Once the desired characteristics of the nominal distance are defined, the search for the nominal distance function can begin. Three possible nominal distance functions are presented:

1) Integer Nominal Distance.

2) Hamming Nominal Distance.

3) Intrinsic Nominal Distance.

3.3.2 Integer Nominal Distance

The nominal variable is arbitrarily mapped to a set of integer values. Each category of the non-numeric variable is mapped to an integer in the interval $[1, |x_{nom}|]$, where $|x_{nom}|$ is the number of points in the categorical input x_{nom}. Then, the distance that is input to the Kriging surrogate is based on this new arrangement of x_{nom} in the integer line.

It is the simplest nominal distance function. The first disadvantage is that it implicitly defines an order between categories. Another drawback is that the random mapping to the integer set does not take into account the relation objective-category.

For a large number of categories, the distance between two integers could be too close which may produce non-positive definite covariance matrices. However, it does not happen for a small number of categories. It is versatile which is useless since the integer distance implicitly defines an order and the covariance matrices can be non-positive.

3.3.3 Hamming Nominal Distance

The most popular nominal distance is the Hamming distance proposed by Hamming [78],

$$d_{ham}\left(\mathbf{x}^{(i)}, \mathbf{x}^{(l)}\right) = \sum_{j=1}^{k} Ih\left(x_j^{(i)}, x_j^{(j)}\right) \tag{68}$$

where a, and b are categories and

$$Ih\left(a, b\right) = \begin{cases} 0 & \text{if } a = b \\ 1 & \text{if } a \neq b \end{cases} \tag{69}$$

It is noticeable that Equation 68 assumes that the distance between points is the same and no order is established. The Hamming distance permits building metric spaces in information theory and general topology as is shown by Ash [6] and Kelley [108], respectively. Also, Aha et al. [2] combine the Hamming and Euclidean distance to include together categorical and continuous variables in instance-based learning algorithms.

The Hamming distance respects the nature of non-numeric variables by not defining an order between categories. It is versatile since the same distance can be used for all the objectives. The main drawback is that the Hamming distance just compares categories of design points without taking into account the objective function. Finally, the Hamming distance keeps radial basis functions meaningful since the distance between points is always one, which avoids having two categories too close.

3.3.4 Intrinsic Nominal Distance

In some instances there is some intrinsic metric for a given objective; so, it may be possible to define other efficient nominal distances. This intrinsic distance can be given by an underlying parameterization or properties of the categorical members.

Many categorical choices have an underlying parameterization that is not explored in conceptual design. For instance, in preliminary or detailed design phases, airfoils can be parametrized into continuous spaces; these parameterizations can only be fully exploited by high-fidelity tools such as CFD codes. However, it does not make sense to use these parameterizations in conceptual design stages, where only the most important system factors are studied and the tools do not include physics with the required details to make the most out of the parameterization. Additionally, in conceptual design, engineers cannot afford these time-consuming parametric studies. Therefore, design teams treat airfoils as categorical variables in conceptual design ignoring their underlying parameterization.

In conceptual design, variables with an underlying continuous space could usually be simplified in terms of categorical properties instead of the hidden continuous parameterization. For instance, in the case of airfoil choices the difference between airfoils is given in terms of aerodynamic curves, which is considered a categorical property. So, these properties can be used to build the intrinsic distance. Categorical properties can be scalar properties or function properties.

A possible example of scalar properties is found when there is a set of materials to study in homogeneous and isotropic elasticity. Each material's Hooke's law in 3D is defined by two scalar parameters. Two possible material defining parameters are the modulus of elasticity and Poisson's ratio. It is reasonable to define a distance between two materials using the value of these defining parameters when the elastic material behavior is pursued. However, for analysis different from elasticity, other parameters turn to be important: fatigue limit for fatigue analysis, ultimate strength

and ductility for plasticity analysis,...

An example of function properties is taken from helicopter design, which is an application of this investigation. Let us assume that the designer has to study the aerodynamic efficiency of a helicopter rotor in hover and is given a set of airfoils. In this problem, the set of airfoils is seen as a non-numeric design input. A simple rotor model using Blade Element Momentum (BEM) theory, details explained by Leishman [123], hints that the influence of the airfoils on the rotor aerodynamic efficiency is introduced by the airfoil aerodynamic characteristic curves: $C_l(\alpha)$, $C_d(\alpha)$, or $C_d(C_l)$, a.k.a. polar curve (the combination of $C_l(\alpha)$ and $C_d(\alpha)$ by removing the angle of attack as parameter). Therefore, it makes sense to base the distance function between two airfoils in terms of these characteristic curves if the rotor aerodynamic efficiency is pursued. Note that the airfoil aerodynamic curves are the properties that influence the objective. These properties are functions unlike in the first example about materials, where the properties were scalars.

In the what follows, a discussion for the assessment of the intrinsic distance function is presented. x_{nom} is a categorical design input whose values have J properties that could define the possible intrinsic distance between category members. For scalar properties, the intrinsic nominal distance between two categories, $x_{nom}^{(i)}$ and $x_{nom}^{(l)}$, could be given by

$$d_{intr}\left(x_{nom}^{(i)}, x_{nom}^{(l)}\right) = \sum_{j=1}^{J} w_j |prop_j\left(x_{nom}^{(i)}\right) - prop_j\left(x_{nom}^{(l)}\right)| \tag{70}$$

where $prop_j\left(x_{nom}^{(l)}\right)$ is the j-th categorical property of the l-th categorical member, and w_j is the weight for the jth property of the categorical design input. w_j weights the relative importance of the jth property compared to other properties. If a sole property defines the intrinsic distance, then, it can be written as

$$d_{intr}\left(x_{nom}^{(i)}, x_{nom}^{(l)}\right) = |prop\left(x_{nom}^{(i)}\right) - prop\left(x_{nom}^{(l)}\right)| \tag{71}$$

Clearly the above metric cannot be applied when the property is a function instead of a scalar. A simple alternative is to choose a representative point of the function via averaging or experts' guess. Specifically for the helicopter instance previously discussed, a desired operating point like α_{opt} can be chosen, so aerodynamic characteristic curve values at α_{opt} are possible scalars to feed to Equation 70, and 71. As is well-known, helicopters usually have different angles of attack in different blade span locations because of not only blade twist, but also different rotational velocity and induced velocity in each span location. As a result, the scalar approach is poor, unless one deals with the unlikely case of an optimal rotor in its optimal design point (in this case all blade sections operate at the optimum angle of attack). For further reference, see[123].

Then, the goal is to look for a distance measure where the properties are in function spaces. Function spaces are the base of "Calculus of variations", see Sagan's work [161] and Gelfand et al.'s work [70]). The space of continuous functions of one variable between points a and b is denoted as S,

$$S \in C[a, b]$$

A more detailed explanation about functional analysis and function spaces is provided by Balakrishnan [10] and Zeidler [212].

There exist several function space metrics based on different norms. Sutherland [181] presents them with examples. Let f and g be two functions. Among the most important metrics, one finds infinity metric based on infinity norm $d(f,g) = \| f-g \|_{\infty}$, metrics over L^p integrable spaces $d(f,g) = \| f-g \|_{L^p}$, and metrics over H^1 Sobolev spaces $d(f,g) = \| f-g \|_{H^1}$. Equation 72 represents the L^1 metric between elements in S.

$$d_{intr}\left(x_{nom}^{(i)}(z), x_{nom}^{(l)}(z)\right) = \int_a^b |prop\left(x_{nom}^{(i)}(z)\right) - prop\left(x_{nom}^{(l)}(z)\right)| dz \qquad (72)$$

where z is the independent variable of the categorical input property $prop\,(.)$.

The definition of the distance for the space of continuous functions of one variable, Equation 72, brings up the issue: the choice of the integration limits. Their definition depends on the expected bounds of the categorical design input property. If one hypothesizes that, for the helicopter problem, $C_d(\alpha)$ is the curve with the major influence in the rotor aerodynamic efficiency, then $prop\left(x_{nom}^{(i)}\,(z)\right)$ turns into $C_d\left(air^{(i)}\,(\alpha)\right)$ (or $C_d^{(i)}\,(\alpha)$ for notation's simplicity). Post-stall angles of attack are not expected to happen because the rotor aerodynamic efficiency is pursued. So, an upper limit for b can be α_{stall}. For the lower limit, α_{inf}, a negative angle of attack, small in magnitude, seems to be a reasonable value. Particularizing Equation 72 to the helicopter problem

$$d_{intr}\left(air^{(i)}\,(\alpha)\,, air^{(l)}\,(\alpha)\right) = \int_{\alpha_{inf}}^{\alpha_{stall}} |C_d^{(i)}\,(\alpha) - C_d^{(l)}\,(\alpha)\,|d\alpha \tag{73}$$

Similarly as in the scalar property case, weighting could generalize Equation 72 when more than one function property is involved in the intrinsic distance. For the sake of brevity, $d_{i,l}$ denotes from now on the nominal intrinsic distance between categorical members i and l unless otherwise said.

Regarding desired nominal distance requirements, the intrinsic distance respects the nature of non-numeric variables by not defining an order between members. By definition, it is in agreement with the relation category-objective. However, it has to be changed for a different objective, so there is no versatility. Finally, the intrinsic distance may result in non-positive covariance matrices in cases where categorical properties of two members are too close.

Finally, a modified intrinsic distance is presented to avoid non-positive covariance matrices. So, this modification of the intrinsic nominal distance targets cases where two categorical members, i and l, are too close, which produces non-positive covariance matrices. In order to avoid this effect, the minimum categorical distance of the

set of categories is increased in value. Let d_{min} denote the minimum nominal distance of the set of categories S_{cat}

$$d_{min} = min\left(d_{i,l}\right) \qquad \forall i, l \in S_{cat}$$

where $d_{i,l} = d_{intr}\left(x_{nom}^{(i)}\left(z\right), x_{nom}^{(l)}\left(z\right)\right)$ is assessed by Equations 70 or 72 (the non-modified intrinsic nominal distance).

The minimum distance can be adjusted to a desired value $d^{\star}$. It is done by translating and scaling $d_{i,l}$ so that a new minimum distance and the maximum distance are set to $d^{\star}$ and 1, respectively.

$$d'_{i,l} = d^{\star} + \frac{1 - d^{\star}}{1 - d_{min}}\left(d_{i,l} - d_{min}\right) \tag{74}$$

where $d'_{i,l}$ is the new nominal distances for avoiding non-positive covariance matrices.

3.4 MIC Surrogate: Leveraging Similar Trends across Categories

As stated in Section 1.6, one of the main goals of this thesis is to build surrogates that leverage similar trends across categories. The current state-of-the-art constructs an independent surrogate for each category. If the proposed surrogate is successful, a better use of the limited computational budget can be done resulting in more optimal and reliable initial designs and categorical selection. As in any usual scientific research, a set of computational experiments is used to test the efficiency of the proposed surrogate.

Surrogates usually cope with continuous and/or integer (or discrete-quantitative) variables. In the presence of categorical inputs, independent surrogates are fitted for each category. The proposed surrogate herein includes not only continuous and discrete-quantitative variables but also non-numeric ones; it is called mixed-integer-categorical (MIC) surrogate. The addition of the categorical variables is done with the idea of cross-using useful observations across categories. It would lead to more efficient

surrogates than the ones currently employed. The Hamming distance is a perfect candidate by which to include the categorical design inputs in the surrogate domain, in a similar manner that Wilson and Martinez [201] do for statistical classification.

In order to build a MIC surrogate it is necessary to have a training set. They are obtained with regular Latin hypercube sampling[96]. The categorical and discrete-quantitative variables are treated as continuous when getting the training set. Specifically, the categorical ones are first mapped to integer variables and later handled as continuous. Once the Latin hypercube sampling produces the set, the categorical and discrete-quantitative variables are rounded-off such that samples are evenly divided among categories and discrete-quantitative values. It is important to emphasize the selection of continuous Latin hypercube techniques that make the most out of the training set; there are different training points in each category, so the cross-use of observations is more efficient than in the case of same training points in each category. In the case of the same training set in each category, observations from other categories are not that much valuable for a given category (the objective values at the sample points of the given category are already known).

The type of surrogate chosen to build the MIC surrogate is Kriging, discussed in Section 2.5, because eventually adaptive sampling techniques are pursued, which requires the surrogate prediction uncertainty. Kriging bases its estimates on its correlation function, given in Equation 9, that needs the distance between the prediction point $\mathbf{x}$ and the sample points. So, the prediction is made according to the proximity to sample points: when the design point $\mathbf{x}$, at which the objective value is pursued, is close to a sample point, then the objective value at this sample point influences the prediction at the design point $\mathbf{x}$.

When dealing with continuous and discrete-quantitative variables, Kriging uses the Manhattan distance between points to assess proximity. The reason for preferring the Manhattan over the Euclidean distance, even in cases with only continuous

design inputs, is to decouple design variables. It allows taking into account that some variables affect more the objective function than others. It is highly recommended to normalize the bounds of the design space in a hypercube $\{0,1\}^k$, if possible, in order to avoid a misleading dependence of the objective function on the design variables.[1]

This work claims that for categorical variables the Hamming distance, defined in Equation 68, is appropriate to be inputted to the Kriging correlation matrix, see Equation 9. Other alternative, if possible, is to use the intrinsic distance for the given objective function, defined either by Equation 70 or 72 depending if the categorical property is a scalar or a function. It is worth reminding that this intrinsic distance is defined with the use of relevant properties of the categorical variable to the given pursued objective function.

3.4.1 Performance Indicators for MIC Surrogates

The MIC surrogate is considered successful if it predicts more accurately the objective function than the current multiple independent surrogates (one for category) in a given range of training points. Also, MIC surrogates based on several nominal distances are to be compared. The comparisons are done in terms of some success indicators[2]. A straight-forward measure of accuracy of a model is the "standardized validation error" between the real and predicted value

$$error\left(\mathbf{x}\right) = \frac{\left|y\left(\mathbf{x}\right) - \hat{y}\left(\mathbf{x}\right)\right|}{y^{\star}_{ch}} \tag{75}$$

where $y\left(\mathbf{x}\right)$ and $\hat{y}\left(\mathbf{x}\right)$ are the real model and its approximation, respectively, and y^{*}_{ch} is the characteristic change of the function in the studied domain. This last value is brought to normalize the error.

[1] It is a good practice to normalize the design space because, as it could be seen, if one of the design inputs x_i has a very large range, it could artificially overpower other design inputs.

[2] Also called performance indicators.

Probabilistic surrogates such as Gaussian processes provide the epistemic uncertainty of the prediction. In Kriging, it is given by the mean squared error, $\hat{s}(\mathbf{x})$; Equation 18 shows it in the interpolating case. Other success indicator is the number of mean squared errors that the predicted value is off the actual value. This indicator called "standardized validation residual" could be written as follows

$$res(\mathbf{x}) = \frac{|y(\mathbf{x}) - \hat{y}(\mathbf{x})|}{\hat{s}(\mathbf{x})} \tag{76}$$

It is known that 99.7% of values drawn from a normal distribution, $\mathcal{N}(\mu, \sigma^2)$, are within the interval $[\mu - 3\sigma, \mu + 3\sigma]$. Thus, the "standardized validation residual" represents the confidence interval in the accuracy of the prediction.

In order to assess the accuracy of the meta-model in its entire domain, a possible option is to calculate the root mean squared (rms) of the $error(\mathbf{x})$ and $res(\mathbf{x})$ over the whole domain as follows

$$error_{rms,global} = \sqrt{\frac{1}{V}\int_V \frac{|y(\mathbf{x}) - \hat{y}(\mathbf{x})|^2}{y_{ch}^{\star 2}} dV} \tag{77}$$

$$res_{rms,global} = \sqrt{\frac{1}{V}\int_V \frac{|y(\mathbf{x}) - \hat{y}(\mathbf{x})|^2}{\hat{s}^2(\mathbf{x})} dV} \tag{78}$$

where V is the design space.

Expressions 77 and 78 are easily evaluated for cheap functions but not for computationally intense objective functions due to the need of assessing the real function value, $y(\mathbf{x})$. The main point of a surrogate is to predict a computationally intense objective function with few real function evaluations. So, an alternative method to assess the success indicators is needed.

Other alternative is to select a few additional points as validation or test points and compare the true and predicted values, but it is still a loss of resources because

it requires sampling beyond the points used to fit the surrogate. A better option is called "cross-validation" that permits to evaluate model accuracy without losing any computational resources.

Jones [99] et al. propose this cross-validation technique. It consists in removing the point i out of the training set; later, the surrogate is built based only on the remaining $N-1$ points; and then, the value of the objective at the point that has been left out is predicted, $\hat{y}_{-i}(\mathbf{x}^i)$ (the subscript $-i$ highlights that the training point $\mathbf{x}^i$ has not been used as a training point in the prediction of $\hat{y}_{-i}(\mathbf{x}^i)$). Instead of tuning over and over again the model hyper-parameters for each reduced training set, the values of the hyperparameters when tuning the whole training set (point $\mathbf{x}^i$ included) are used. It is because little changes are expected in those unless few observations exist. In the Kriging surrogate, the i-th column and row of the correlation matrix $\mathbf{\Psi}$ are removed, and so is the i-th component of the vectors ψ and $\mathbf{y}$. Similarly, cross-validated mean squared error, $\hat{s}_{-i}(\mathbf{x}^i)$, is assessed at the i-th point. Cross validation has been employed as well by other authors, see Kleijnen et al.'s work [113].

New local success indicators are defined: "standardized cross-validation error" by Equation 79, and "standardized cross-validation residual" by Equation 80,

$$error_{cvi,i} = \frac{|y(\mathbf{x}^i) - \hat{y}_{-i}(\mathbf{x}^i)|}{y^{\star}_{ch}} \tag{79}$$

$$res_{cvi,i} = \frac{|y(\mathbf{x}^i) - \hat{y}_{-i}(\mathbf{x}^i)|}{\hat{s}_{-i}(\mathbf{x}^i)} \tag{80}$$

In order to get global success indicators, the rms of all the sample points are added

$$error_{rms,cvi} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} error^2_{cvi,i}} \tag{81}$$

$$res_{rms,cvi} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} res^2_{cvi,i}} \tag{82}$$

Also, the influence of the size of the training set on the surrogate performance is sought, see Section 3.2. Specifically, how quickly the rms of the surrogate "standardized error" converges to zero as the training set becomes larger and larger. A power relationship is fitted to the predicted rms of the "standardized validation error" for analyzing this convergence behavior.

$$error_{rms} = \frac{error_{rms}\left(N_{tr.set} = 0\right)}{N_{tr.set}^{\alpha}} \tag{83}$$

where N is the number of training set samples, and α tells how quickly the rms of the error converges to zero. The value of α enables to see the effect of an increase in the training set size on the rms of the error.

Note that the MIC surrogate and its performance indicators explained so far in this section assume a noise-free function. However, a similar procedure for noisy functions is developed, where just the values of $\hat{y}$ and $\hat{s}$ are taken from Subsections 2.5.1 and 2.5.2, respectively.

3.5 ECMF Surrogate: Leveraging Similar Trends from Previous Concepts

Section 2.6 explains in details multi-fidelity Gaussian random field meta-models that combine two different fidelities. The multi-fidelity framework to be presented in this Section re-uses observations of previously sampled concepts in new ones when there exist similar trends. Nonetheless, current surrogates are fitted independently for each concept. Concepts with similar trends are common in conceptual design stages, where the iterative process keeps incrementally evolving the configuration of the solution ("evolutionary incremental concept"). The multi-fidelity surrogate presented herein is targeted to these evolutionary concepts; thus, it is called evolutionary multi-fidelity (ECMF) surrogate.

The addition of previous concept observations via multi-fidelity approach is done

with the idea that the reuse of the observations results in more efficient surrogates than the ones that just use observations of the concept itself. It allows designers to make a better use of the limited computational budget. Again, computer experiments are used to test the efficiency of reusing computationally intense observations from previous concepts. The concept previously sampled is called concept 1, whereas the new concept is called concept 2. Two main assumptions are made while building the ECMF surrogate: 1) concept 1 design space is included in concept 2 design space; and 2) there is a constant behavior in the concept 1 training set extension along the concept 2 exclusive design variables. They are discussed in more detailed in this Section.

Traditional multi-fidelity methods deal with variable-fidelity codes that have the same design space. However, concepts have different design spaces: dimensionalities and design variables may differ across concepts. Space mapping, which is developed in the microwave circuit community [12] [155], is a method of linking variable-fidelity models. This technique is capable of linking different design spaces of the same concept from variable fidelities that may or may not share design variables.

For simplicity's sake, it is assumed herein that the new concept has all the design variables of the previous concept plus some new ones exclusive for the new concept. It matches with the conceptual design iterative scenarios, where concepts evolve from one to the other incrementally (which is previously called "evolutionary incremental concepts"). The reason for the previous assumption is that the main purpose is to study the effectiveness of reusing computationally intense observations from previous concepts; thus, the applicability of the ECMF surrogate to a wider range of design spaces is left as a secondary factor to study in the future. Good candidates for this extension to a wider range of design spaces are space mapping techniques [12]

Firstly, a Gaussian surrogate is fitted to the concept 1, $\hat{y}_c^{(D1)}$, in its designed space, denoted as $D1$. It means that the hyper-parameters $\mu_c^{(D1)}$, $\sigma_c^{(D1)}$, and $\boldsymbol{\theta}_c^{(D1)}$ are

found. Note that the super-index indicates the design space of the hyper-parameter or training data-sets.

Then, the concept 1 training set is extended from $D1$ to the concept 2 design space, denoted by $D2$. There are two possible ways for carrying this extension out. The first one is to assume a constant behavior of the training set along the concept 2 exclusive design variables. The second option is to use known data to assume a given behavior along these concept 2 exclusive design variables. The present work assumes that the only information available on the new concept is the following: a) its observations ($\mathbf{y}_e^{(D2)}$), and b) its similar trends to concept 1 (about which some observations $\mathbf{y}_c^{(D1)}$ are known); thus, the constant behavior for the concept 1 training set extension to $D2$ seems to be the best option. Also, it is important to realize that the incremental changes from one concept to another are likely to produce small changes in the response compared to those of the main design factors, which are present in both concepts.

In order to carry out this concept 1 training data-set extension to $D2$, $\mathbf{X}_c^{(D1)}$ and $\mathbf{y}_c^{(D1)}$ are placed at the ends of the concept 2 exclusive design variables, i.e., $\mathbf{X}_c^{(D2)} = \{\left[\mathbf{X}_c^{(D1)}\mathbf{0}^{(exclD2)}\right], \left[\mathbf{X}_c^{(D1)}\mathbf{1}^{(exclD2)}\right]\}$, and $\mathbf{y}_c^{(D2)} = \{\mathbf{y}_c^{(D1)}, \mathbf{y}_c^{(D1)}\}$.

Then, a new mono-fidelity surrogate for concept 1, $\hat{y}_c^{(D2)}$, can be calculated in the new concept design space $D2$. A surrogate computation from scratch is a waste of resources because the surrogate is just a constant extension from the one in $D1$ to a new one in $D2$. The hyper-parameters of the surrogate $\hat{y}_c^{(D2)}$ can be obtained from the ones for $\hat{y}_c^{(D1)}$, specifically $\mu_c^{(D2)} = \mu_c^{(D1)}$, $\boldsymbol{\theta}_c^{(D2)} = \{\boldsymbol{\theta}_c^{(D1)}, \boldsymbol{\theta}_c^{(exclD2)}\}$; once $\mu_c^{(D2)}$ and $\boldsymbol{\theta}_c^{(D2)}$ are obtained, $\sigma_c^{(D2)}$ is given by Equation 14 for the interpolating case. Since the new surrogate is a constant extension along concept 2 exclusive design variables, denoted by $exclD2$, $\boldsymbol{\theta}_c^{(exclD2)}$ should be set to $\mathbf{0}^{(exclD2)}$. However, this option produces numerical problems such as singularities in the covariance matrices $\boldsymbol{\Psi}_c\left(\mathbf{X}_c^{(exclD2)}, \mathbf{X}_c^{(exclD2)}\right)$, and $\boldsymbol{\Psi}_c\left(\mathbf{X}_e^{(exclD2)}, \mathbf{X}_e^{(exclD2)}\right)$, see Equation 29. In order to solve this problem, the

zero values are substituted for a fraction of the minimum component of the vector $\boldsymbol{\theta}_c^{(D1)}$, i.e., $\boldsymbol{\theta}_c^{(D2)} = \{\boldsymbol{\theta}_c^{(D1)}, \frac{\min \boldsymbol{\theta}_c^{(D1)} \mathbf{1}^{(exclD2)}}{30}\}$ where $\min \boldsymbol{\theta}_c^{(D1)}$ is the minimum value of the vector $\boldsymbol{\theta}_c^{(D1)}$.

Once the hyper-parameters of the cheap model in $D2$ are tuned, Equations 31 and 32 provide the value of the surrogate hyper-parameters $\mu_d^{(D2)}$ and $\sigma_d^{(D2)}$ for the concept 2 surrogate, $\hat{\mathbf{y}}_e^{(D2)}$. The hyper-parameter $\boldsymbol{\theta}_d^{(D2)}$ is obtained by a GA search as explained in Section 2.6. Figure 9 shows a diagram with the proposed multi-fidelity approach.

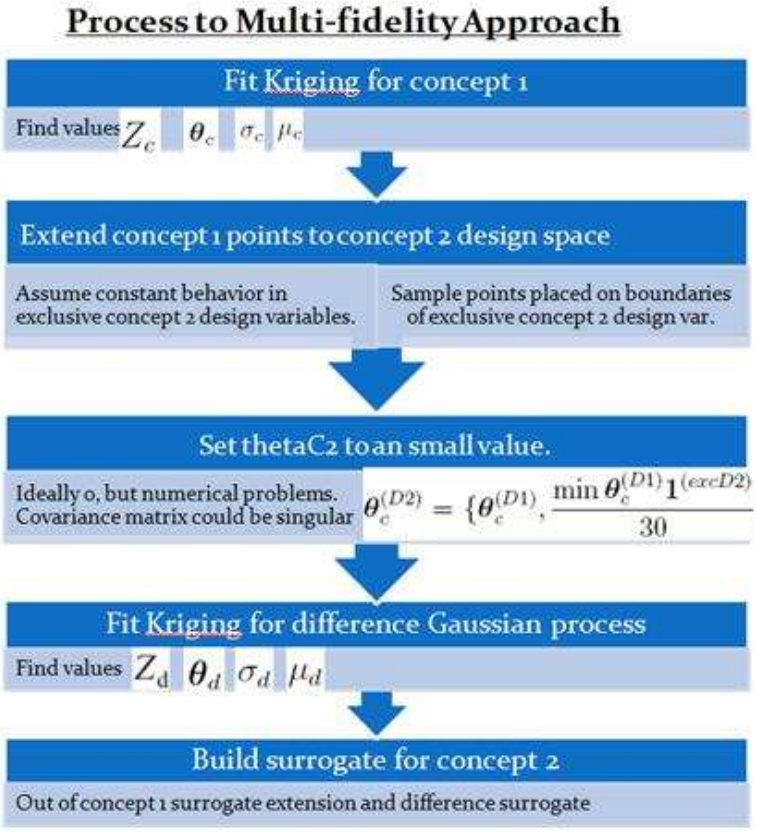


Figure 9: Process for ECMF Surrogate

Finally, the prediction and the mean squared error of the multi-fidelity surrogate for concept 2 could be assessed with Equations 33 and 34, respectively.

Note that the multi-fidelity approach explained so far in this section assumes the objective function is noise-free. However, a similar procedure for noisy functions is

developed with the addition of the hyper-parameter λ_e. Also, equations for $\hat{\mu}$ and $\hat{\sigma}$ are taken from Subsections 2.6.1 and 2.6.2.

3.5.1 Performance Indicators for ECMF Surrogates

The method is considered successful if the ECMF surrogate predicts more accurately the objective function of the new concept than the current mono-fidelity surrogate for a given range of concept 1 and 2 training data-set sizes. The success indicators employed to measure success are the same ones as in Subsection 3.4.1: rms of the "standardized validation error", and "standardized validation residual".

If the objective function is available in the design space, the rms of the "standardized validation error" and "standardized validation residual" are given by

$$error_{rms,global} = \sqrt{\frac{1}{V}\int_V \frac{|y_e(\mathbf{x}) - \hat{y}_e(\mathbf{x})|^2}{y_{e,ch}^{\star 2}} dV} \tag{84}$$

$$res_{rms,global} = \sqrt{\frac{1}{V}\int_V \frac{|y_e(\mathbf{x}) - \hat{y}_e(\mathbf{x})|^2}{\hat{s_e}^2(\mathbf{x})} dV} \tag{85}$$

If the cross-validation technique is used (see Subsection 3.4.1), the rms of the "standardized validation error" and "standardized validation residual" are given by

$$error_{rms,cvi} = \sqrt{\frac{1}{N_{2,tr.set}} \sum_{i=1}^{N_{2,tr.set}} \left(\frac{|y_e(\mathbf{x}^i) - \hat{y}_{e,-i}(\mathbf{x}^i)|}{y_{e,ch}^{\star}} \right)^2} \tag{86}$$

$$res_{rms,cvi} = \sqrt{\frac{1}{N_{2,tr.set}} \sum_{i=1}^{N_{2,tr.set}} \left(\frac{|y_e(\mathbf{x}^i) - \hat{y}_{e,-i}(\mathbf{x}^i)|}{\hat{s}_{e,-i}(\mathbf{x}^i)} \right)^2} \tag{87}$$

3.6 Adaptive Sampling of MIC and ECMF Surrogates

One of the goals of this work is to develop adaptive sampling techniques for surrogates that leverage similar trends. The ExI, explained in Subsection 2.7.1, is the infill criterion chosen herein to adaptively sample design spaces. Typical objective

functions in design are unknown and likely to be multi-modal and non-convex. ExI permits exploring regions of the design domain with high uncertainty and exploiting potentially good performing regions.

As previously said, the ExI is optimized throughout the whole domain. For a modest size of training datasets, multiple valleys are expected in ExI functions; therefore, gradient techniques are useless and are ruled out. Instead global search techniques are needed to look for the maximum ExI. This investigation chooses a GA as a global ExI optimizer due to its high popularity and easy availability. The adaptive sampling algorithm is halted in this research when a given number of updates is reached.

A MICGA is developed to adaptively sample MIC surrogates whose domains have continuous, discrete-quantitative, and non-numeric variables. The current state-of-the-art EGO algorithm to compare with is the simultaneous adaptive sampling of independent surrogates.

Also, adaptive sampling on ECMF surrogates is performed. However, it is not a technical challenge because only one concept is handled at a time and its domain is at most MIC. In the last case, a new concept with a MIC domain, the adaptive sampling algorithm gets reduced to the same problem, adaptively sampling a MIC surrogate.

Regarding the adaptive sampling of MIC surrogates, the first step is to make sure that the optimizer MICGA can drive the adaptive sampling in MIC domains. So, computer experiments are done to see if there is EGO behavior in the sampling process. EGO behavior means that there is a higher concentration of updates in globally high-performing areas and not just in high performing regions of each category.

3.6.1 Performance Indicators for Adaptive Sampling

The second step is to define success indicators for multi-objective adaptive sampling on MIC and ECMF surrogates. Optimal solutions for multi-objective scenarios are

Pareto fronts. So, the success indicators are given by the quality of Pareto fronts. The adaptive sampling algorithms provide a non-dominated set that intends to approach the ‘ “real” Pareto set. Two performance indicators are used for this purpose:

Number of points in the non-dominated set It measures the level of detail of the non-dominated set.

The average distance between the non-dominated set and “real” Pareto front It quantifies the proximity between the approximating non-dominated set and the real Pareto front. The average distance $\bar{d}(F, PF)$ is given by

$$\bar{d}(F, PF) = \sum_{i=1}^{q_{real}} \min(\mathbf{y}_i, PF) \tag{88}$$

where $\mathbf{y}_i$ is the i-th point of the approximating non-dominated set, PF the “real” Pareto front which is assessed by the method NSGA-II developed by Deb [41], and q_{real} is the number of Pareto members of PF.

Once the performance indicators of the adaptive sampling experiments are defined, the MICGA that would search mixed-integer-categorical design spaces is presented.

3.6.2 Mixed-Integer-Categorical Genetic Algorithm

The ExI infill criterion drives the adaptive sampling of MIC surrogates. The next sample point is obtained from the solution of an optimization problem in the MIC design space, which is composed with attributes of several nature: continuous, discrete-quantitative, and categorical. Section 2.8 reviews the MIO techniques: mixed-integer programming, and stochastic MIO searches. The desired qualities for the optimizer of the ExI in the MIC domain are:

1) Global search. ExI function is multi-modal.

2) Non-linear functions. ExI is non-linear.

3) Category handling.

4) Availability and open source.

5) Low computational expense.

The first three qualities are crucial due to the nature of the optimization problem to solve (the ExI optimization in a MIC domain). Figure 10 shows the qualities satisfied by mixed-integer programming and stochastic MIO searches. Stochastic MIO searches satisfy two of the three crucial qualities; therefore, they are chosen to optimize the ExI, specifically a mixed-integer genetic algorithm is selected. However, the optimization tool needs to deal with categories, i.e., a mixed-integer-categorical optimizer is needed. The question is: how can a MIGA be extended to search mixed-integer-categorical design spaces?

Types / Requir.	MI Programming	Stochastic Searches
Global search	×	✓
Non-linear func.	×	✓
Categories	×	×
Availability/Open source	✓	≈
Computational expense	✓	×

Figure 10: Qualities of the MIO Methods

Note that in Figure 10, the symbol ≈ means that the tools are available but they are restricted-source.

First of all, it is explained the conversion from continuous GA to MIGA. The first possible option to obtain a MIGA is to treat all the variables as continuous; then,

once the optimal solution is obtained, it is rounded-off. Also Karoui et al. [103] propose a progressive rounding-off of discrete variables. Capitanescu and Wehenkel [27] point out the problems for these two approaches: a) a likely deterioration of the objective value, and b) to ensure feasibility a method to restore it is necessary, which results in additional coding. These approaches are ruled out due to these downsides. Better suited generation, crossover, and mutation functions are developed to properly treat the integer variables. These functions have led to good performing MIGA [44]. Several modification on MIGA will be proposed in this section to convert them into mixed-integer-categorical GA.

Deep et al. [44] develop a GA to solve mixed-integer problems. However, its access is strictly limited. MATLAB® 2011b provides a GA with mixed-integer capabilities based on Deep et al.'s work. Again, the functions to carry out the optimization process are encrypted in private files, so no modifications can be done to build a MICGA out of the MATLAB® MIGA. On the other hand, all the MATLAB® functions that are called by the MATLAB® continuous GA solver are accessible. Therefore, the first step towards MICGA is to build a MIGA out of the MATLAB® continuous GA, and later the MICGA will be implemented.

Modifications on the generation, mutations, and crossover functions are carried out to convert the continuous GA into MIGA according to the literature available. The initial population of continuous and integer variables are uniformly randomly generated. Gaussian and geometric mutation distributions are used for continuous and integer variables, respectively. Regarding crossovers, both continuous and integer values are crossovered intermediately. If the resulting integer crossover value is not integer, rounding is applied with equal probability for ceiling and flooring. As a reminder the mutation and crossover functions are briefly classified in Subsection 2.8.3. The resulting solver is tested against typical mixed-integer problems and compared with the source-restricted MATLAB® 2011b MIGA.

MIGAs can be used to pursue the global optimum on a MIC domain. The MIGA can optimize the ExI of one category at a time. Later, a comparison between categorical optimum selects the global optimum. However, a more efficient handling of the categorical variable is to modify the generation, cross-over, and mutation of the categorical part of the population genome. In such a way, just one optimizer call is needed, and the population is expected to focus on the best categories and, therefore, save ExI calls in categories with low ExI.

Once the MIGA is available, modifications are performed to turn it into a MICGA. Regarding the initial population generation, the categorical part of the genomes is produced using full factorial techniques with as many levels as members in the categorical variable. Then, the resulting sampling plan in the categorical dimension is repeated to obtain the desired number of sample points for each combination of categories. It assures that all categories are present in the initial population evenly, i.e., similar number of initial population members for all points in the non-numeric subset domain, denoted by $D_1 \times D_2 \times ... \times D_{n_n}$. D_i represents the set of possible members for the non-numeric variable i. The choice of full factorial techniques for categorical inputs is made due to the lack of proper distance between members of the same categorical variable. It is also made sure that the size of the population is bigger than the number of members in the non-numeric subset $|D_1 \times D_2 \times ... \times D_{n_n}|$. $|A|$ represents the number of members in the set A.

Nominal values are scatteredly crossovered because intermediate points can not be defined when there is no order between points. Mutation-wise, uniform distributions are utilized, which agrees with the Hamming distance: the distance between all categories is the same.

3.6.3 Mutation Study

Once the MICGA has been built, a possible spin-off question is the following: is there any categorical mutation operator that outperforms others for the optimization of the ExI functions?

Mutation is the process that genetically alters chromosomes. ESs use mutations as search operators to look for a better solution. An usual method to implement mutation operators involves generating a random variable for each chromosome. This random variable determines which value the chromosome will take after the mutation process. The ES user can modify the mutation operator by defining a mutation probability over the possible set of values.

Authors have implemented numerous mutation strategies for continuous and discrete-quantitative design variables [44, 69]. Popular choices are the normal and geometrical distributions for continuous and integer variables, respectively. These last mutation distributions are obtained from "Principle of Maximum Entropy" [159, 127]. Uniform distributions for the non-numeric variables are commonly employed.

Several mutation techniques are developed herein for categorical variables. It enables the study of the influence of several mutation operators in the performance of the MICGA when optimizing the ExI over a MIC domain. The mutation distributions to study are:

1) Uniform Distribution. When a non-numeric variable is to mutate, the likelihood of the child value is uniformly distributed, i.e., all the other categories have equal chance to appear in the genome of the child.[3]

2) Inverse Distance Distribution. The nominal part of the child genome follows a distribution inversely proportional to the intrinsic distances between the father

[3]This technique could be considered a maximum entropy technique for categorical variables whose metric is given by the Hamming distance.

categorical member and the remaining members.

3) Maximum Entropy. The nominal part of the child genome follows the maximum entropy distribution given the intrinsic distances between the father categorical member and the remaining members.

The concept of neighborhood between nominal points could be laid out thanks to the intrinsic distance of some non-numeric variables, as explained in Subsection 3.3.4. In this case, distances from one point to others in the categorical dimension is given by a discrete set of real numbers. The second and third mutation distributions of the above enumeration, "Inverse Distance Distribution" and "Maximum Entropy", are based on this discrete set of real numbers. In the following, a discussion about these two mutation distributions is presented. Finally, a discussion is presented regarding the performance indicators while testing the mutation distributions of all types of variables on functions with MIC domains.

3.6.3.1 Inverse Distance Distribution

It represents a simple and straight-forward method to build a mutation operator based on the intrinsic distance of a categorical set given a objective function.

Without loss of generality a problem with only one categorical variable is supposed. Let $\mathbf{p}^j$ be the probability mutation distribution for a parent whose nominal chromosome takes the j-th value of the n_d possible chromosome values, i.e., the non-numeric variable can take on n_d categories and the parent takes the j-th category. Notice that the super-index j emphasizes the fact that each discrete point in the categorical dimension has a different mutation distribution. Also, let p_k^j be the probability that the mutation child gets the k-th category given the parent takes the j-th category.

For the parental category j, the inverse distance distribution assigns mutation

probabilities, p_k^j, proportionally to the distance between the j-th and k-th categories

$$p_k^j \propto \frac{1}{d_{j,k}}$$

where $d_{j,k}$ represents the intrinsic distance between the j-th and k-th categories in the given categorical variable.

Normalization is required for probability distributions[4]. The normalization condition is given by Equation 89

$$\sum_{k=1}^{n_d \setminus j} p_k^j = 1 \tag{89}$$

Taking into account the normalization condition, the probability distribution for a parent that belongs to the j-th category could be written as

$$p_k^j = \frac{1}{\sum_{l=1}^{n_d \setminus j} \frac{1}{d_{j,l}}} \frac{1}{d_{j,k}} \qquad \text{for } k = 1...n_d \setminus j \qquad p_j^j = 0 \tag{90}$$

This method clearly makes more likely mutations to categories k that are close to the parental category j. It represents a simple way to build the mutation distribution. However, there exist more sophisticated mutation distributions.

3.6.3.2 Maximum Entropy Distribution

As mentioned in Subsection 2.8.3, Rudolph proposes that the mutation distribution should maximize the entropy if no additional knowledge about the objective function is known [159].

As is well-known from statistical thermodynamics and information theory, the entropy definition for a discrete probabilistic distribution $\mathbf{p}^j$ is

$$H\left(\mathbf{p}^j\right) = -\sum_{k=1}^{n_d \setminus j} p_k^j \log\left(p_k^j\right) \qquad \text{for } j = 1...n_d \tag{91}$$

[4]It is necessary for a probability distribution that the sum of the probabilities over all the possible states is equal to one.

where the index k goes through all the categorical points different from j.

As a probability distribution, normalization is again required. Also, the mean of the $\mathbf{p}^j$ distribution, μ^j, is to be specified to control the shape of the distribution. These two conditions are written as follows

$$\begin{aligned} \sum_{k=1}^{n_d \backslash j} p_k^j &= 1 \qquad \text{for } j = 1...n_d \\ \sum_{k=1}^{n_d \backslash j} d_{j,k} p_k^j &= \mu^j \qquad \text{for } j = 1...n_d \end{aligned} \tag{92}$$

The values $d_{j,k}$ can be better arranged by performing a translation, setting the minimum distance to other categories to zero,

$$\hat{d}_{j,k} = d_{j,k} - \min_k \left(d_{j,k} \right)$$

where $\min_k$ is the minimum distance to the category j (in this minimization the value of $k = j$ is excluded). Notice that the entropy function, see Equation 91, is invariable to translation in the sample space, which is the set of all possible outcomes for the random process. It is because the entropy definition only involves the probability and not the sample space as Equation 91 shows. Only the value of the mean gets affected.

The maximization of the entropy leads to solve a nonlinear constrained optimization problems, one per category $j = 1, 2, ...n_d$. Constraints are treated with the help of Lagrangian multipliers. For a given j, the optimal values can be obtained by partial differentiation of the objective function. The entropy, augmented with the constraints given by Equations 92, can be written as

$$L\left(p_k^j, \lambda_1^j, \lambda_2^j\right) = -\sum_{k=1}^{n_d \backslash j} p_k^j \log\left(p_k^j\right) + \lambda_1^j \left(\sum_{k=1}^{n_d \backslash j} p_k^j - 1 \right) + \lambda_2^j \left(\sum_{k=1}^{n_d \backslash j} \hat{d}_{j,k} p_k^j - \mu^j \right) \qquad \text{for } j = 1...n_d \tag{93}$$

where λ_1^j and λ_2^j are the Lagrange multipliers for the normalization, and the mean μ^j

constraints, respectively. Notice that there is no coupling of the mutation probability distribution between different parental points j of the categorical design variable.

The main details of the derivation are omitted for compactness' sake. Finally, a non-analytically tractable equation is found.

$$\mu^j = \frac{\sum_{k=1}^{n_d \setminus j} \hat{d}_{j,k} \exp\left(\lambda_2^j \hat{d}_{j,k}\right)}{\sum_{k=1}^{n_d \setminus j} \exp\left(\lambda_2^j \hat{d}_{j,k}\right)} \qquad \text{for } j = 1...n_d \tag{94}$$

If μ^j is chosen, Equation 94 is solved numerically for λ_2^j. Once λ_2^j is obtained, the other Lagrange multiplier λ_1^j and the probability distribution $\mathbf{p}^j$ for the parental category j are

$$\lambda_1^j = 1 - \log\left(\sum_{k=1}^{n_d \setminus j} \exp\left(\lambda_2^j \hat{d}_{j,k}\right)\right) \qquad \text{for } j = 1...n_d \tag{95}$$

$$p_k^j = \exp(-1 + \lambda_1^j + \lambda_2^j \hat{d}_{j,k}) \qquad \text{for } k = 1...n_d \setminus j \qquad \text{for } j = 1...n_d \tag{96}$$

It is worth realizing that, for each $\mathbf{p}^j$, the mean parameter μ^j is still free. Several remarks should be stated regarding the choice of this last parameter.

1) In order to guarantee that the mutation probability distribution, p_k^j, decays with d_k^j, the maximum value of the distribution mean, μ^j, has to be smaller than the center of mass of the translated distance set to the j-th category, $\hat{d}_{j,k}$,

$$\mu_{\text{limit}}^j = \frac{1}{n_d - 1} \sum_{k=1}^{n_d \setminus j} \hat{d}_{j,k}$$

2) The parameter μ^j controls the shape and mean of the distribution, smaller values provides a mutation distribution that highly benefits categories close to the j-th categorical member.

With the previous remarks being said, the mutation distribution is controlled by a mutation parameter, $\nu_{\text{mut}} \in [0, 1]$, that sets the value of the probability distribution mean as follows,

$$\mu^j = \nu_{\text{mut}} \mu^j_{\text{limit}} \tag{97}$$

Therefore, mutation distributions that are wished to highly benefit categories close to the parental one have values of ν_{mut} close to 0, whereas those intended not to take into account the distance between categorical points have values of ν_{mut} close to 1. The limit $\nu_{\text{mut}} \longrightarrow 1$ represents an uniform mutation probability distribution; the same one obtained with the Hamming distance. It is convenient for non-numeric variables where there is no a clear intrinsic distance between values.

3.6.3.3 Performance Indicators Mutation Study

Interest is in the combined mutation strategies for continuous, discrete-quantitative, and categorical variables. Two performance indicators measure the success of the method. The first one is the error of the maximization process given by

$$error = |\log_{10}\left(max\left(E\left[I\left(\mathbf{x}\right)\right]\right)_{est}\right) - \log_{10}\left(max\left(E\left[I\left(\mathbf{x}\right)\right]\right)_{real}\right)| \tag{98}$$

where the sub-indexes $\{.\}_{est}$ and $\{.\}_{real}$ represent the estimated and real value of the maximization process, respectively.

The error is given in terms of the negative $\log_{10}$ of the maximum $E\left[I\left(\mathbf{x}\right)\right]$. The second indicator is the generation at which the last successful improvement larger than 1% is produced in the optimization process. These parameters indicate the accuracy of the solution and the efficiency of each mutation strategy in the optimization process, respectively.

The mean and standard deviation of the two indicators are assessed because of the statistical nature of the GA outputs.

3.7 Fenestron Configuration as the New Concept

The choice of the fenestron tail as the new concept to test the ECMF surrogate is made due to several constraints on the author. A possible new concept is a coaxial rotor. However, the author could not find a coaxial control routine for rotors in FLIGHTLAB, the commercial software employed to model the helicopter. The remaining options are to include a weapon as a external load in the basic UH60A model or to substitute the conventional tail with a fenestron one.

Both possibilities have some drawbacks. The first option could be seen more a new flight condition or mission for the helicopter than a new concept. The second option is a new concept, but the UH60A is heavier than the largest helicopter that has been equipped with the shrouded fan. On the other side, the fenestron tail option allows the study of the trade-off between the increase in weight due to the new tail configuration and the increase in anti-torque device efficiency because of the ducted fan.

The fan-in-fin tail is chosen as the new concept for two reasons. First, it is interesting and practical to study the previously mentioned trade-off between the weight increase and the enhancement of tail rotor efficiency[123]. The second reason is the expected proximity between the old concept (UH60A with conventional tail) and the new concept (UH60A with fenestron tail); it is reasonable to think that they would experience similar trends.

The main challenge on building a UH60A with fenestron tail FLIGHTLAB model is the estimation of the fenestron baseline parameters and its overall weight. These estimations are assessed in the following Subsections.

3.7.1 Fenestron Baseline Values

Rand et al. [150] perform a statistical analysis on a helicopter database. The aim of the study is to estimate geometry parameters, weight of components, preliminary

power, and flight performance. Among the targeted geometry parameters, one finds the rotor diameter, rotor angular speed, horizontal tail surface area, and vertical tail average chord. The useful part of Rand et al.'s research to the present work is that in the database some helicopters have a fenestron configuration. Thus, the regression fits can be used to estimate the fenestron baseline parameters.

The desired fenestron parameters to assess are: R_{tr}^{fen}, b_{vt}^{fen}, c_{vt}^{fen}, Ω_{tr}^{fen}, $N_{b,tr}^{fen}$, and c_{tr}^{fen}. Rand et al. [150] find that the fenestron rotor diameter is correlated with the gross weight W_0 as follows

$$D_{tr}^{Fen} = 0.3081 W_0^{0.154} \tag{99}$$

The resulting fenestron diameter is 1.2852 m, where $W_0 = 10660$ kg is assumed. This value is too small compared with Super Puma SA330, one of the heaviest helicopter with fenestron. Super Puma diameter is 1.6 m with a gross weight of $W_0 = 7000$ kg. Therefore, a more reliable estimation, according to the author, could be assessed by a parallel fit to the one in Equation 99 that passes through the Super Puma point ($D_{tr}^{Fen} = 1.6$ m and $W_0 = 7000$ kg). The resulting UH60A diameter for the fenestron tail configuration is $D_{tr}^{Fen} = 1.7071$ m.

The addition of the fenestron tail usually results in an increase in the vertical tail span b_{vt}^{fen}. No correlation has been found, but herein it is assumed that the vertical tail span is increased one fenestron radius with respect to that of the UH60A with conventional tail. So, for the baseline case one obtains $b_{vt}^{fen} = b_{vt}^{Conv} + \dfrac{D_{tr}^{Fen}}{2} = 3.14$ m.

The correlation found by Rand et al. [150] for the fenestron vertical tail average chord is

$$c_{vt}^{fen} = 0.909 D_{tr}^{fen0.927}$$

that results in $c_{vt}^{fen} = 1.4923$ m, which is 13.6% larger than that of the UH60A with the conventional tail.

Regarding the fenestron angular speed, Rand et al.'s study does not provide a statistical fitting for the fenestron case. However, the two fenestron helicopters seem to validate the regressive fit calculated for the angular velocity of the conventional tail rotor-crafts. Therefore, the conventional tail angular velocity correlation is applied to estimate the fenestron angular velocity

$$\Omega^{tr} = \frac{364}{{D_{tr}^{fen}}^{0.828}}$$

that leads to a rotor angular speed of $\Omega_{tr}^{fen} = 233.77\ rad/s$.

Neither is a sizing correlation for the number of fenestron blades found. Normally, the number of blades go from 7 to 14. In this study it is assumed the lowest range, i.e., from 7 to 9 fenestron blades, being $N_{b,tr}^{fen} = 8$ the baseline value for the fenestron case.

The tail rotor blade chord correlation presented by Rand et al. [150] does not include fenestron information either. However, it is used in the estimation of c_{tr}^{fen} because it is assessed for a situation that also applies to fenestron rotors: the ability of the tail rotor to balance the torque of the main rotor in a full power vertical climb at a specified altitude [150]. The correlation formula is

$$c_{tr} = 0.0058 \frac{W_0^{0.506}}{N_{b,tr}^{0.72}}$$

using the previous values of $N_{b,tr}^{fen} = 8$ and $W_0 = 10660$ kg, the resulting chord is $c_{tr}^{fen} = 0.1417$ m.

Table 1 shows the baseline values for the UH60A with fenestron tail rotor.

3.7.2 Weight Estimation

There exist several models for rotor-craft weight estimation. Among them, one can find investigations on weight estimation for preliminary sizing of rotor-craft by Prouty [146], Rand et al. [150], Tishchenko et al. [185], NASA design and analysis of rotor-craft (NDARC) [97], and Kalra et al. [102]. These references provide formulas for the

Table 1: Fenestron Baseline Values

Fenestron Baseline Parameters	Values
R_{tr}^{fen}	1.7071 m
b_{vt}^{fen}	3.140 m
c_{vt}^{fen}	1.4923 m
Ω_{tr}^{fen}	2333.77 rad/s
$N_{b,tr}^{fen}$	8
c_{tr}^{fen}	0.1417 m

weight estimation of several parts of the rotor-craft in terms of some basic preliminary information, such as empty weight, rotor diameter, number of blades, ...

However, not much sizing information is available for the fenestron weight estimation; the main reference for this purpose is Rand et al.'s work [150]. Nevertheless, the fenestron data is for rotor-craft gross weights up to 8000 kg. The UH60A configuration is heavier than the information in the database, but it is assumed that the statistical correlations and weight estimation formulas still apply.

The change in rotor-craft weight due to the fenestron is broken down in 3 main contributions:

1) Vertical fin weight

2) Blades weight

3) Hub weight

In the above three contributions, the same procedure is used to calculate the change in weight of the fenestron tail with respect to the conventional case. The item contributions of the conventional tail can be known or estimated based on the basic UH60A information. Also, the item contributions of the fenestron configuration are estimated with Rand et al.'s work[150].

3.7.2.1 Vertical Fin Weight Estimation

First, the weight of the conventional vertical tail, W_{vt}^{Conv}, is assessed with the formula provided by NDARC [97]

$$W_{vt}^{Conv} = 1.05 A_v^{0.94} AR_v^{0.53} N_{tr,gb}^{0.71} \tag{100}$$

where A_v is the vertical fin area, AR_v the vertical fin aspect ratio, and $N_{tr,gb}$ the number of tail rotor gearboxes.

The resulted vertical tail weight for the conventional UH60A obtained is $W_{vt}^{Conv} =$ 61.0287 lb. With the Prouty's equivalent formula one gets a similar weight $W_{vt}^{Conv} =$ 60.4474 lb.

The next step is to estimate the vertical fin weight for the fenestron configuration. The fenestron diameter and vertical fin chord, estimated in Subsection 3.7.1, are brought to the weight estimation process: $D_{tr}^{Fen} = 1.7071$ m and $c_{vt}^{Fen} = 1.4923$ m. It is convenient to remember that the vertical fin chord is around 13.6% larger than that of the conventional tail configuration, i.e., $\Delta c_{vt} = 13.6\%$.

Finally, the estimation of the fenestron vertical fin is done assuming that the relative thickness of the fin is kept constant. Also, as previously explained, the length of the fenestron configuration fin, b_{vt}^{Fen}, is one fan radius longer than the UH60A conventional tail, $b_{vt}^{Conv} = 7.5$ ft

$$W_{vt}^{Fen} = W_{vt}^{Conv} \Delta c_{vt}^2 \frac{b_{vt}^{Conv} + 0.5 D_{tr}^{Fen}}{b_{vt}^{Conv}} \tag{101}$$

W_{vt}^{Fen}, given by Equation 101, is in general a function of the value of D_{tr}^{Fen}. For the baseline case ($D_{tr}^{Fen} = 1.7071$ m), its value is $W_{vt}^{Fen} = 49.047$ kg.

3.7.2.2 Blades Weight Estimation

Kalra et al. [102] propose a detailed equation for the weight of main rotor blades that is used in this work for the conventional and fenestron tail rotor blade weight

estimation, $W_{tr,bl}$.

$$W_{tr,bl} = k_{tr,bl} \left(\frac{N_b}{4} \right)^{0.5348} \frac{\sigma_{sol} R^{2.7}}{\left(\frac{AR}{18} \right)^{0.7}} \tag{102}$$

where σ_{sol} is the rotor solidity, and $k_{tr,bl}$ is a constant of value 15.

Applying Equation 102 to the conventional tail configuration results in a $W_{tr,bl}^{Conv} =$ 22.45 kg ($W_{tr,1bl}^{Conv} = 5.6125$ kg per blade). Regarding the fenestron case, the value of $W_{tr,bl}^{Fen}$ is determined by fenestron parameters, specifically D_{tr}^{Fen}, c_{tr}^{Fen}, and $N_{b,tr}^{Fen}$. For the baseline case (parameters are given by Table 1), the resulting weight is $W_{tr,bl}^{Conv} =$ 21.529 kg.

3.7.2.3 Hub Weight Estimation

Tishchenko et al. [185] propose the following tail rotor hub weight estimation $W_{tr,hub}$

$$W_{tr,hub} = k_{tr,hub} N_{b,tr} f_{z,tr,bl} N_{cf,tr,bl}^{1.35} \tag{103}$$

where $k_{tr,hub}$ is a constant weight coefficient for the tail rotor hub equals to 0.5, $N_{b,tr}$ the number of tail rotor blades, $f_{z,tr,bl}$ a factor to account the influence of the number of blades in the hub weight, and $N_{cf,tr,bl}$ the centrifugal force on the tail rotor blades.

The value of $N_{cf,tr,bl}$ is assessed by the integration of the centrifugal force along the blade

$$N_{cf,tr,bl} = \int_{r_{cut}}^{R_{tr}} \rho_{tr,bl} \Omega^2 r dr$$

where $\rho_{tr,bl} = \dfrac{W_{tr,1bl}}{R_{tr}}$ is the blade weight per unit of length, and r_{cut} the root cut-off, which is set to $\frac{1}{3}$ for the fenestron case in this work.

The value of $f_{z,tr,bl}$ is to be determined. The helicopter example provided in Appendix A in Prouty's work [146] is used to estimate its value. The parameter values of this helicopter example are $N_{b,tr}^{Prou} = 3$, $c_{tr}^{Prou} = 1$ ft, $\Omega_{tr}^{Prou} = 100$ rad/sec, and $R_{tr}^{Prou} = 6.5$ ft. In order to calculate the blade weight, Equation 102 is used, which provides $W_{tr,bl}^{Prou} = 24.4$ kg.

It is known that the total weight of the whole tail rotor assembly in the Prouty's example is 82.1 kg. Thus, the difference between the rotor assembly and the blade weight is the weight of the tail rotor hub, which is $W_{tr,hub}^{Prou} = 57.7$ kg. Now all the factors in Equation 103 are known except $f_{z,tr,bl}$, so one can solve for it, obtaining a value of $f_{z,tr,bl} = 9.15 \cdot 10^{-6}$.

Once the value of $f_{z,tr,bl}$ is obtained, Equation 103 provides $W_{tr,hub}^{Conv} = 67.36$ kg. Also, Equation 103 is employed to obtain $W_{tr,hub}^{Fen}$; it is a function of D_{tr}^{Fen}, c_{tr}^{Fen}, and $N_{b,tr}^{Fen}$. For the baseline case the resulting weight is $W_{tr,hub}^{Conv} = 90.058$ kg.

Finally, the total change in weight, ΔW, due to the fenestron tail is given by Equation 104

$$\Delta W = W_{vt}^{Fen} - W_{vt}^{Conv} + W_{tr,bl}^{Fen} - W_{tr,bl}^{Conv} + W_{tr,hub}^{Fen} - W_{tr,hub}^{Conv} \tag{104}$$

which in general is a function of the fenestron design variables. When particularized to the the baseline case, its value is $\Delta W = 43.143$ kg.

3.8 Research Methodology Diagrams

Once the methods and the desired studies have been explained, it is helpful to plot the adaptive sampling algorithms, that include the developed surrogates as well, for categories and concepts. Adaptive sampling on MIC and ECMF surrogates are shown in Figures 11, and 12, respectively. These Figures also contain the studies on the influence of training set sizes, nominal distances, and mutation strategies.

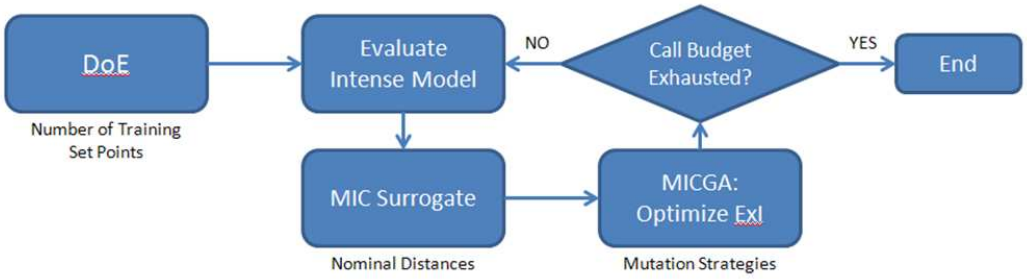


Figure 11: Diagram of Adaptive Sampling on MIC Surrogates

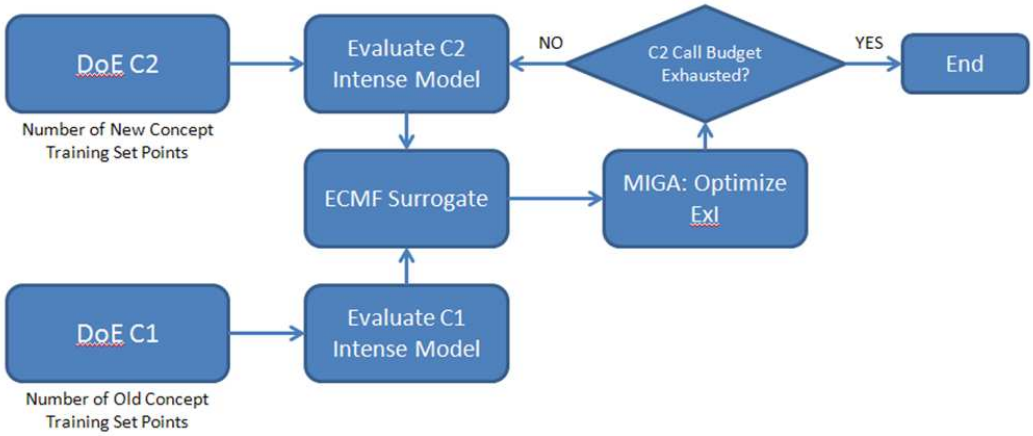


Figure 12: Diagram of Adaptive Sampling on ECMF Surrogates

Figures 13 and 14 show the whole methodology diagrams for the categories and concepts, respectively. They contain the motivation, technical challenges, research questions, hypothesis, and experiments to achieve the purpose of this thesis.

Figure 13: MIC Methodology Review

Figure 14: ECMF Methodology Review

CHAPTER IV

FLIGHTLAB UH60A MODEL

A practical computational model is pursued to measure in the following Chapters the efficiency of the proposed surrogates, MIC and ECMF. They are intended as tools in conceptual design of engineering systems where there exists a choice of discrete alternatives. The targeted application is helicopter design.

Also, the proposed methods are aimed to computationally expensive functions. A possible example of these functions is the performance of a helicopter; the assessment of the performance requires to combine several disciplines such as aerodynamics, structural mechanics, and controls. Thus, even mid-fidelity codes for each discipline produce intense function calls due to the necessary iterations between disciplines. The author has no access to the mid and high-fidelity commercial software for helicopters RCAS or OVERFLOW. However, the software FLIGHTLAB [1] is available.

Typical performance measures used in conceptual design of helicopters are the engine shaft horsepower in hover and forward flight. Herein, these two measures are the objective functions to optimize the helicopter design.

The choice of the helicopter to model has to satisfy two conditions: availability of a reliable FLIGHTLAB model and some experimental information to validate the model. These restrictions lead the author to pick the UH60A as the baseline helicopter.

In the remaining part of the Chapter, a review of the UH60A parameters is presented. Followed by the validation of the hover and forward flight models. The noisy nature of the FLIGHTLAB output is discussed. Finally, the FLIGHTLAB fenestron modeling is presented.

Table 2: Parameters of the UH60A Baseline Model [23]

$W = 17252.17$ lb
$Vel_{hov} = 0$ knots
$Vel_{for} = 139$ knots (Adv. Ratio = 0.3218)
UH60A Fuselage
Position$_{fus,c.g.}$[1]= $[345.5, 234.0, 0]$ in
2 Horizontal Stabilizers
Surface Each Hor. Stab. = 22.5 ft^2
$AR = 4.6$
Airfoils=NACA 0014
Position$_{attach,hs} = [700.1, 244, 0]$ in
Surface Vert. Stab. = 32.3 ft^2
$AR = 1.92$
Airfoils=NACA 0021
Position$_{attach,vs} = [695, 273, 0]$ in
T700-GE-700 Engine
$N_{eng} = 2$
Nominal Engine Torque = 355 $lbf \cdot ft$
Main Rotor to Engine Gear Ratio = 0.012336
UH60A Landing Gear

[1]Position = {Fuselage, Waterline, Buttline}

4.1 UH60A Parameters

The FLIGHTLAB commercial package includes a example helicopter which is really close to the UH60A helicopter. Taking a look at the literature, one can find experimental data on UH60A hover and forward flight performance, see Bousman and Kufeld [23], Yeo et al. [205] Lawrence et al. [121], and Shinoda et al. [168]. The baseline helicopter and rotor configurations are given in Tables 2, and 3, respectively.

4.2 Hover Model Validation

The hover model consists of:

1) Main Rotor. Articulated rigid rotor. Flapping and lead-lag dynamics are included. The aerodynamics is brought by a quasi-steady aerodynamic model with look-up tables for the blade sections ("SC 1095" and "SC 1094R8"). Aerodynamic tables obtained from JAVAFOIL[84], assuming the Re of the 75% of the

Table 3: Parameters of the UH60A Baseline Rotors [23]

$R = 26.83$ ft
$\Omega = 27$ rad/sec
$N_b = 4$
$\bar{c} = 1.73$ ft. Rectangular blade
Blade Twist of the UH60A Blade
Airfoils = {SC 1095, SC 1094R8, SC 1095}
Airfoil boundary = $[0.0; 0.485; 0.835; 1.0]\, R$
Hinge Offset = 1.25 ft
Rotor Position = $[342.215, 315, 0]$ in
Weight One Blade = 256.91 lb
Blade Moment Inertia About Hinge = 1512.6 $slug \cdot ft^2$
Longitudinal Shaft Tilt = -3 degrees
Articulated rotor: Flap and lead-lag hinge
$R_{tr} = 5.5$ ft
$\Omega_{tr} = 124.62$ rad/sec
$N_{b,tr} = 4$
$\bar{c}_{tr} = 0.81$ ft. Rectangular blade
$\theta_{tr} = -18$ degrees
$c_{l,\alpha,tr} = 5.73$
$c_{d0,tr} = 0.0087$
$c_{d1,tr} = -0.0216$
$c_{d2,tr} = 0.4$
Tail Rotor Position = $[732, 324.7, -14]$ in

blade span at hover. A blade tip loss factor of 0.97 is assumed. The Peters-He three states inflow computes the induced velocity. Interference between main rotor and fuselage is allowed. Regarding discretization, there are a set of 20 aerodynamic segments and another of 6 structural segments, which are both distributed by an equal annuli area.

2) Tail Rotor. It is a Bailey rotor with blade tip loss factor of 0.92. It is modeled with a quadratic airfoil drag polar and a linear lift curve.

3) A rigid fuselage with empirical airloads.

4) Aerodynamic surfaces. Rigid surfaces with aerodynamic look-up tables for the surface section.

5) Propulsion. Two ideal engines without power losses.

6) Control. Standard flight control unit with a longitudinal, lateral and collective stick, and pedals. Also, two attitudes as pseudo-controls.

This model is compared with the experimental data from "UH60A Airloads Program", specifically the "UH60A Airloads Catalog" by Bousman and Kufeld [23]. They measured the UH60A hover performance in terms of the total helicopter C_P for several C_W. Figure 15 shows the "UH60A Airloads Program" experimental results and the corresponding FLIGHTLAB model simulations. The FLIGHTLAB model does not take into account power losses like transmission, interference between rotors, electrical or operation losses; thus, the FLIGHTLAB model has been corrected with a helicopter efficiency, η, of 90%.

According to the results in Figure 15, the UH60A model represents decently the hover performance of the UH60A.

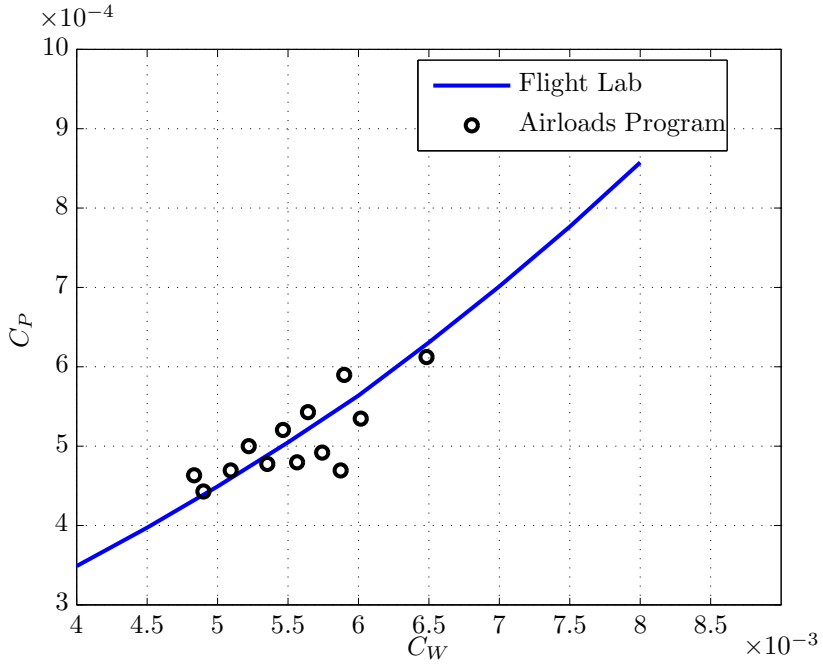


Figure 15: Total C_P vs C_W. Validation of UH60A Hover Shaft Power. FLIGHTLAB Model

4.2.1 Fenestron Modeling in FLIGHTLAB

FLIGHTLAB features a ducted fan airload model [1]. It is derived from momentum theory combined with blade element analysis. Enhancement features are present in the model, such as duct-fan interference, and fan wake contraction. These enhancements require the use of empirical corrections obtained either by experimental measurements or advance CFD. However, the author does not have access to these data; therefore, the ducted fan is kept in the most simple way: the fan wake contraction and the duct-fan interference are considered negligible. A proper duct lip design for specific flight conditions can diminish the fan wake contraction effect.

4.3 Forward Model Validation

The forward model consists of:

1) Main Rotor. Articulated rigid rotor. Flapping and lead-lag dynamics are included. The aerodynamics is brought by a quasi-steady aerodynamic model with look-up tables for the blade section ("SC 1095" and "SC 1094R8"). Aerodynamic tables obtained from JAVAFOIL[84], assuming the Re of the 75% of the blade span at hover. A blade tip loss factor of 0.97 is assumed. The inflow model is a Glauert one with a nonuniform correction of 15%. No interference between main rotor and fuselage is allowed. Regarding discretization there are a set of 20 aerodynamic segments and another of 6 structural segments, which are both distributed by an equal annuli area.

2) Tail Rotor. It is a Bailey rotor with blade tip loss factor of 0.92. It is modeled with a quadratic airfoil drag polar and a linear lift curve.

3) A rigid fuselage with empirical airloads.

4) Aerodynamic surfaces. Rigid surfaces with aerodynamic look-up tables for the surface section.

5) Propulsion. Two ideal engines without power losses.

6) Control. Standard flight control unit with a longitudinal, lateral and collective stick, and pedals. Also, two attitudes as pseudo-controls.

Once again, the model is compared with the experimental data from "UH60A Airloads Program", specifically the "UH60A Airloads Catalog" by Bousman and Kufeld [23]. They measured the UH60A forward flight performance in terms of the total helicopter C_P vs advance ratio for several C_W. In this work the weight coefficient chosen to validate the forward flight model is $C_W = 0.0074$. Figure 16 shows both

the "UH60A Airloads Program" experimental results and the FLIGHTLAB model simulations. As in the hover case, see Section 4.2, the FLIGHTLAB model does not take into account power losses. The losses are the same as in hover, but the rotors interference is supposed to be larger for forward flight.

The FLIGHTLAB model has been corrected with a helicopter efficiency, η, of 83%. It could be thought as a slightly low helicopter efficiency; however, it is convenient to mimic the experiment measures from the "UH60A Airloads Program". Figure 16 shows some confidence in the UH60A forward flight FLIGHTLAB model.

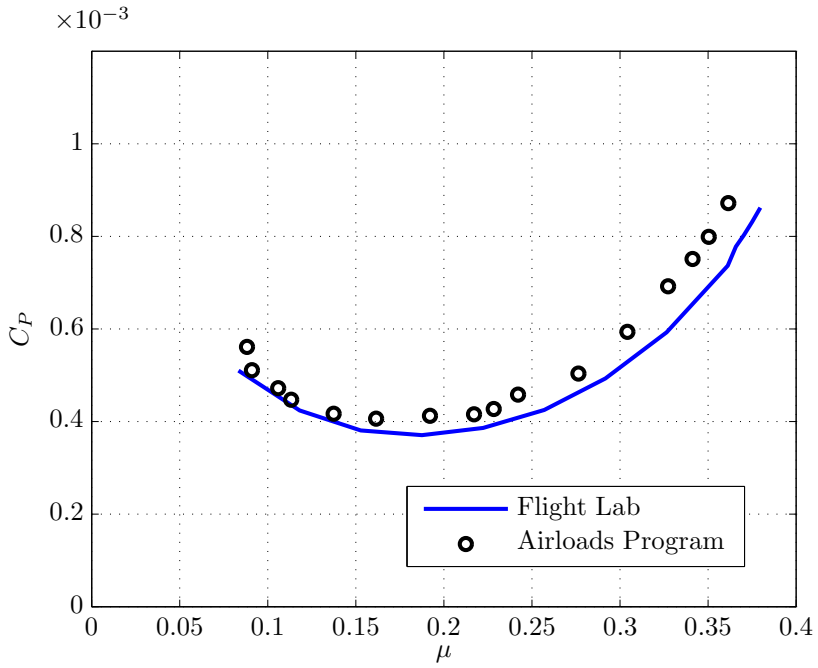


Figure 16: Total C_P vs Advance Ratio. Validation of UH60A Forward Flight Shaft Power. FLIGHTLAB Model

Notice that this forward flight validation was carried out with a weight different from the baseline, see Table 2. It is done because no experimental data was found for the exact baseline weight.

Also, the fenestron modeling in forward flight is similar as the one in the hover case, see Subsection 4.2.1.

4.4 UH60A Power Consumption as a Noisy Function

It is well known that numerical solvers require discretization and iterative methods. The resulting solutions are noisy, which represents a problem for design teams when exploring and exploiting objective functions [66]. The noisy nature of FLIGHTLAB solutions are shown in this Section.

In order to demonstrate the existence of noise in the UH60A hover power consumption, the landscape of the function is shown in some part of the domain space, specifically along the main rotor linear twist design variables: θ_1 and θ_2. The function is computationally expensive, so a detailed sweep in a large design space is not practical. Thus, a reduced design space is chosen to evaluate the hover power consumption. The reduced domain where the function is evaluated is:

$$\theta_1 = -3.5 + 3L^{\circ}_{x1}/m \qquad \theta_2 = -1.75 + 2.25L^{\circ}_{x2}/m$$

where $L_{x1} = L_{x2} = 10^{-2}$.

Evaluations of the UH60A hover power consumption are performed on a grid of this two-dimensional domain space. The samples on the grid are given by parameters $N_{sampl,x1} = 101$, and $N_{sampl,x2} = 101$. Figure 17 exhibits the contours of the function. It can be seen the noisy nature of the function.

In order to get a better insight of the noise of the FLIGHTLAB simulation, the fast Fourier transform (FFT) of the function, $FT\left(f_{x1}, f_{x2}\right)$, is carried out to see which frequencies are present in the hover power consumption. The values of f_{x1} and f_{x2} are given by the spacing of the evaluation grid.

$$f_{x1} = \frac{1}{\Delta x_1}\{0, 1, 2, ..., floor\left(\frac{N_{sampl,x1}}{2}\right) - 1\}$$
$$f_{x2} = \frac{1}{\Delta x_2}\{0, 1, 2, ..., floor\left(\frac{N_{sampl,x2}}{2}\right) - 1\}$$

where Δx_1 and Δx_2 are the separation between the grid evaluation points in the dimensions x_1 and x_2, respectively.

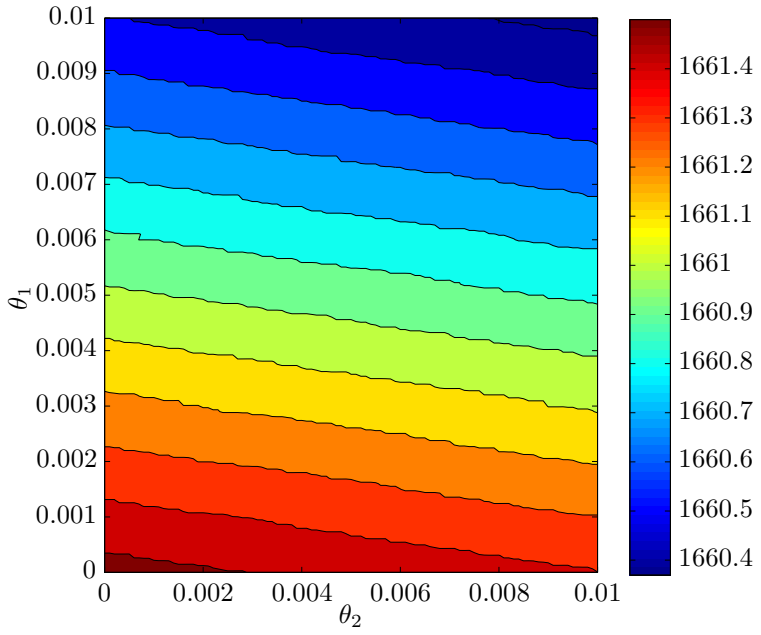


Figure 17: Noise on UH60A shp_{hov}. $L_{x1} = L_{x2} = 10^{-2}$

The amplitude of the FFT, $|FT\,(f_{x1}, f_{x2})\,|$, is plotted in Figure 18. It is seen that the constant term is the dominating one. However, when zooming in high frequency areas of the $FT\,(f_{x1}, f_{x2})$, non-zero values are seen, see Figure 19. Even though the values of the amplitude are much smaller than that of the zero frequency, the ExI criterion is expected to set points close to one other when a local optimum is found. Interpolating surrogates cannot deal with noisy observations. Therefore, this high frequency noise can result in flawed surrogates when the adaptive sampling process finds a local optimum.

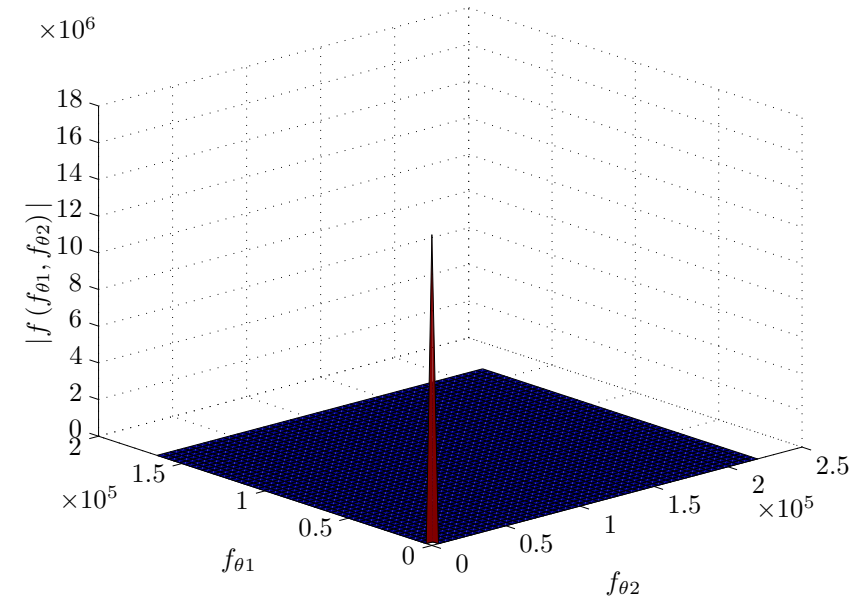


Figure 18: Fast Fourier Transform UH60A shp_{hov}

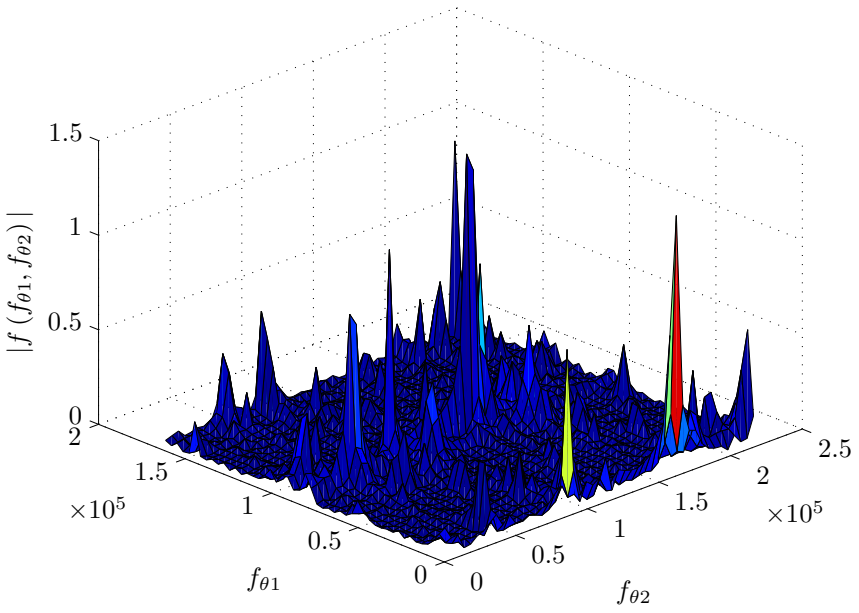


Figure 19: Fast Fourier Transform UH60A shp_{hov}. Close-Up to Large Frequency

CHAPTER V

MIC EXPERIMENTS: TRAINING SIZE AND NOMINAL DISTANCE

As mentioned in previous sections, one of the goals of this research is to develop surrogates that take advantage of similar trends across categorical alternatives. Two Kriging surrogates are brought in this Chapter:

MIC Surrogates They are developed to leverage similar trends across categories. They include categorical variables in its domain, so it hosts all kinds of design attributes: continuous, discrete-quantitative and non-numeric. The definition of a nominal distance is necessary to include the categorical variables in the surrogate domain. A study is carried out to see the influence of several nominal distances in the surrogate performance. However, the MIC surrogate that is compared against the current state-of-the-art is based on the Hamming distance.

Independent Surrogates It is the current state-of-the-art method. A surrogate is fitted independently for each category.

Both surrogates, the independent surrogate and the MIC surrogate, are compared to assess the MIC surrogate efficiency. The quality of these surrogates is expected to increase as the number of the training set size does. Two functions are tested: the noise-free and cheap disturbed Branin function; and the noisy and intense UH60A hover *shp*. As was mentioned in Section 1.7, tests on computationally expensive models will not help to better interpret the canonical test results or support the research hypothesis or predictions; however, it supports the practicality of the application.

The surrogate comparison is given in terms of the success indicators introduced in Subsection 3.4.1: rms of the "standardized validation error", and rms of the "standardized validation residual", see Equations 77 and 78, respectively.

5.1 Disturbed Branin Function

The first function tested is the disturbed Branin function. Some disturbances are applied on the first Fourier modes of the Branin function to obtain an objective function that has similar categorical trends. Also, the second variable of the original Branin function, x_2, is transformed into a discrete-quantitative one by discretizing it into 11 equally distanced discrete points. Appendix A provides a detailed explanation on the generation of the disturbed Branin function. The landscape of the function is plotted in Figure 20. For clarity's sake, the variable x_2 is shown as continuous, but it is necessary to remind the reader that it is a discrete-quantitative one whose possible values are $\left[0, \frac{1}{10}, \frac{2}{10}, ...1\right]$.

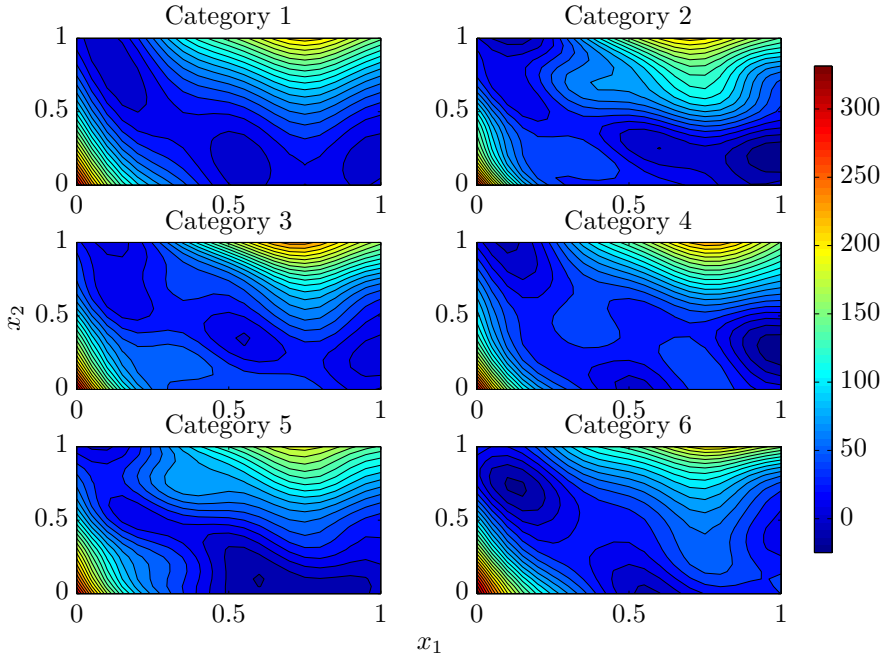


Figure 20: Contours of Disturbed Branin Function

5.1.1 Influence of Training Size on Performance

The two meta-models are compared in a range of training set sizes from 20 to 400 points. When building the independent surrogates, the training set is equally divided among the categories. Since both meta-models are probabilistic, two performance indicators are used for the comparison: rms of the surrogate "standardized validation error", and rms of surrogate "standard validation residual". A set of validation points are evaluated to assess the performance indicators.

Curves are fitted to try to understand the tendency of the results for each type of surrogate. For the rms of the "standardized surrogate error" the fitting curve chosen is of the form

$$\frac{B}{N_{tr.set}^{\alpha}} \tag{105}$$

it asymptotically tends to zero since the r.m.s. of the "standardized surrogate error" is expected to do so. The curve fit for the rms of the standardized surrogate residual is

$$A + \frac{B}{N_{tr.set}^{\alpha}} \tag{106}$$

it asymptotically tends to A since the r.m.s. of the "standardized surrogate residual" is expected to tend to a number, specifically to one. A, B, and α are constants.

Note that, as explained in Subsection 3.4.1, the value of the constant α determines how quickly performance indicators converge to their final value; thus, α enables the study of the convergence as the training set size increases.

The success indicator rms of the "standardized validation error" is plotted in Figure 21. It shows the comparison between the MIC Hamming surrogate versus the current state-of-the-art (one independent surrogate for category).

Figure 21 shows that the MIC Hamming surrogate results in more accurate meta-models than fitting independent surrogates for each category for the small range of sampling sizes. The reason is that, in the case of low number of points, only a few

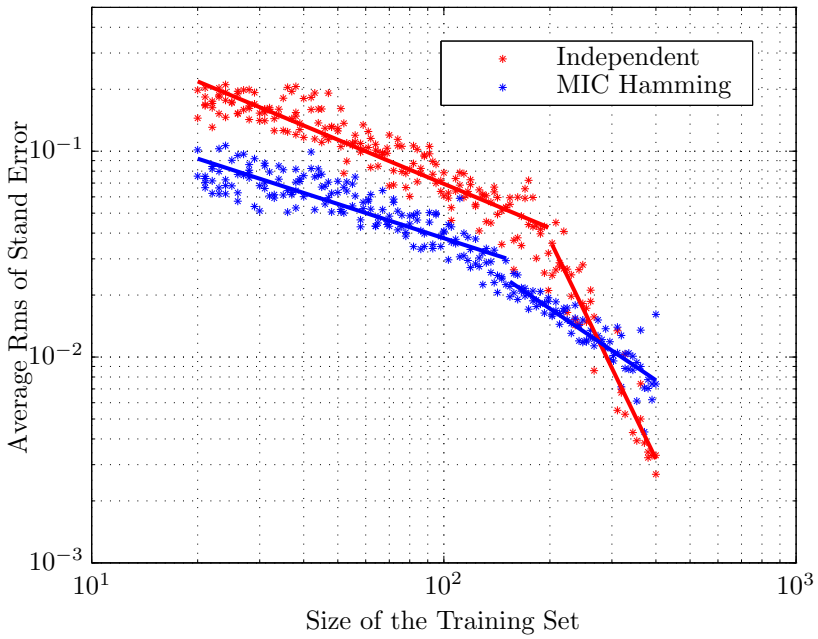


Figure 21: Surrogate "Standardized Validation Error". MIC vs Independent Surrogate. Disturbed Branin Function

points are sampled for each category; thus, the rms of the "standardized validation error" for the independent surrogate has little information, resulting in a poorer performance than the MIC Hamming surrogate, where observations are cross-used across categories. Specifically, for 24 sample points and six categories, only four points are used for building the two-dimensional surrogate of each category.

For high training set sizes, independent surrogates gain performance up to a point where they produce a better surrogate than the MIC Hamming one. It implies that when the training set is large enough, the cross-use of observations matters less. This cross-use of observations at this high range of sample points are even disadvantageous, producing surrogates less accurate than the ones that do not cross-use observations. The reason could be that, in this large training set case, observations hold not only similar trend information but also high frequency information for each category which is different across categories. Figure 21 shows that, for training sets larger than 280 samples, the current state-of-the-art outperforms the proposed MIC surrogate.

Table 4: Values of Constant α for the Fitting Curves of Rms of the "Standardized Validation Error". Independent vs MIC Surrogate. Disturbed Branin Function

	Low $N_{tr.set}$	Large $N_{tr.set}$
Independent Surr.	0.7122	3.533
MIC Hamming Surr.	0.553	1.156

The gain in performance of the independent surrogate with respect to the MIC Hamming surrogate is quantified by the values of the constant α in the fitting curve of the rms of the surrogate error (See Equation 105). Table 4 contains the values of αs. It is seen that independent surrogates converge more quickly than the MIC Hamming ones. Also, it is seen that if two fitting curves are used (one for small and another for large training sets), then the gain in performance of the independent surrogate in large training sets is higher than in small ones.

Finally the "standard validation residual" is plotted in Figure 22 for the independent surrogate and the MIC Hamming one. The independent surrogate "standardized validation residual" is higher than one for small training sets, which implies that the uncertainty estimated from the surrogate is lower than the real error, i.e., the surrogate uncertainty is underestimated. However, the "standardized validation residual" for the MIC Hamming surrogate is around one for the whole studied range of training set sizes.

Results are similar to the ones obtained for the "standardized validation error": the MIC Hamming surrogate outperforms the state-of-the-art for small training set sizes while applied on the disturbed Branin function. For more details about the tested function, see Appendix A.

5.1.2 Influence of Nominal Distance on Performance

MIC surrogates can be based on several nominal distances. The influence of these distances in the performance of the MIC surrogate is studied in this Subsection. The key to include the categorical input into the meta-model is the definition of a distance

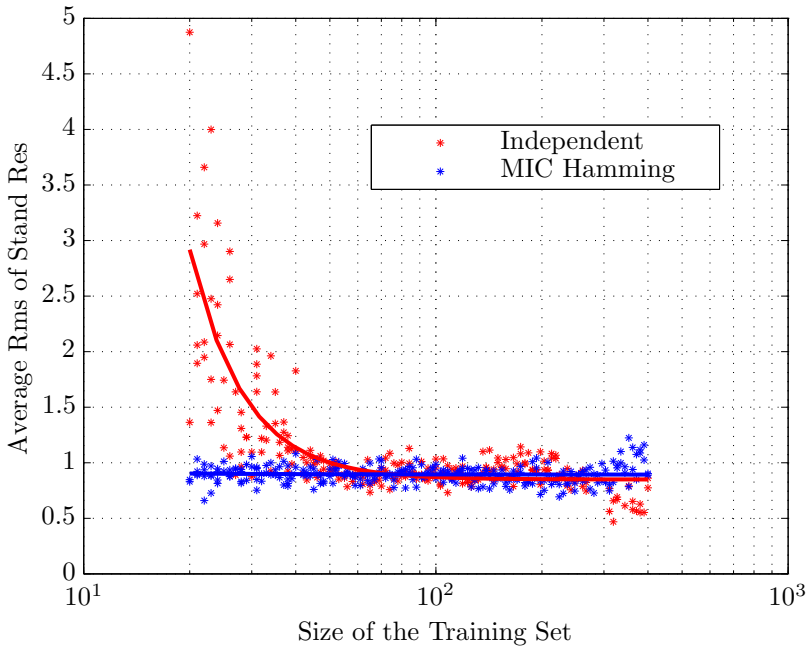


Figure 22: Surrogate "Standardized Validation Residual". MIC vs Independent Surrogate. Disturbed Branin Function

between categories that allows interpolating and regressing techniques, as is explained in Section 3.4.

Several distance functions were presented in Section 3.3 to include the categorical inputs into the Gaussian process meta-model. They were the following: integer, Hamming, and intrinsic. The intrinsic-based nominal distances for the disturbed Branin function are based on phases and amplitudes of the first nine Fourier modes of the disturbed Branin function. The possible nominal distances (four of them intrinsic-distances) are the following:

Hamming distance The Hamming distance is employed for the categorical variables.

Integer distance The nominal variable is arbitrarily mapped to a set of integer values. Each category of a non-numeric variable is mapped to an integer in the interval $[1, |x_{nom}|]$, where $|x_{nom}|$ is the number of points in the input x_{nom}. Then,

the distance brought to the Kriging surrogate is based on this new arrangement of x_{nom} in the integer line.

Phase-based distance 1 The distance between categorical members i and l is given by the phase difference in the first nine Fourier modes.

$$d\left(x_{nom}^{(i)}, x_{nom}^{(l)}\right) = \sum_{j=1}^{9} \left| \angle FT_{x_{nom}^{(i)}}\left(n_{1j}, n_{2j}\right) - \angle FT_{x_{nom}^{(l)}}\left(n_{1j}, n_{2j}\right) \right| \tag{107}$$

Then, distances given by Equation 107 are normalized to set the largest distance between categories to 1.

Phase-based distance 2 As "phase-based distance 1" but with an adjustment to set the minimum distance between categories, $d^{\star}$, to 0.5 according to Equation 74.

Amplitude-based distance 1 The distance between categorical members i and l is given by the amplitude difference in the first nine Fourier modes.

$$d\left(x_{nom}^{(i)}, x_{nom}^{(l)}\right) = \sum_{j=1}^{9} \left| \left| FT_{x_{nom}^{(i)}}\left(n_{1j}, n_{2j}\right) \right| - \left| FT_{x_{nom}^{(l)}}\left(n_{1j}, n_{2j}\right) \right| \right| \tag{108}$$

Then, distances given by Equation 108 are normalized to set the largest distance between categories to 1.

Amplitude-based distance 2 As "amplitude-based distance 1" but with an adjustment to set the minimum distance between categories, $d^{\star}$, to 0.5 according to Equation 74.

In Equations 107 and 108 $FT_{x_{nom}^{(i)}}$ is the Fourier transform of the i-th category of the disturbed Branin function; n_{1j} and n_{2j} are the frequencies for the jth mode in the first and second coordinate, respectively.

The rms of the "standardized validation error" is plotted in Figure 23. It shows the comparison between MIC surrogates based on the previously listed nominal distances.

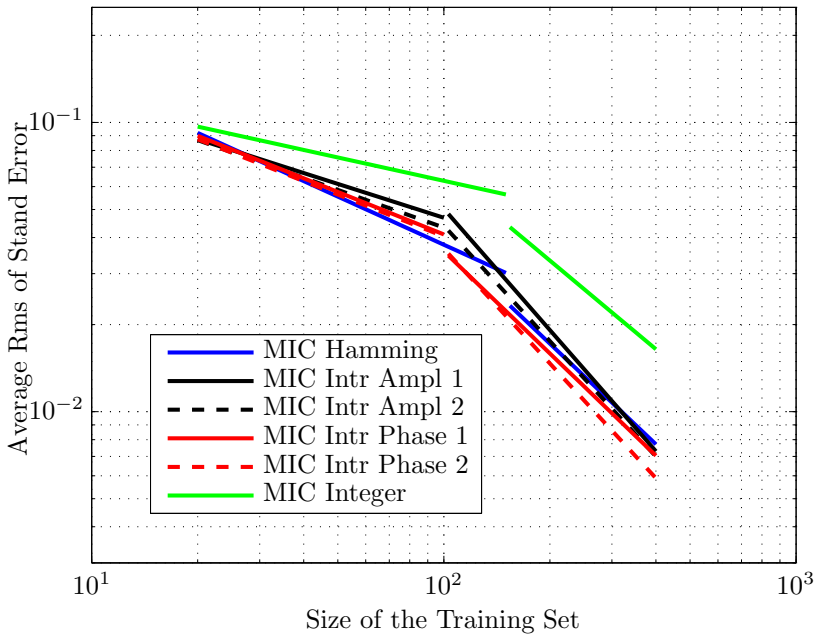


Figure 23: Surrogate "Standardized Validation Error". Comparison Several MIC surrogates. Disturbed Branin Function

Once it has been shown that for small ranges of training set sizes the MIC surrogate is convenient (see Subsection 5.1.1), interest is in the performance of MIC surrogates based on several nominal distances, see Figure 23. This Figure immediate shows that the MIC integer performs badly in comparison with the other nominal distance-based MIC surrogates. The performance of the Hamming and the intrinsic models are similar according to "standardized validation error". It is noticeable that models that adjust the minimum distance (MIC intrinsic amplitude 2 and phase 2) behave better than their non-adapted counterparts (MIC intrinsic amplitude 1 and phase 1). Also, MIC surrogates based on the Fourier amplitude distance provide slightly poorer rms of the "standardized validation error" than the ones based on Hamming and the Fourier phase distances.

Table 5 contains the values of αs for all the MIC surrogates. It is seen that all the MIC surrogates gain performance with the same speed, except the MIC integer that is the slowest in both small and large training set sizes. It is important to realize

Table 5: Values of Constant α for the Fitting Curves of Rms of the "Standardized Validation Error". All MIC Surrogates. Disturbed Branin Function

	Low $N_{tr.set}$	Large $N_{tr.set}$
MIC Hamming	0.553	1.156
MIC Ampl. 1	0.389	1.393
MIC Ampl. 2	0.433	1.326
MIC Phase 1	0.488	1.174
MIC Phase 2	0.481	1.318
MIC Integer	0.268	1.017

that the exact values of αs depend on where the splitting point between the large and small training seats. Thus, a different choice of the splitting point would change the values of αs; however, these changes are expected to be small.

Regarding the "standardized validation residual", its behavior for several nominal distance-based MIC surrogates are similar to the one of MIC Hamming surrogate.

5.2 UH60A Hover Shaft Power. Screened Model

Once the "MIC Surrogate" has been successfully tested on a noise-free canonical function, the next step is to test it in a practical application. Rotor-craft design is the selected application to apply the meta-modeling and adaptive sampling techniques. The multidisciplinary nature of the rotor-craft, the complexity of the rotor-craft physics, and the multi-objective desire of design teams make typical function evaluations time-consuming. Physics based rotor-craft models are noisy due to iterations, discretizations, ... Therefore, rotor-craft objective functions are a great scenario to test MIC surrogates on noisy functions. In order to learn the MIC characteristics when applied to noisy functions, the first rotor-craft design tests are done in screened domains that are more manageable for testing purposes.

5.2.1 Screening of UH60A Hover Shaft Power

In order to better understand the nature of the objective function, a screening of the function is done to reduce the number of design variables. The objective function is

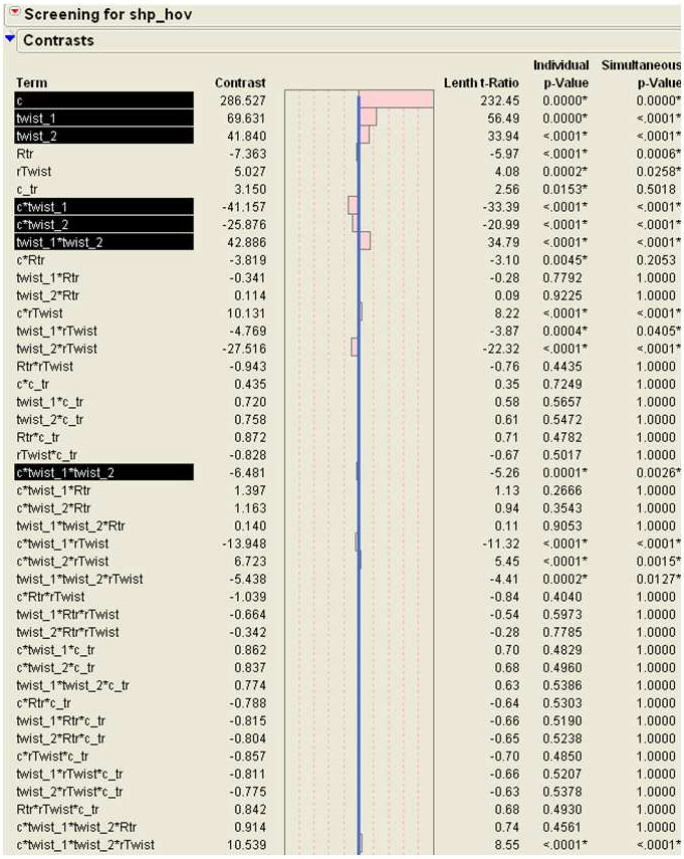
the power consumption of the whole UH60A helicopter, shp_{hov}. For further details of the model, see Section 4.2. Typical design variables are inner θ_1 and outer θ_2 main rotor twists, main rotor chord c, radial position of the twist change r_{tw}, main rotor airfoil, tail rotor radius R_{tr}, and tail rotor chord c_{tr}. The design limits of the full design space are shown in Figure 24.

Variable	Lower Limit	Upper Limit
Inner Linear Twist	-3.4 deg/m	-0.5 deg/m
Outer Linear Twist	-1.5 deg/m	0 deg/m
Main Rotor Chord	1.56ft	3.03ft
Radial Pos. Twist Change	0.72	0.85
Tail Rotor Radius	0.75 ft	0.9 ft
Tail Rotor Chord	5.4ft	6.5ft

Figure 24: Limits of the Design Space to Screen

In order to reduce the domain, a full factorial DoE of 2 levels is carried out in JMP®. Note that the main rotor airfoil is not included in the screening because it is the categorical variable chosen to later test MIC surrogates. The full factorial DoE of 2 levels is done for the airfoil "SC 2110". Figure 25 shows the screening results: the main features of the functions are captured by keeping the inner θ_1 and outer θ_2 main rotor twists; the main rotor chord c; and the interactions between θ_1, θ_2, and c.

Fitting a regressive model in JMP® with θ_1, θ_2, and c as independent variables results in a $RSquare = 0.9852$. The value of $RSquare$ measures the proportion of the function variation around the mean explained by the model rather than random errors. The remaining variation is not explained by the model and is attributed to random error. Therefore, the screened model that keeps θ_1, θ_2, and c catches much of the function behavior (98.52% of the function variation).



Term	Contrast	Lenth t-Ratio	Individual p-Value	Simultaneous p-Value
c	286.527	232.45	0.0000*	0.0000*
twist_1	69.631	56.49	0.0000*	<.0001*
twist_2	41.840	33.94	<.0001*	<.0001*
Rtr	-7.363	-5.97	<.0001*	0.0006*
rTwist	5.027	4.08	0.0002*	0.0258*
c_tr	3.150	2.56	0.0153*	0.5018
c*twist_1	-41.157	-33.39	<.0001*	<.0001*
c*twist_2	-25.876	-20.99	<.0001*	<.0001*
twist_1*twist_2	42.886	34.79	<.0001*	<.0001*
c*Rtr	-3.819	-3.10	0.0045*	0.2053
twist_1*Rtr	-0.341	-0.28	0.7792	1.0000
twist_2*Rtr	0.114	0.09	0.9225	1.0000
c*rTwist	10.131	8.22	<.0001*	<.0001*
twist_1*rTwist	-4.769	-3.87	0.0004*	0.0405*
twist_2*rTwist	-27.516	-22.32	<.0001*	<.0001*
Rtr*rTwist	-0.943	-0.76	0.4435	1.0000
c*c_tr	0.435	0.35	0.7249	1.0000
twist_1*c_tr	0.720	0.58	0.5657	1.0000
twist_2*c_tr	0.758	0.61	0.5472	1.0000
Rtr*c_tr	0.872	0.71	0.4782	1.0000
rTwist*c_tr	-0.828	-0.67	0.5017	1.0000
c*twist_1*twist_2	-6.481	-5.26	0.0001*	0.0026*
c*twist_1*Rtr	1.397	1.13	0.2666	1.0000
c*twist_2*Rtr	1.163	0.94	0.3543	1.0000
twist_1*twist_2*Rtr	0.140	0.11	0.9053	1.0000
c*twist_1*rTwist	-13.948	-11.32	<.0001*	<.0001*
c*twist_2*rTwist	6.723	5.45	<.0001*	0.0015*
twist_1*twist_2*rTwist	-5.438	-4.41	0.0002*	0.0127*
c*Rtr*rTwist	-1.039	-0.84	0.4040	1.0000
twist_1*Rtr*rTwist	-0.664	-0.54	0.5973	1.0000
twist_2*Rtr*rTwist	-0.342	-0.28	0.7785	1.0000
c*twist_1*c_tr	0.862	0.70	0.4829	1.0000
c*twist_2*c_tr	0.837	0.68	0.4960	1.0000
twist_1*twist_2*c_tr	0.774	0.63	0.5386	1.0000
c*Rtr*c_tr	-0.788	-0.64	0.5303	1.0000
twist_1*Rtr*c_tr	-0.815	-0.66	0.5190	1.0000
twist_2*Rtr*c_tr	-0.804	-0.65	0.5238	1.0000
c*rTwist*c_tr	-0.857	-0.70	0.4850	1.0000
twist_1*rTwist*c_tr	-0.811	-0.66	0.5207	1.0000
twist_2*rTwist*c_tr	-0.775	-0.63	0.5378	1.0000
Rtr*rTwist*c_tr	0.842	0.68	0.4930	1.0000
c*twist_1*twist_2*Rtr	0.914	0.74	0.4561	1.0000
c*twist_1*twist_2*rTwist	10.539	8.55	<.0001*	<.0001*

Figure 25: Screening Results of the UH60A Hover Power Consumption

5.2.2 Landscapes of Screened Domain

The design variables are: the inner blade twist θ_1 (continuous variable), the outer blade twist θ_2 (continuous variable), the chord length c (discrete-quantitative variable), and the type of airfoil (categorical variable). Four possible airfoils are available: "NACA 0012", "SC 2110", "NACA 23012", and "SC 1095". The objective function is summarized as follows:

$$shp_{hov}\left(\theta_1, \theta_2, c, \text{airfoil}\right) \tag{109}$$

subjected to

$$\theta_1 \in [-3.5^\circ/m, -0.5^\circ/m] \in \mathbb{R} \qquad \theta_2 \in [-1.75^\circ/m, 0.5^\circ/m] \in \mathbb{R} \tag{110}$$

$$\frac{c}{\bar{c}} \in 0.75 + \left[0, \frac{1}{3}, \frac{2}{3}, 1\right] \cdot 0.5 \qquad \text{airfoil} \in [\text{NACA } 0012, ..., \text{SC } 1095] \tag{111}$$

where $\bar{c} = 1.73ft$ is the baseline chord.

The remaining parameters are $W = 16994lb$ ($C_T = 5.99 \cdot 10^{-3}$), $R = 26.83ft$, $\Omega = 27.0063\frac{rad}{sec}$, and $\frac{r_{\text{cut-off}}}{R} = 0.047$. As a numerical function this function is noisy. A regression model is fitted to visualize and understand its landscapes better. Figures 26, 27, 28, and 29 show the regression of the objective function for different chord values. The four airfoils clearly experience similar trends. It is seen that the regression of shp_{hov} looks smoother than the Branin function studied in Section 5.1; however, it is important to remember that shp_{hov} is a noisy function, and the Figures show a regression fit that absorbs the noise.

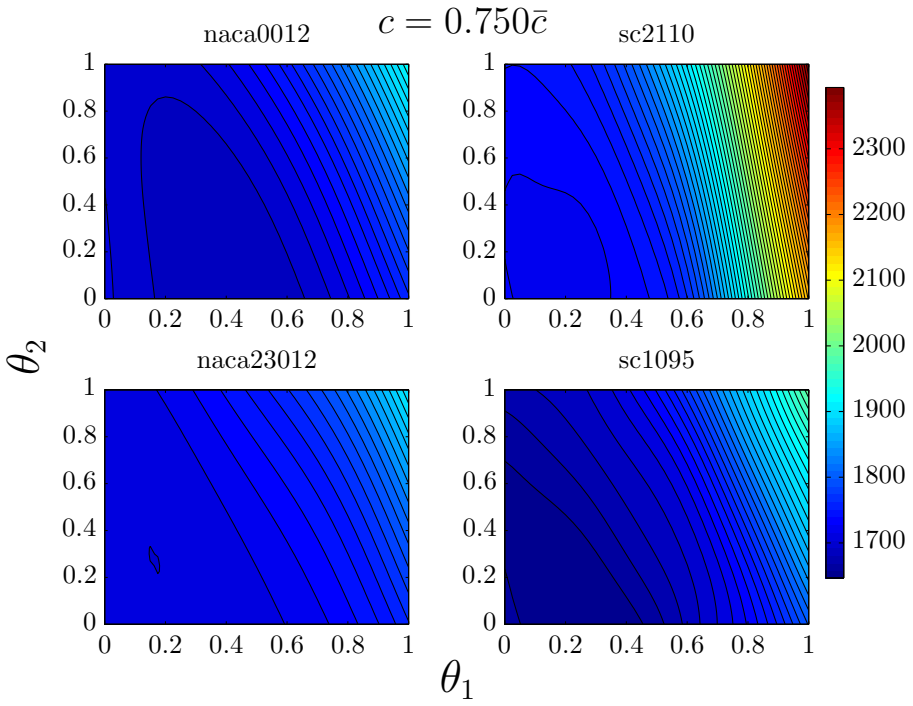


Figure 26: UH60A Total shp_{hov} versus θ_1 and θ_2. $c = 0.750\bar{c}$

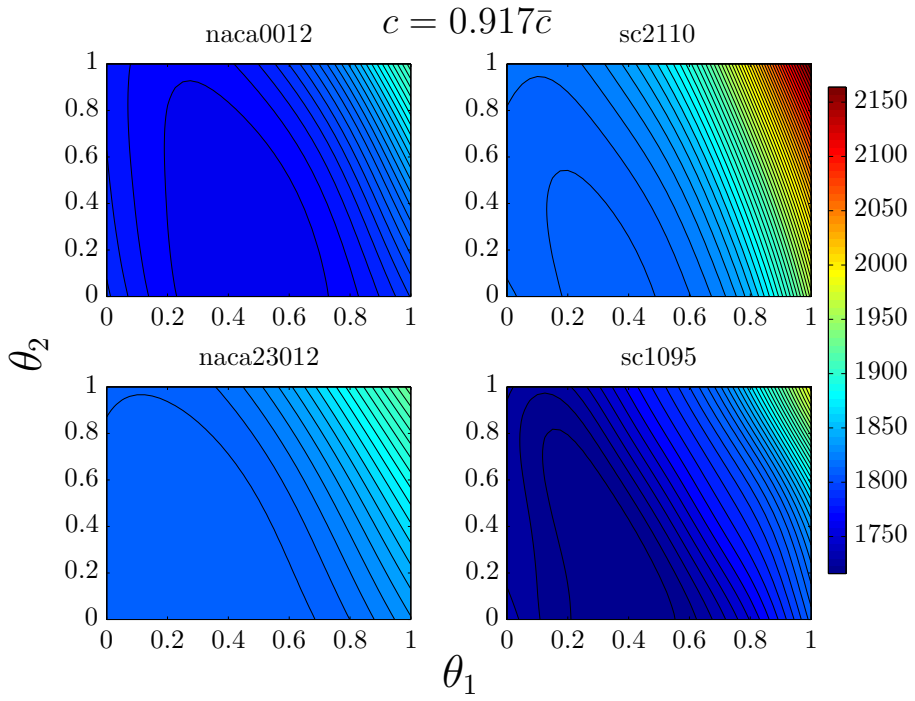


Figure 27: UH60A Total shp_{hov} versus θ_1 and θ_2. $c = 0.917\bar{c}$

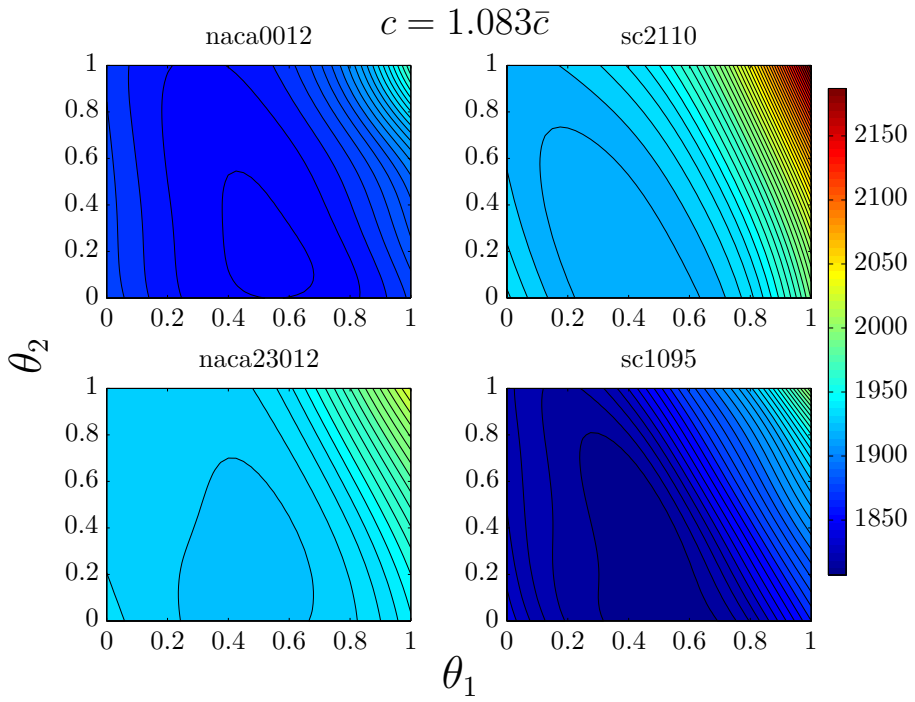


Figure 28: UH60A Total shp_{hov} versus θ_1 and θ_2. $c = 1.083\bar{c}$

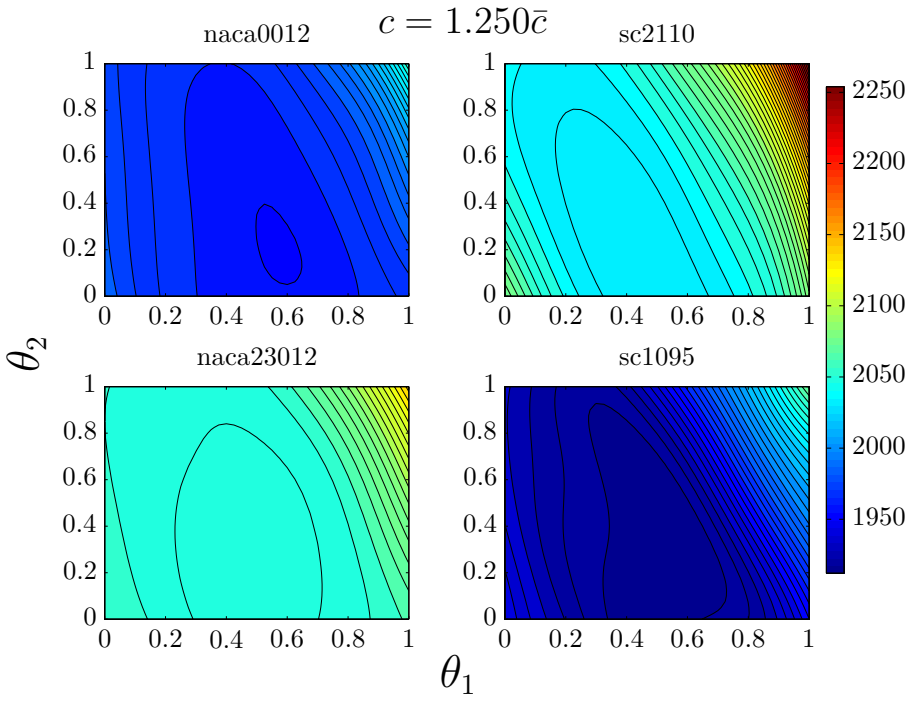


Figure 29: UH60A Total shp_{hov} versus θ_1 and θ_2. $c = 1.250\bar{c}$

5.2.3 Influence of Training Size on Performance

The goal of this section is to see the efficiency of MIC surrogates with respect to independent surrogates (one for each categorical alternative). The objective to test these surrogates is the UH60A hover power in the screened domain. The two meta-models are compared in a range of training set sizes from 36 to 380 points. When building the independent surrogates, the training set is equally divided among the four airfoil surrogates. Since both meta-models are probabilistic, two performance indicators are used for the comparison: rms of the surrogate "standardized validation error", and rms of surrogate "standard validation residual" given by Equations 77, and 78, respectively. A set of 1024 validation points, 256 for each airfoil, are evaluated to assess the mentioned performance indicators.

Again, curves are fitted to try to understand the tendency of the results for each type of surrogate. For the rms of the "standardized validation error", the curve chosen is Equation 105, and for the rms of the "standardized validation residual" Equation 106.

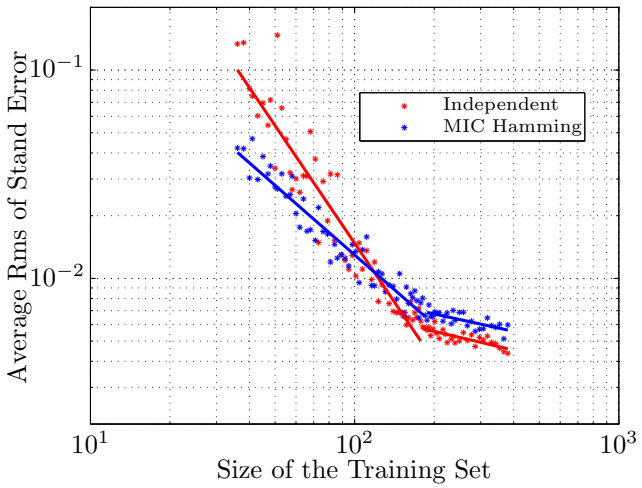


Figure 30: Surrogate "Standardized Validation Error". MIC vs Independent Surrogate. UH60A shp_{hov} Screened Domain

Figure 30 shows that the MIC Hamming surrogate produces a more accurate

surrogate than fitting an independent surrogate for each airfoil in the small range of sampling sizes, specifically from 36 to 125 points. The independent surrogate gains performance, as training set size increases, up to a point where it produces a better surrogate than the MIC one. It appears to reflect that, when the training set is large, the cross-use of observations matters less, as experienced in Subsection 5.1.1.

The reason is that, for low number of points, only a few points are sampled for each category; thus, the rms of the "standardized validation error" for the independent surrogate has a poor performance. Specifically for 36 training points, only nine points are used to build a surrogate in the remaining three design variables: θ_1, θ_2, and c. Nine points are not enough to construct an accurate three dimensional surrogate. However, the MIC surrogate, although it also has only nine observations for airfoil, successfully cross-uses observations from other airfoils, producing more accurate surrogates.

For large training set sizes, the independent surrogate gains in performance until it produces a better surrogate than the MIC Hamming one; this occurs with training sets of approximately 125 points. The reason could be that, in this large training set case, observations hold not only similar trend information but also high frequency information for each airfoil which are different across airfoils. This share of observations at this large range of points even produces surrogates less accurate than the ones that do not share observations.

The gain in performance of independent surrogates with respect to the MIC Hamming surrogates is again quantified by the values of the constant α in the fitting curve of the rms error, see Equation 105. Table 6 contains the values of αs. For small training sets (see "low $N_{tr.set}$" column in Table 6), the independent surrogate converges more quickly than the MIC Hamming one as $N_{tr.set}$ increases. For large training sets, both surrogates converge at a similar speed (similar values of α).

Figure 31 exhibits the rms of the "standardized validation residual" of the UH60A

Table 6: Values of Constant α for the Fitting Curves of Rms of the "Standardized Validation Error". Independent vs MIC Surrogate. UH60A shp_{hov} Screened Domain

	Low $N_{tr.set}$	Large $N_{tr.set}$
Independent Surrogate	1.876	2.993
MIC Hamming	1.11	2.78

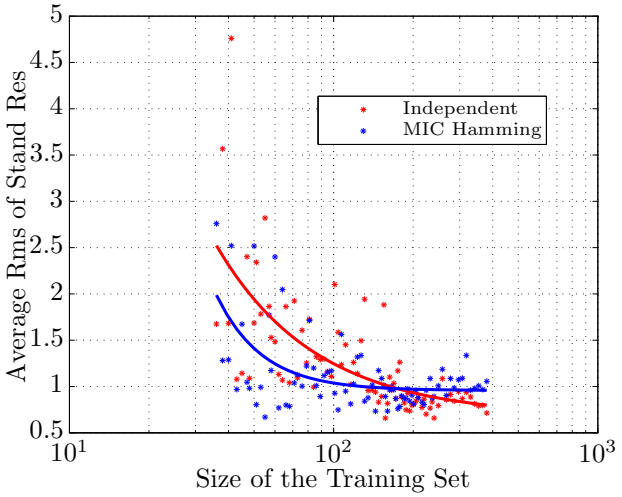


Figure 31: Surrogate "Standardized Validation Residual". MIC vs Independent Surrogate. UH60A shp_{hov} Screened Domain

shp_{hov} for the independent surrogate and the MIC Hamming one. The independent surrogate "standardized validation residual" is higher than one for small training sets. It implies that the uncertainty estimated with the independent surrogate is lower than the real error, i.e., the surrogate uncertainty is underestimated.

When the training set becomes larger, the behavior of the independent surrogate is the opposite: the rms of the "standardized validation residual" is lower than one, implying that the uncertainty estimated from the surrogate is higher than the real error. It is worth noticing that the curve fit shown in Figure 31 is decreasing; however, a closer look into large training set sizes shows that the local tendency of the independent surrogate residual is to increase towards one. One can explain the erroneous behavior of the fitting curve at large training sets with the choice of the fitting curve, $A + \frac{B}{x^{\alpha}}$,that only allows for monotonicity.

The rms of the "standardized validation residual" of the MIC Hamming surrogate

is also higher than one for the small range of $N_{tr.set}$. It also tends to one as the training set increases. Compared with the independent surrogate case, the MIC Hamming residual is closer to one for the majority of the studied range of training set sizes, so the MIC surrogate estimates the uncertainty of the UH60A shp_{hov} better than the independent surrogate.

Both performance indicators demonstrate that the MIC Hamming surrogate outperforms the state-of-the-art for small training set sizes when applied on the noisy UH60A shp_{hov}.

5.2.4 Influence of Nominal Distance on Performance

MIC surrogates can be based on several nominal distances. The influence of these nominal distances in the performance of the MIC surrogate is studied in this Subsection. A similar studied was done in Subsection 5.1.2 for the case of the noise-free disturbed Branin function. For the noisy UH60A shp_{hov}, three nominal distances are tested:

Hamming distance The Hamming distance is employed for the categorical variables.

Integer distance The nominal variable is arbitrarily mapped to a set of integer values. Each category of a non-numeric variable is mapped to an integer in the interval $[1, |x_{nom}|]$, where $|x_{nom}|$ is the number of points in the categorical input x_{nom}. Then, the distance brought to the Kriging surrogate is based on this new arrangement of x_{nom} in the integer line.

C_d-based distance The distance between airfoils i and l is based on the drag coefficient. Since the aerodynamic drag curves are functions, the L^1 distance is brought, see Section 3.3. The distance takes the form of

$$d\left(air^{(i)}\left(\alpha\right), air^{(l)}\left(\alpha\right)\right) = \int_{\alpha_{inf}}^{\alpha_{stall}} |C_d^{(i)}\left(\alpha\right) - C_d^{(l)}\left(\alpha\right)| d\alpha \tag{112}$$

Then, distances given by Equation 112 are normalized to set the largest distance between categories to one.

Results on the performance of MIC surrogates based on several nominal distances are shown in Figure 32. This figure exposes similar results as in the free-noise disturbed Branin function studied in Subsection 5.1.2. Again, the MIC integer performs badly in comparison with the other nominal distance-based MIC surrogates. The reason could be that the integer distance sets the distance randomly and an order among categories, which is not in the nature of nominal variables.

The performance of the MIC Hamming and MIC intrinsic models are quite close according to "standardized validation error". The MIC intrinsic provides only slightly better performance than the MIC Hamming surrogate. However, as mentioned in Section 3.3, the intrinsic distance is not versatile, because it is based in the relationship category-objective function. Also, the intrinsic nominal distance could produce ill-defined Kriging covariance matrix unlike the Hamming distance. Therefore, the Hamming distance is the one used to build MIC surrogates for practical problems.

For blade airfoils there is an underlying space that is materialized in terms of aerodynamic curves at the fidelity level used in the FLIGHTLAB analysis. An intrinsic distance based on these curves provides a similar result to just modeling airfoils as purely categorical variables (using the Hamming distance). The intrinsic distance based on the underlying continuous parameterization for the airfoil choice may be effective while doing CFD analysis. However, it is not the purpose of the research because the multi-disciplinary nature of rotor-crafts makes unaffordable the use of high-fidelity tools like CFD in conceptual design.

Table 7 contains the values of αs for all the MIC surrogates. It is seen that all MIC surrogates gain performance at similar speed. However, the MIC integer has a larger value of the rms of the surrogate "standardized validation error", as shown in Figure 32.

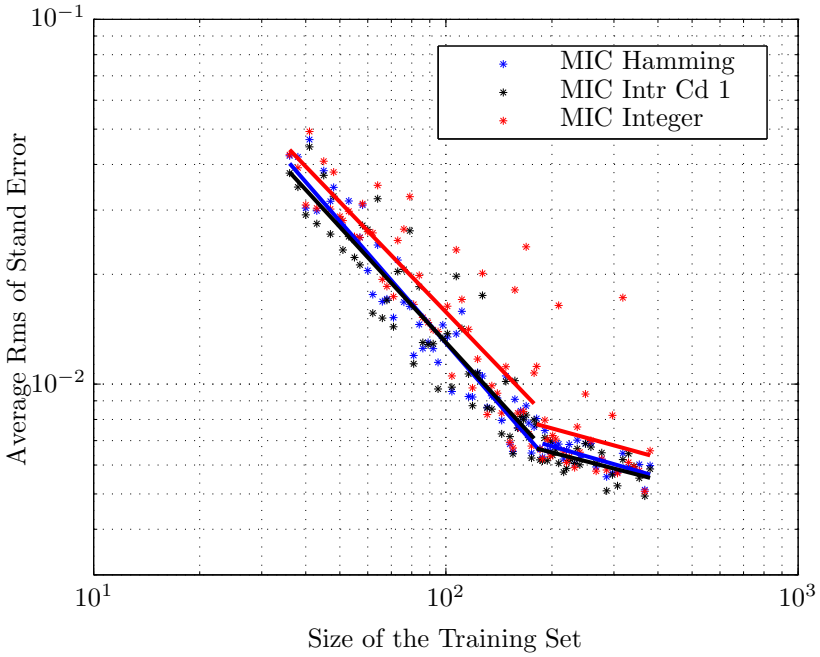


Figure 32: Surrogate "Standardized Validation Error". Comparison Several MIC surrogates. UH60A shp_{hov} Screened Domain

Table 7: Values of Constant α for the Fitting Curves of Rms of the "Standardized Validation Error". All MIC Surrogates. UH60A shp_{hov} Screened Domain

	Low $N_{tr.set}$	Large $N_{tr.set}$
MIC Hamming	1.11	2.78
MIC C_d	1.051	2.498
MIC Integer	1.002	2.653

Regarding the "standardized validation residual", the behaviors for different nominal distance-based MIC surrogates are similar to the one of the MIC Hamming surrogate, as happened with noise-free functions (see Subsection 5.1.2).

CHAPTER VI

MICGA AS THE EGO INFILL CRITERION SOLVER

The goal is to check if the stochastic search algorithm MICGA, developed in Subsection 3.6.2, can drive the adaptive sampling process on a MIC surrogate, which has a mixed-integer-categorical domain. A proper behavior of the "Expected Improvement" criterion is characterized by adaptive sampling updates behaving as EGO optimizers: explore and exploit the MIC surrogate. For more details, see Subsection 2.7.1.

First the validation of the modified MIGA is done to get confidence on the solver before building the MICGA. Several mixed-integer canonical functions from the literature are brought to test the modified MIGA.

Later, the MICGA is tested to see its capability to drive the adaptive sampling process on MIC surrogates. Performance of MIC surrogates with respect to independent ones was tested in Chapter 5. Results showed evidence that, for small training set sizes, modeling a function dependent on non-numeric variables by MIC surrogate is more efficient than building an independent surrogate for each category. Also, experiments on MIC surrogates showed that the Hamming distance is an appropriate nominal distance to build MIC surrogates. Therefore, adaptive sampling algorithms are run on MIC surrogates, which are based on the Hamming distance and whose warm-up sampling plans are in the efficient range of training set sizes pointed out in Chapter 5.

Runs on MIC surrogates with a fixed number of updates are performed to see where the adaptive sampling algorithm places the updates. The goal is to see if the layout of the updates corresponds to a typical EGO pattern: high density of updates located in globally good performing regions. The layout of updates is compared for

MIC Hamming and MIC integer surrogates of the disturbed Branin function (see Appendix A).

Later, the UH60A hover *shp* is adaptively sampled for several warm-up sampling plans. EGO behavior is again searched by looking at the layout of updates.

6.1 Validation Mixed-Integer Genetic Algorithm

Subsection 3.6.2 proposes an algorithm to optimize the ExI criterion on a design space composed of continuous, integer, and categorical design inputs. In order to get to this MICGA, an intermediate step is needed: a MIGA. MIGA stems from the MATLAB® continuous GA by properly adjusting generation, mutation, and crossover functions, as explained in Subsection 3.6.2. Several canonical mixed-integer problems are found in the literature; see Floudas, Pardalos et al. [62], and Himmelblau [87], among others. The tested optimization problems are provided in detail in Appendix C. Results for the proposed MIGA are compared with those given by the source-restricted MIGA provided by MATLAB® 2011b.

Each problem is optimized 100 times with the two algorithms in order to marginalize the intrinsic randomness of GA. Three parameters measure the success of the run: the percentage of success, the average number of function evaluations, and the average time (in seconds) on obtaining the optimal solution. The last two measures are computed only for successful runs.

An optimization is considered successful if all these conditions are satisfied: a) the optimal objective function is within 2% of the known optimal value; b) the average cumulative change in value of the fitness function over "StallGenLimit" generations is less than "TolFun"; and c) the constraint violation is less than "TolCon". In case the optimal value of the objective is zero, an optimization whose optimum absolute value is less than 0.02 is branded as a success. The value of the GA parameters for both MIGA are: TolFun = 10^{-6}, TolCon = 10^{-6}, StallGenLimit = 50, and

Table 8: Solutions Obtained by Using Modified MATLAB® GA and Restricted-Source MATLAB® Mixed-Integer GA

	MATLAB® Modified MIGA			MATLAB® Restricted-Source MIGA		
Problem	% Success	ave Calls	time	% Success	ave Calls	time
1	0.94	5459.7	1.5977	0.56	3004.6	0.7664
2	1	4022.5	1.3054	0.93	1310.1	0.3773
3	1	1561.5	0.3074	1	1561	0.3855
4	1	4258.2	1.4002	0.91	1302.9	0.3951

PopulationSize = 25 and 30 for problems of 2 and 3 independent variables, respectively. The optimization results for the canonical problems are given in Table 8.

Table 8 shows that the percentage of success is better for the modified MIGA, whereas the time consumption is generally lower for the MATLAB® restricted-source MIGA. However, for the test problem 3, the time consumption is lower for the modified MIGA. MIGA parameters have not been tuned because of the corresponding high cost. It is worth noting that the intention of this Section is just to show that the modified MIGA can successfully be used to solve mixed-integer problems; it is a sanity check before building the desired MICGA out of this modified MIGA. An advantage of the modified MIGA is the accessibility to change the mutation, generation, and crossover functions unlike in the restricted-source MIGA provided by MATLAB® where it is not possible due to privacy issues. This freedom is necessary to convert the modified MIGA into a MICGA. This accessibility also allows a case study about the influence of possible categorical mutations in the optimization of the ExI functions on MIC domains.

6.2 Adaptive Sampling of the Disturbed Branin Function

The adaptive sampling algorithm is run for 40 updates starting with initial warm-up sampling plans of 30, 54, and 78 points. The adaptive sampling technique is run for the MIC Hamming, MIC integer and MIC phase-based 1. The problem chosen is the disturbed Branin function explained in Appendix A. The update points are plotted to see if the layout of the update points follows the typical pattern of EGO type optimizer, see Subsection 2.7.1.

An example of update points for a MIC Hamming with 54 warm-up observations is shown in Figure 33. The updates, green circumferences filled with black, are plotted together with contours for each nominal variable; however, it is important to remark that the variable x_2 is discrete-quantitative even though the contour plots are filled. It could be seen that all updates are on tenth values of the x_2 design input.

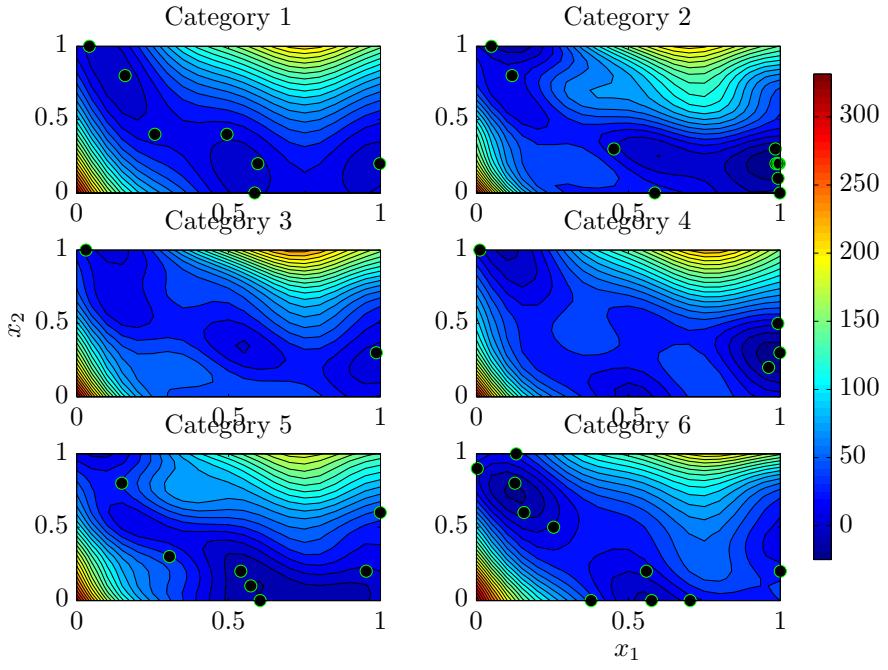


Figure 33: Adaptive Sampling Disturbed Branin Function. Warm-up Size 54. Updates 40. MIC Nominal Metric Hamming.

Figure 33 shows EGO type optimizer's behavior, i.e., samples are concentrated in globally optimal areas of design (categories 2, 4, 5 and 6 as discussed in Appendix A).

So, one can conclude that the MICGA can successfully drive the adaptive sampling process on MIC surrogates.

The 40 updates of the adaptive sampling algorithm on the MIC Integer surrogate are plotted in Figure 34. Compared with the MIC Hamming one, Figure 33, the adaptive sampling on the MIC integer surrogate places less updates than that on the MIC Hamming one in interesting regions of the domain. The former surrogate invests too many resources in category 1 that has a worse performance than other disturbed Branin function categories such as two and four. Also, in the integer metric case, updates are located in useless corners; it does not happen in the Hamming case. Figures 33 and 34 provide evidence that the Hamming metric is more appropriate than the integer one for building a surrogate of the Branin disturbed problem; similar results were found in Subsection 5.1.2 while comparing the efficiency of MIC surrogates based on several nominal distances.

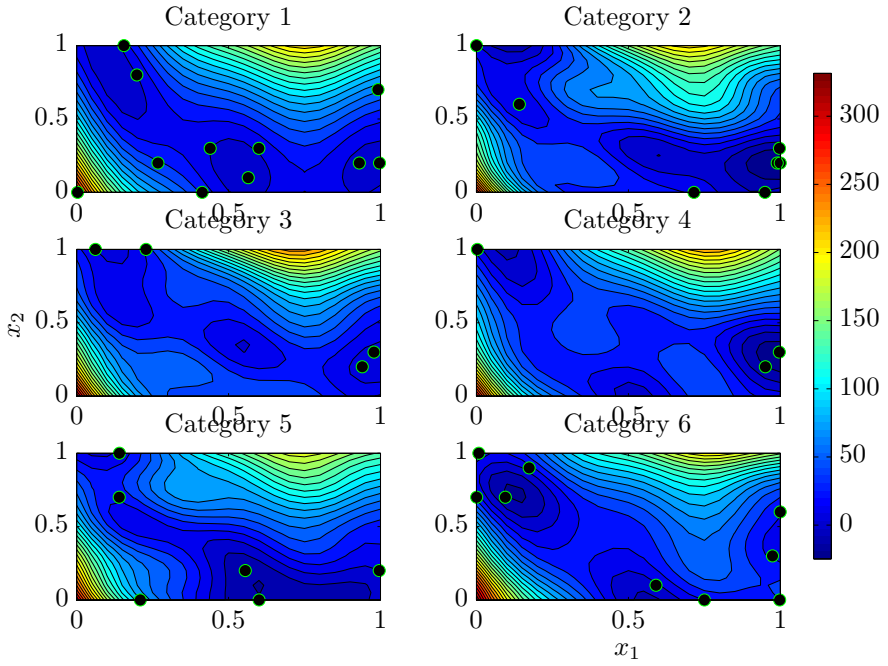


Figure 34: Disturbed Branin Function. Warm-up Size 54. Updates 40. MIC Nominal Metric Integer.

6.3 Adaptive Sampling of the UH60A Hover Shaft Power

In the present example the UH60A hover power consumption is adaptively sampled. Also, the resulting best sample is compared with the baseline case (given by Tables 2, and 3). The optimal solution is searched by adaptive sampling the MIC surrogate. The MICGA optimizes the ExI to search for the next update. The baseline value of the objective function is

$$shp_{hov} = 1901.7hp$$

The optimization problem is defined by

$$shp_{hov}\left(\theta_1, \theta_2, c, \text{airfoil}\right) \tag{113}$$

and subjected to

$$\begin{gathered}
\theta_1 \in [-3.4^\circ/m, -0.5^\circ/m] \qquad \theta_1 \in \mathbb{R} \\
\theta_2 \in [-1.5^\circ/m, 0.5^\circ/m] \qquad \theta_2 \in \mathbb{R} \\
\frac{c}{\bar{c}} \in \{0.9 + \left[0, \frac{1}{3}, \frac{2}{3}, 1\right] \cdot 0.85\} \\
\text{airfoil} \in \{\text{NACA 0012}, ..., \text{SC 1095}\}
\end{gathered} \tag{114}$$

The adaptive sampling algorithm is run on MIC surrogates with the following warm-up sampling sizes: 36, 48, 60, 72, 84 and 96. Table 9 contains the results of the adaptive sampling runs. The indicators of the results are: the percentage of function samples in the best 3%, the optimal design ($\theta_{1,opt}, \theta_{2,opt}, c_{opt}, \text{Airfoil}_{opt}$), and the optimal objective (y_{opt}).

Results in Table 9 demonstrate that the MICGA can drive the EGO algorithm on a MIC meta-model of the UH60A hover *shp*. All the runs provide similar optimal

Table 9: Solution EGO Runs on UH60A shp_{hov}. MIC Surrogate

Warm-up Size	**Updates**	**% in Best** 3%	$\theta^*_{1,opt}$	$\theta^*_{2,opt}$	c^*_{opt}	**Airfoil**$^*_{opt}$	y_{opt}
36	5	1	0.3352	0.08208	0	SC 1095	1709.0356
	20	0.45	0.3352	0.08208	0	SC 1095	1709.0356
	40	0.275	0.3352	0.08208	0	SC 1095	1709.0356
	60	0.35	0.2743	0.3454	0	SC 1095	1708.1737
48	5	0.8	0.2106	0.0002473	0	SC 1095	1711.4107
	20	0.7	0.229	0.4948	0	SC 1095	1708.1578
	40	0.525	0.241	0.4743	0	SC 1095	1708.0969
	60	0.3667	0.241	0.4743	0	SC 1095	1708.0969
60	5	1	0.295	0.2788	0	SC 1095	1708.3896
	20	0.5	0.2837	0.2942	0	SC 1095	1708.3214
	40	0.425	0.2837	0.2942	0	SC 1095	1708.3214
	60	0.3833	0.2837	0.2942	0	SC 1095	1708.3214
72	5	1	0.301	0.1717	0	SC 1095	1708.7405
	20	0.7	0.2847	0.2591	0	SC 1095	1708.4440
	40	0.525	0.2782	0.3248	0	SC 1095	1708.2285
	60	0.45	0.2782	0.3248	0	SC 1095	1708.2285
84	5	1	0.2	0.55	0	SC 1095	1708.7854
	20	0.7	0.2283	0.4658	0	SC 1095	1708.1822
	40	0.525	0.2437	0.4174	0	SC 1095	1708.0962
	60	0.4	0.2437	0.4174	0	SC 1095	1708.0962
96	5	1	0.3302	0.097	0	SC 1095	1708.9790
	20	0.7	0.3302	0.097	0	SC 1095	1708.9790
	40	0.4	0.3302	0.097	0	SC 1095	1708.9790
	60	0.35	0.2525	0.3803	0	SC 1095	1708.1316

results, which represent an improvement of around 10% with respect to the baseline case. A large subset of the updates lands in globally high-performing areas as the indicator "% in Best 3%" demonstrates. Also, once a good approximation to the optimal value is obtained, the value of "% in Best 3%" begins to decrease indicating that the adaptive sampling algorithm starts to explore the unknown parts of the domain. This is again a typical EGO behavior.

6.3.1 EGO Update Location

Once it has been demonstrated that updates are located on globally good performing areas instead of on good performing areas of each category, the adaptive sampling technique is run for 30 updates on MIC Hamming surrogates with several warm-up

sampling sizes: 32, 52, 66, 84, and 136. The purpose is to actually plot the updates and see their location instead of relying on the indicator "% in Best 3%". Layouts of the adaptive sampling updates are shown in Figures 35, 36, 37, 38, 39, and 40. It is worth noticing that the globally best performing areas are designs in $c = 0.9\bar{c}$, and airfoils "SC 1095" and "NACA 0012".

A 32 warm-up sampling plan is adaptively sampled 30 times. Figure 35 shows the 26 updates at $c = 0.9\bar{c}$. The remaining updates are two at $c = 1.183\bar{c}$, and two at $c = 1.75\bar{c}$. It is seen that some of the updates explore bad performing categories, whereas the majority of them are in "SC 1095", the best performing category.

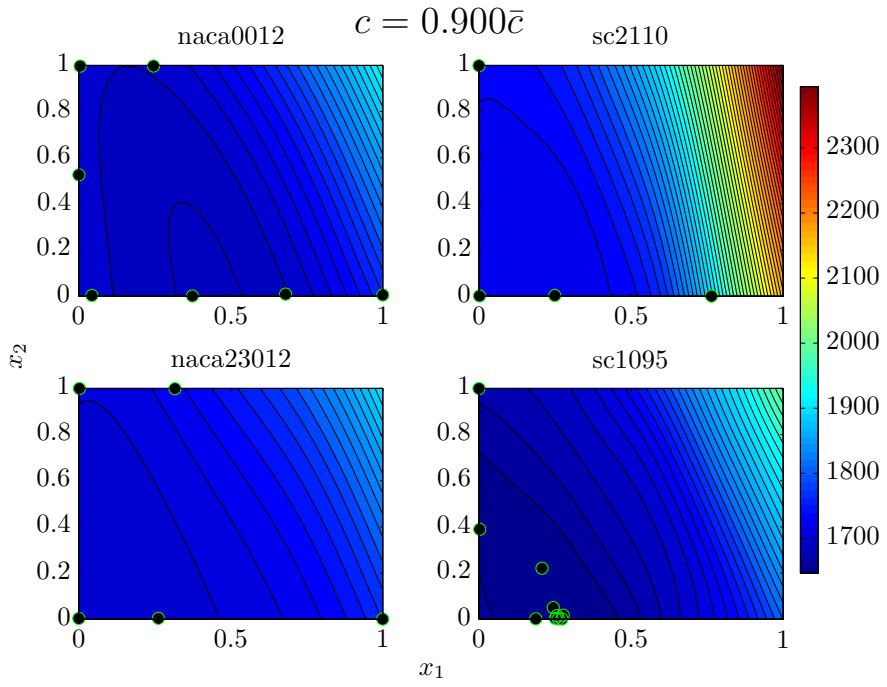


Figure 35: Adaptive Sampling of the UH60A shp_{hov}. Warm-up Size 32. Updates 30. $c = 0.9\bar{c}$. MIC Surrogate

For an initial sampling plan of 52 observations, Figures 36 and 37 show the first 20 and 30 updates that seat on $c = 0.9\bar{c}$, respectively. It is interesting to see that all the 20 first updates are at $c = 0.9\bar{c}$, see Figure 36, i.e., the adaptive sampling algorithm focuses in a globally good performing area, $c = 0.9\bar{c}$.

As the number of updates increases, there is less uncertainty in the globally good

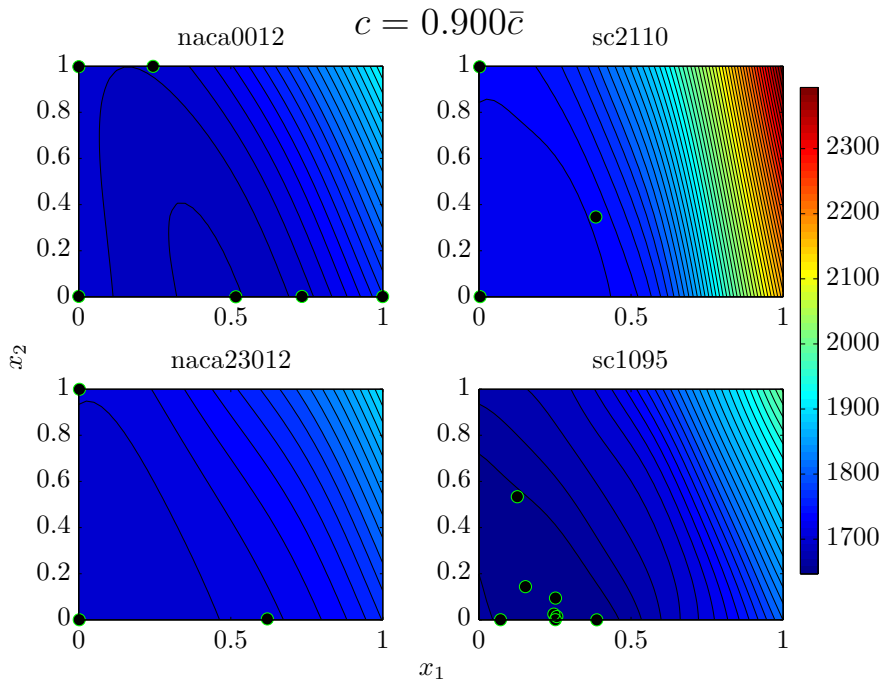


Figure 36: Adaptive Sampling of the UH60A shp_{hov}. Warm-up Size 52. Updates 20. $c = 0.9\bar{c}$. MIC Surrogate

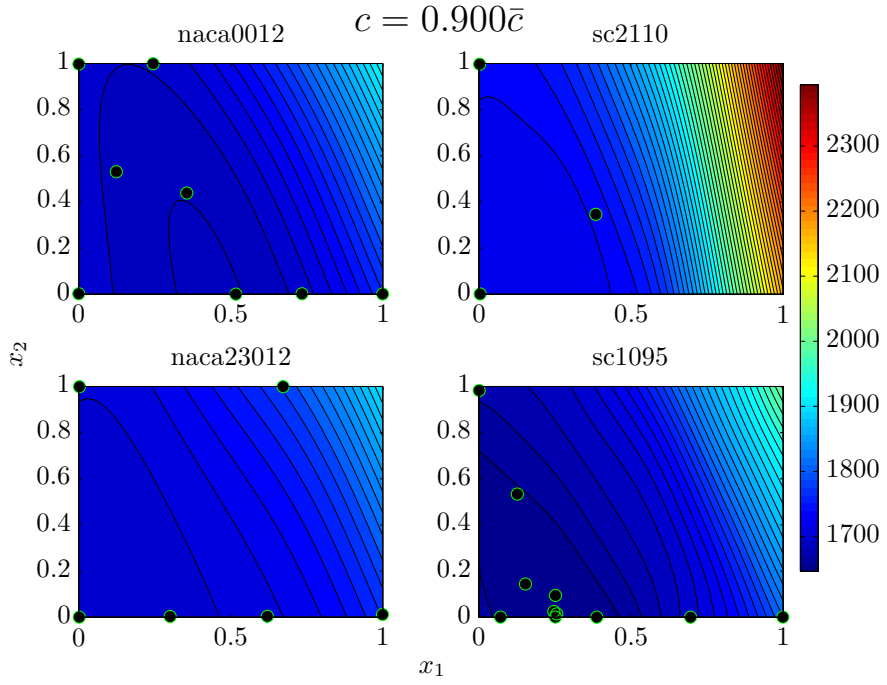


Figure 37: Adaptive Sampling of the UH60A shp_{hov}. Warm-up Size 52. Updates 30. $c = 0.9\bar{c}$. MIC Surrogate

performing regions, such as $c = 0.9\bar{c}$, because of the high density of updates. Consequently, two updates out of the next 10 are in chord values different from $c = 0.9\bar{c}$ (one at $c = 1.183\bar{c}$ and another at $c = 1.75\bar{c}$). Figure 37 exhibits the 28 updates out of the 30 that lay on $c = 0.9\bar{c}$.

19 out of the first 20 updates for a MIC surrogate with 66 warm-up samples are at $c = 0.9\bar{c}$, as Figure 38 depicts. The remaining update is at $c = 1.183\bar{c}$.

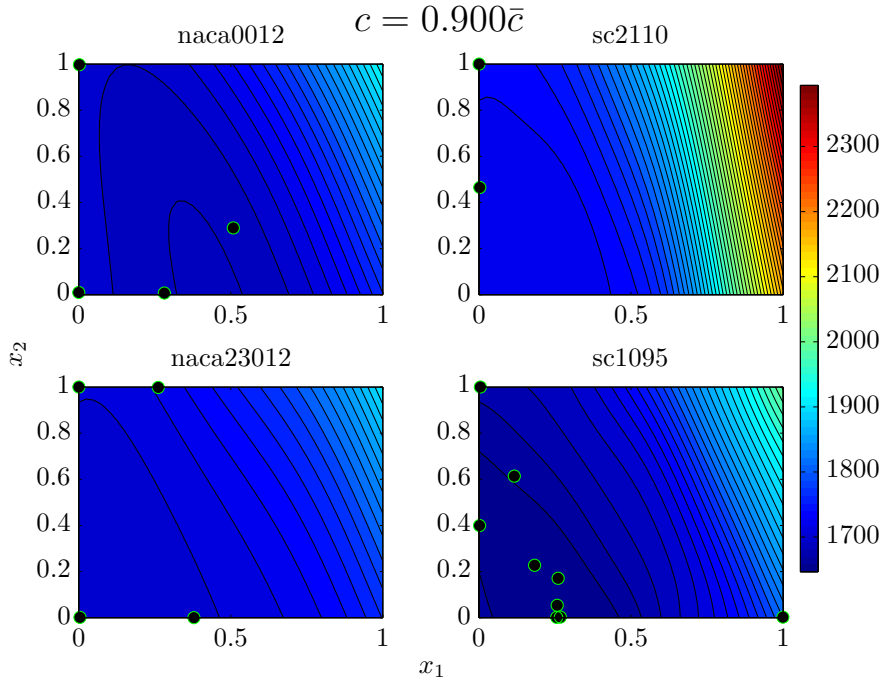


Figure 38: Adaptive Sampling of the UH60A shp_{hov}. Warm-up Size 66. Updates 20. $c = 0.9\bar{c}$. MIC Surrogate

Figure 39 exhibits the updates at $c = 0.9\bar{c}$ starting with 84 warm-up samples. It is seen that, out of the first 10 updates, nine are in the smallest chord. The remaining one is at $c = 1.183\bar{c}$.

The last adaptive sampling algorithm is run on a MIC Hamming surrogate starting with 136 warm-up observations. Figure 40 shows that the 10 first updates are all at $c = 0.9\bar{c}$. There exist enough warm-up points to disregard bad performing areas and fully focus on exploitation.

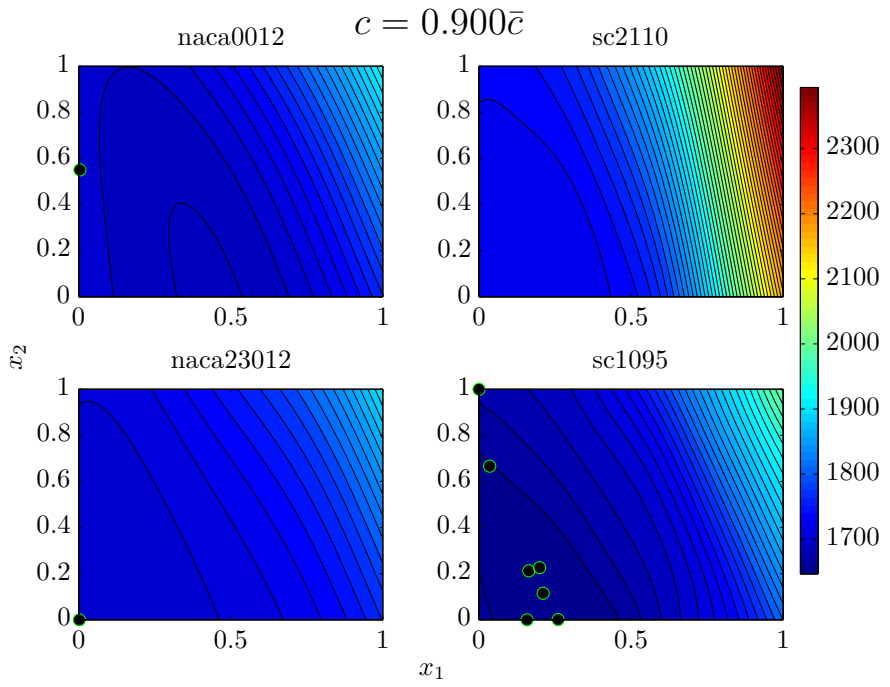


Figure 39: Adaptive Sampling of the UH60A shp_{hov}. Warm-up Size 84. Updates 10. $c = 0.9\bar{c}$. MIC Surrogate

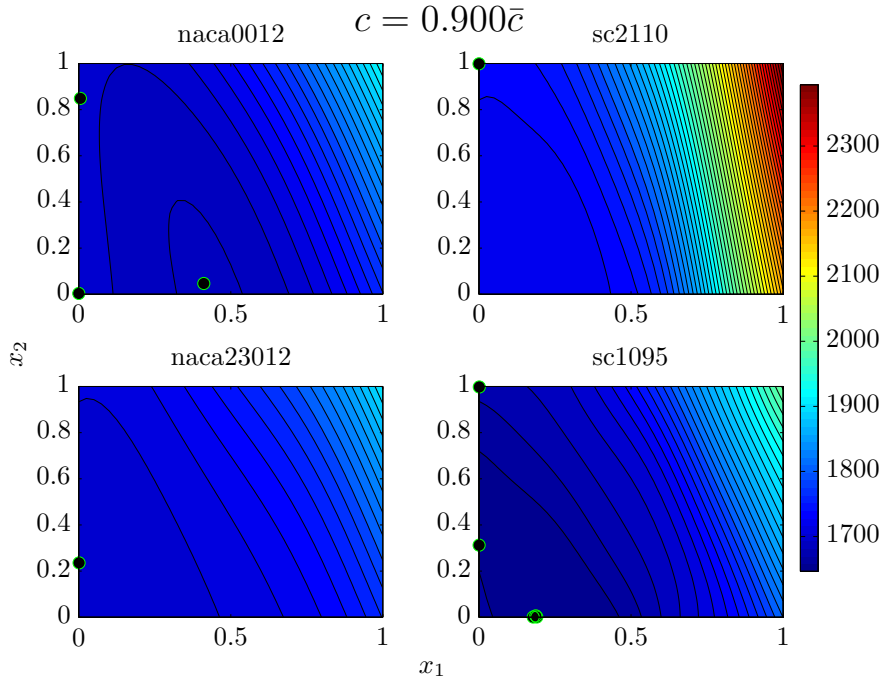


Figure 40: Adaptive Sampling of the UH60A shp_{hov}. Warm-up Size 136. Updates 10. $c = 0.9\bar{c}$. MIC Surrogate

In all the shown cases, the EGO type behavior occurs: the samples are concentrated in the best performing areas across the whole design space, which are $c = 0.9\bar{c}$, and airfoils "SC 1095" and "NACA 0012".

6.4 Study of Mutation Strategies

Once new nominal distance measures have been tested for building MIC surrogates, see Chapter 5, new questions come up: could the distance metrics be efficiently used in the mutation functions of MICGA when optimizing ExI on MIC domains? which combination of mutation strategies for all types of variables provides best optimization results?

Herein, MICGA, presented in Subsection 3.6.2, maximizes the ExI over the MIC design space for obtaining the next update point. Normally, expected improvement landscapes have a characteristic multi-modality with hills between points already sampled as pointed out by Jones [99]. These characteristics make possible to extrapolate the results of ExI optimization tests on canonical objectives to other objectives.

As explained in Subsection 3.6.3, several possibilities can be chosen for the categorical mutation function of the MICGA: uniform, inverse distance, and maximum entropy. The following overall mutation strategies for each type of variable are tested:

1) Normal, Geometrical, and Uniform. Normal, geometrical, and uniform distributions for continuous, integer, and categorical variables, respectively.

2) Normal, Geometrical, and Inverse Distance. Normal, geometrical, and inverse distance distributions for continuous, integer, and categorical variables, respectively. Subsection 3.6.3 explains the inverse distance distribution for categorical variables.

3) Normal, Geometrical, and Maximum Entropy. Normal, geometrical and maximum entropy distributions for continuous, integer, and categorical variables, respectively. Subsection 3.6.3 explains the maximum entropy distribution for categorical variables.

4) Uniform, Uniform, and Uniform. Uniform distributions for all kinds of variables.

5) Uniform, Uniform, and Inverse Distance. Uniform distributions for continuous, and integer variables; and inverse distance distribution for categorical ones.

6) Uniform, Uniform, and Maximum Entropy. Uniform distributions for continuous, and integer variables; and maximum entropy distribution for categorical ones.

The success of the method is given in terms of the error of the maximization process, and the generation at which the last improvement larger than 1% is achieved in the optimization process, defined in Subsection 3.6.3.3. These parameters express the precision of the solution, and the efficiency of each overall mutation strategy in the optimization process.

The ultimate goal of this Subsection is to figure out which mutation strategy is the most appropriate for the maximization of the ExI. However, before dealing with it, an study of the performance of each mutation strategy in the maximization of a canonical function is carried out. The canonical function chosen is the disturbed Branin function, see Appendix A. Five distinct disturbed Branin functions are minimized with the MICGA 150 times. This is done to marginalize the intrinsic random nature of stochastic search. The success parameters are presented in Figures 41, and 42.

According to Figures 41 and 42 normal and geometrical distributions seem generally to provide better results than constant distributions for continuous and integer variables, respectively, i.e., the mean and standard deviation of the success parameters are generally smaller for normal and geometrical distributions than for uniform distributions. Also, it is noticeable that the difference between the mutation strategies with different categorical mutations are not large.

The next step is to study the mutation strategies in the maximization of a ExI on MIC domains. Several ExI landscapes have been generated at different stages of

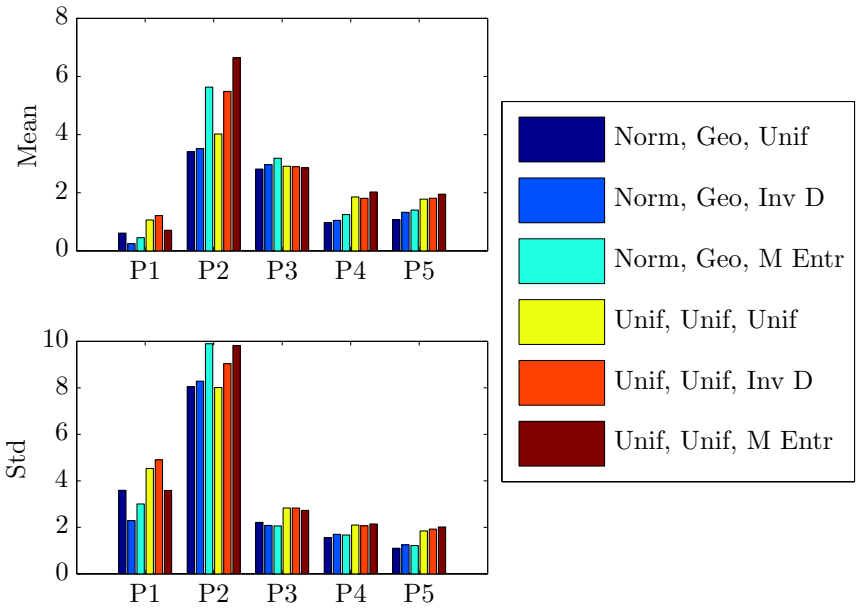


Figure 41: Statistics of Solution Error for Five Disturbed Branin Problems. MICGA Mutation Study

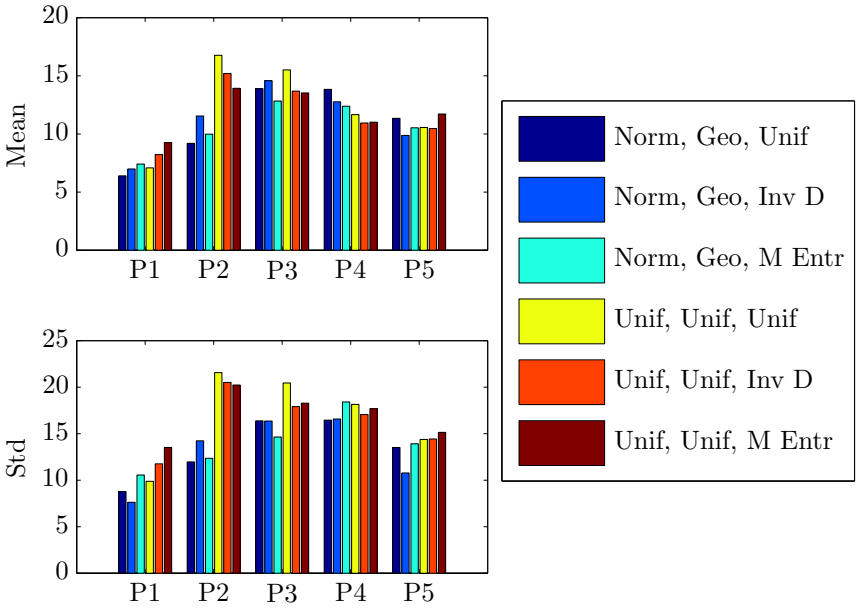


Figure 42: Statistics of Generation of the Last Improvement Larger than 1% for Five Disturbed Branin Problems. MICGA Mutation Study

the adaptive sampling process on the disturbed Branin function plotted in Figure 20. Again, the optimization is run 150 times to marginalize the randomness of the MICGA.

Figures 43 and 44 exhibit the error of the maximization process and the generations at which the last improvement larger than 1% is achieved in the optimization process, respectively.

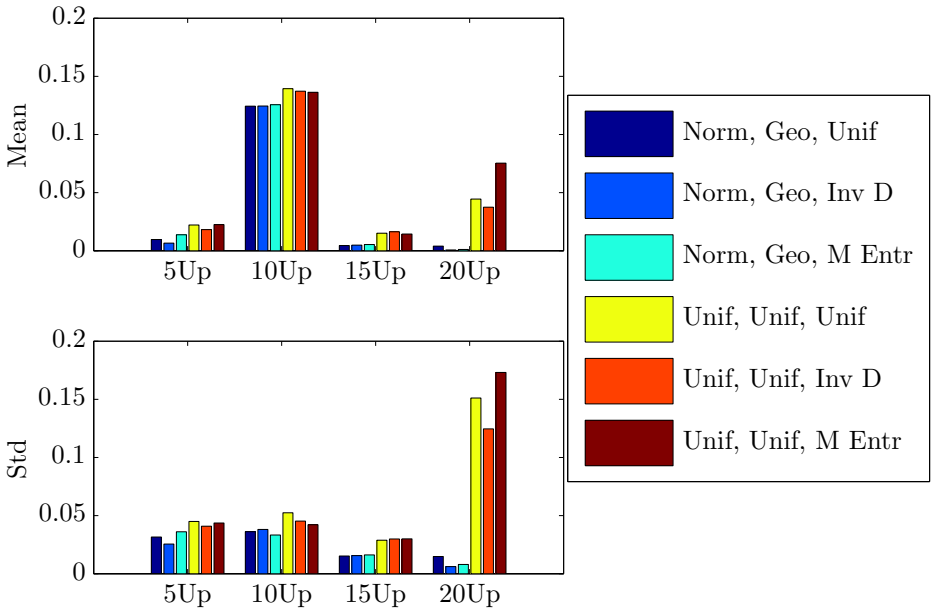


Figure 43: Statistics of Solution Error for ExI Optimization with Several Updates. MICGA Mutation Study

Results in Figure 43 and 44 reveal again that normal and geometrical mutation strategies for continuous and integer variables perform normally slightly better than uniform mutation strategies in terms of both success parameters. Also, little performance differences are found across the categorical mutation strategies, as experienced in the direct optimization of the disturbed Branin function.

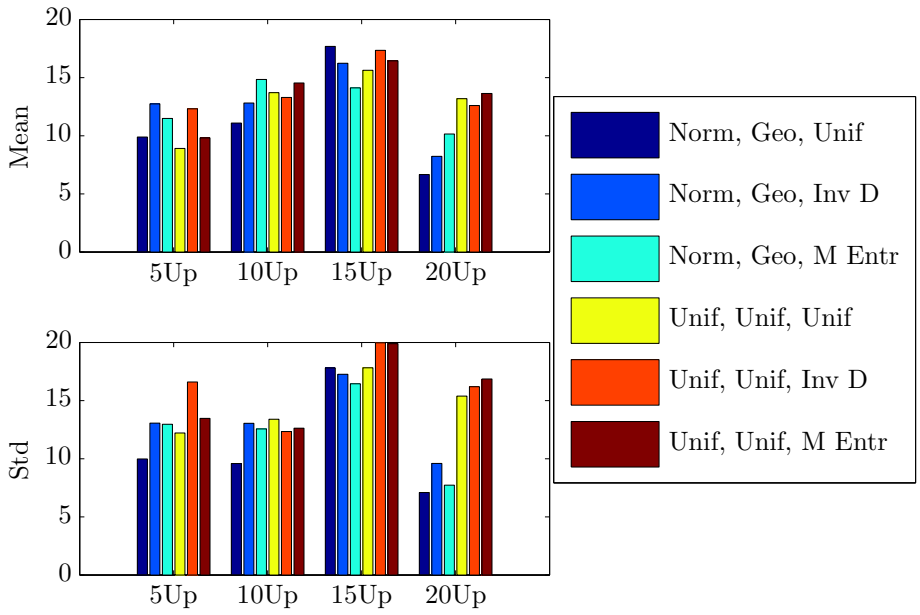


Figure 44: Statistics of Generation of the Last Improvement Larger than 1% for ExI Optimization with Several Updates. MICGA Mutation Study

CHAPTER VII

ECMF EXPERIMENTS: NEW CONCEPT AND OLD TRAINING SIZE

One of the goals of this research is to leverage similar trends between concepts to enhance the current poor strategy of fitting a surrogate independently for each concept. Two Gaussian surrogates are brought in this Chapter:

ECMF Surrogates Surrogates that are developed to leverage similar trends on incrementally evolved concepts. A multi-fidelity framework reuses observations of previously sampled concepts when building a new concept surrogate, see Section 3.5.

Mono-fidelity Surrogates It is the current state-of-the-art method. A surrogate is fitted for the new concept with no reuse of observations from previous concepts.

It is expected that the multi-fidelity surrogate provides better results than the independent surrogate for some range of the training set sizes of concept 1 and 2. A case study is carried out to compare concept 2 ECMF surrogates with concept 2 mono-fidelity surrogates for several sizes of the training sets. In order to take advantage of the similar trends, concept 1 has to be already sampled, so a concept 1 surrogate with certain accuracy is available. Thus, the size of the concept 1 training data-set must be chosen above a certain size to assure that a precise concept 1 surrogate is available.

A canonical set of concepts is tested to see the efficiency of the proposed ECMF method. The canonical set of functions is the Michalewicz one, which is a noise-free set of functions. Appendix B explains in more detail the construction of this canonical

set of functions. Then, a test on a more practical scenario is done: the UH60A with fenestron hover power when observations from the conventional tail UH60A hover power are available. Both hover power simulations are noisy and computationally intense functions. As it was mentioned in Section 1.7, tests on computationally expensive models do not help to better interpret the canonical test results or support the research hypothesis or predictions; however, they support the practicality of the application.

The comparison of the ECMF surrogate with the current state-of-the-art is made in terms of the success indicators introduced in Subsection 3.5.1: rms of the "standardized validation error", and rms of the "standardized validation residual", see Equations 84, and 85, respectively.

7.1 Michalewicz Canonical Function

7.1.1 Influence of New Concept Training Size on Performance

The concept 1 Michalewicz canonical function has one independent variable; its landscapes can be seen in Figure 76 in Appendix B. An old concept training set of seven observations produces a decent old concept surrogate. The concept 2 training dataset studied in this Subsection ranges in the interval $N_{2,tr.set} \in [6, 30]$. The success indicators are: concept 2 rms of the "standardized validation error", and concept 2 rms of the "standardized validation residual" are plotted in Figures 45 and 46, respectively.

Both success indicators show that, in the small range of $N_{2,tr.set}$, the proposed concept 2 ECMF surrogate outperforms the concept 2 surrogate with no reuse of information (mono-fidelity surrogate). The range in which the ECMF surrogate provides better rms of the "standardized validation error" is smaller than that range in the rms of the "standardized validation residual" case. It is worth realizing that there is a value of $N_{2,tr.set}$ above which the mono-fidelity surrogate outperforms the ECMF one; thus, the observations from previous concepts must not be used beyond

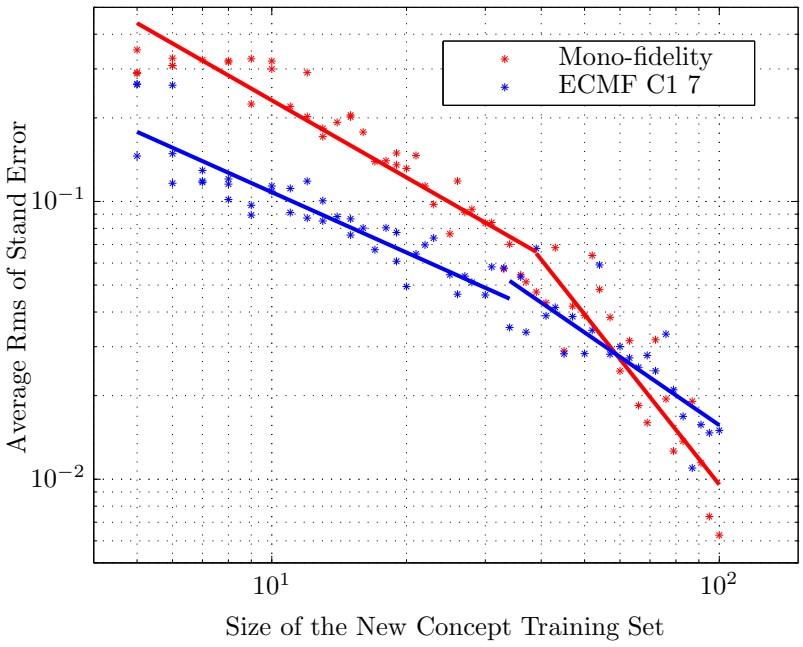


Figure 45: Concept 2 Surrogate "Standardized Validation Error". ECMF $N_{1,tr.set} = 7$ vs Mono-Fidelity Surrogate. Michalewicz Function.

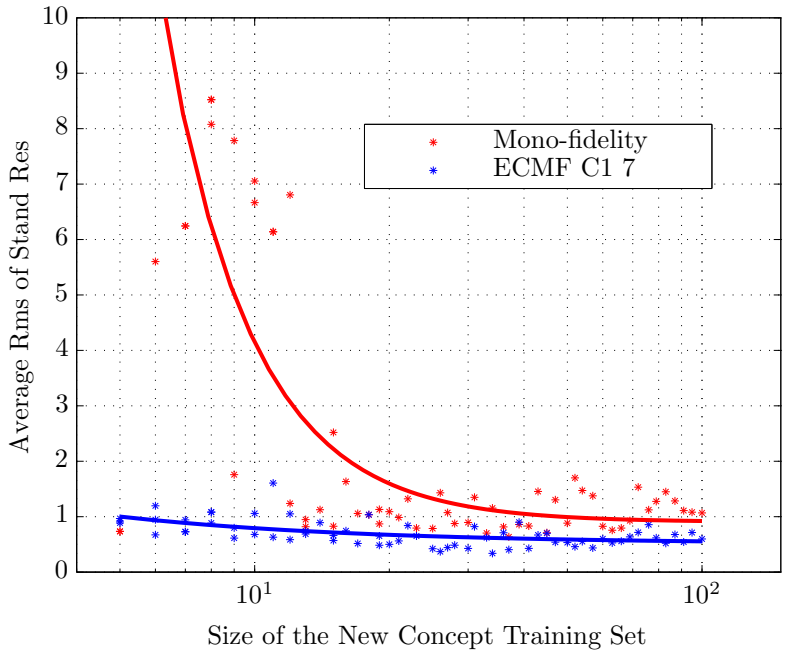


Figure 46: Concept 2 Surrogate "Standardized Validation Residual". ECMF $N_{1,tr.set} = 7$ vs Mono-Fidelity Surrogate. Michalewicz Function.

Table 10: Values of Constant α for the Fitting Curves of Rms of the "Standardized Validation Error". Mono-fidelity vs ECMF Surrogate. Michalewicz Function.

	Low $N_{2,tr.set}$	Large $N_{2,tr.set}$
Mono-fidelity Surr.	0.9231	2.0357
ECMF $N_{1,tr.set} = 7$ Surr.	0.7218	1.1092

this value of $N_{2,tr.set}$.

The gain in performance of the mono-fidelity surrogate with respect to the ECMF one is quantified by the values of the constant α in the fitting curves of the rms of the "standardized validation error", similarly as in Subsection 5.1.1 (see Equation 105). Table 10 contains the values of α. It is seen that the mono-fidelity surrogate converges more quickly to zero than the ECMF with $N_{1,tr.set} = 7$ as the size of the new concept training set. Also, it is seen that if two fitting curves are used (one for small and another for large $N_{2,tr.set}$), the gain in performance of the mono-fidelity surrogate in large training sets is higher than in the small ones (see Table 10).

7.1.2 Influence of Previous Concept Training Size on Performance

Also, it is interesting to see the influence of the $N_{1,tr.set}$ on the ECMF surrogate performance. The studied range for $N_{1,tr.set}$ is $[5, 11]$. Figure 47 exhibits the rms of the "standardized validation error" for ECMF surrogates with several values of $N_{1,tr.set}$.

For small values of $N_{2,tr.set}$, the success indicator performance increases as $N_{1,tr.set}$ does. It is noticeable that the rms of the "standardized validation error" for ECMF surrogates with $N_{1,tr.set} = 9$ and $N_{1,tr.set} = 11$ are quite similar in this small range of $N_{2,tr.set}$.

On the high range of $N_{2,tr.set}$, it seems to be a tendency that as $N_{1,tr.set}$ increases, the surrogate rms of the "standardized validation error" becomes larger. Figure 47 demonstrates that around $N_{2,tr.set} \approx 100$, the ECMF surrogate with $N_{1,tr.set} = 9$ outperforms the ECMF surrogate with $N_{1,tr.set} = 11$. Thus, the rms of the "standardized

validation error" dependence on $N_{1,tr.set}$ seems to reverse for large values of $N_{2,tr.set}$. It appears to reflect that, when the new concept training set is large, the re-use of observations matters less. It is because the new concept sampling plan provides the necessary features to fit an accurate surrogate.

Also, as the number of previous concept observations, $N_{1,tr.set}$, increases, the concept 1 (low-fidelity) surrogate captures features beyond common trends. This high frequency features can mislead the new concept ECMF surrogate, specially for high $N_{2,tr.set}$ where new concept high frequency information is already captured by the new concept sampling plan.

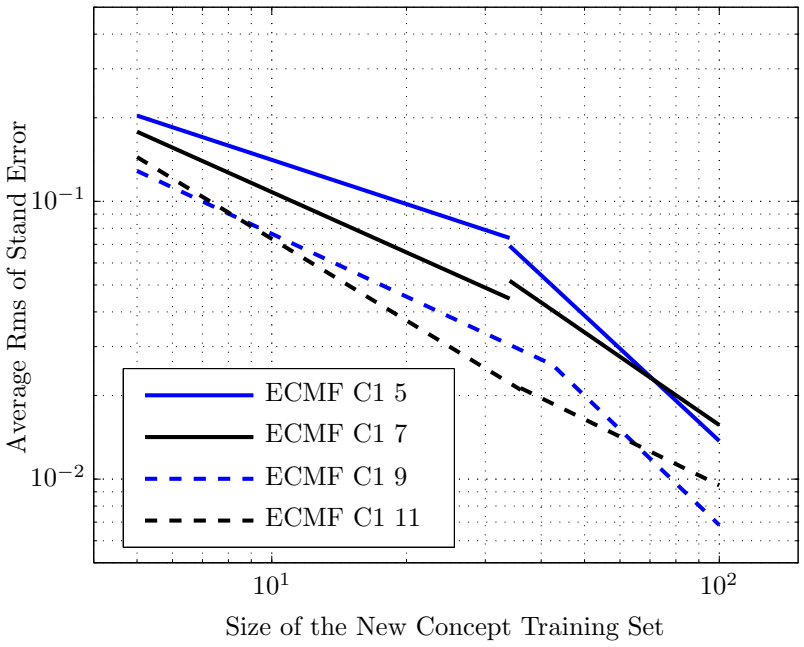


Figure 47: Concept 2 Surrogate "Standardized Validation Error". ECMF Several $N_{1,tr.set}$. Michalewicz Function.

Table 11 contains the values of α for ECMF surrogates built with several values of $N_{1,tr.set}$. It is seen that all the ECMF surrogates gain performance with more or less similar speed. Looking carefully, one can see the previously discussed rms of the error dependence on $N_{1,tr.set}$ for the low and high ranges of $N_{2,tr.set}$.

Table 11: Values of Constant α for the Fitting Curves of the Rms of the "Standardized Validation Error". All ECMF Surrogates. Michalewicz Function

	Low $N_{2,tr.set}$	Large $N_{2,tr.set}$
ECMF $N_{1,tr.set} = 5$	0.5290	1.4988
ECMF $N_{1,tr.set} = 7$	0.7218	1.1092
ECMF $N_{1,tr.set} = 9$	0.7517	1.5479
ECMF $N_{1,tr.set} = 11$	0.9773	0.7891

7.2 UH60A with Fenestron Tail Hover Shaft Power. Screened Domain

Before testing the ECMF surrogate on a UH60A model with a large design space, some tests are done in a screened design space to better understand the ECMF surrogate performance on noisy functions with manageable design spaces. The objective function is the power consumption of the whole UH60A helicopter with a fenestron tail, denoted by shp_{hov}^{fen}. The previous or old concept is the UH60A with a conventional tail.

The design variables are: the inner blade twist (continuous variable), the outer blade twist (continuous variable), the number of fan blades (discrete-quantitative variable), and the chord length (discrete-quantitative variable). One can see that the new concept design variables are the ones obtained in the screening done in Subsection 5.2.1, plus the number of fan blades. This last variable is included to make the new concept design space different from that of the old concept.

The objective function is summarized as follows:

$$shp_{hov}^{fen}\left(\theta_1, \theta_2, N_{b,tr}, c\right) \tag{115}$$

subjected to

$$\begin{aligned} \theta_1 \in & [-3.5^\circ/m, -0.5^\circ/m] \qquad \theta_1 \in \mathbb{R} \\ \theta_2 \in & [-1.75^\circ/m, 0.5^\circ/m] \qquad \theta_2 \in \mathbb{R} \\ N_{b,tr} \in & 7, 8, 9 \\ \frac{c}{\bar{c}} \in & 0.75 + \left[0, \frac{1}{3}, \frac{2}{3}, 1\right] \cdot 0.5 \end{aligned} \tag{116}$$

where $\bar{c} = 1.73ft$ is the baseline chord.

The remaining parameters are $W = 16994lb$ ($C_T = 5.99 \cdot 10^{-3}$), $R = 26.83ft$, $\Omega = 27.0063\frac{rad}{sec}$, and $\frac{r_{\text{cut-off}}}{R} = 0.047$. The main rotor blade section is "NACA 0012".

The numerical and iterative nature of the shp_{hov}^{fen} make it a noisy function. Thus, a regression model is fitted to plot its landscapes. Figures 48, 49, 50, and 51 show the regression of the objective function for the feasible chords. The regression of the function looks less multi-modal than the Michalewicz function studied in Section 7.1. However, it is worth realizing that the noise is not seen in Figures 48, 49, 50, and 51 because they are regressing meta-models, but the function is noisy.

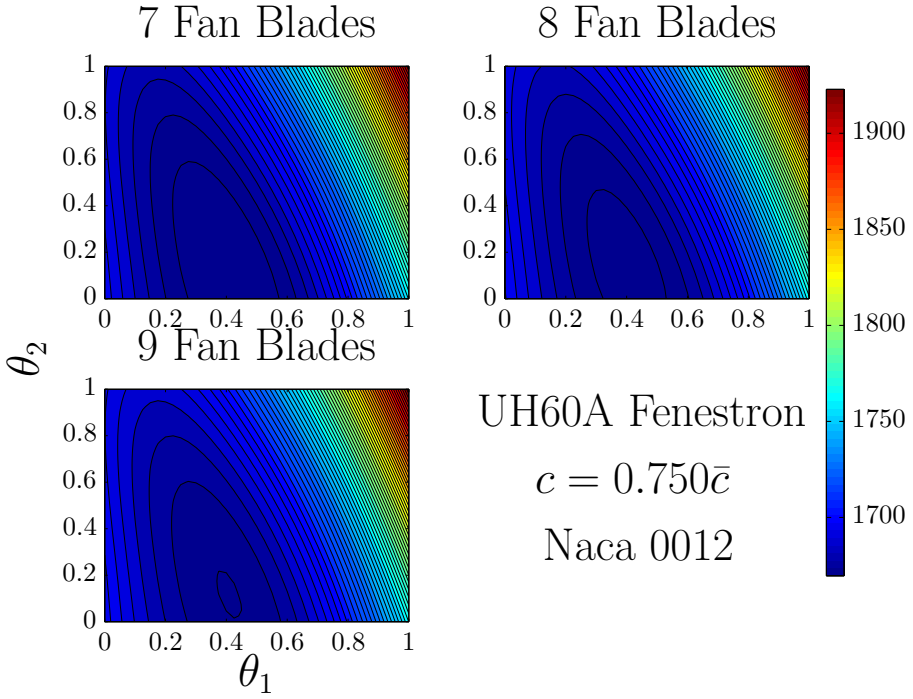


Figure 48: UH60A with Fenestron Total shp_{hov}^{fen} versus θ_1 and θ_2. $c = 0.750\bar{c}$

The previous concept from which observations are reused is the UH60A with conventional tail shp_{hov}^{conv}. It is defined in Equations 109, 110, and 111 with the only change that just the airfoil "NACA 0012" is allowed.

Notice that each concept design space for the power consumption (shp_{hov}^{fen} and shp_{hov}^{conv}) is different. The UH60A with the conventional tail has one less design variable, $N_{b,tr}$, than the UH60A with fenestron. Therefore, when building the fenestron concept ECMF surrogate, the previous concept behavior along $N_{b,tr}$ is assumed to be constant, as explained in Section 3.5.

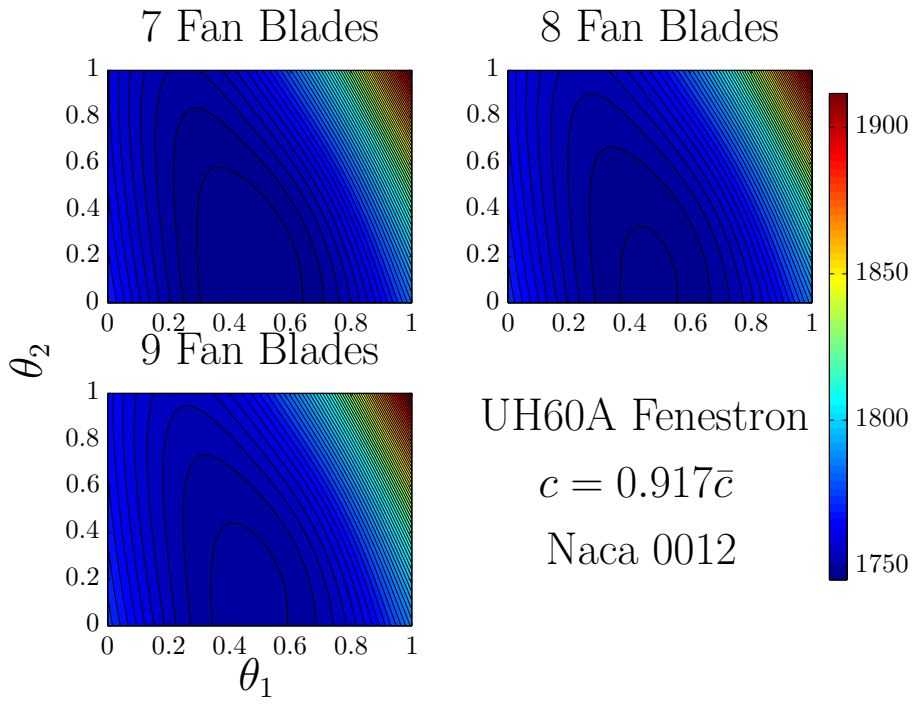


Figure 49: UH60A with Fenestron Total shp_{hov}^{fen} versus θ_1 and θ_2. $c = 0.917\bar{c}$

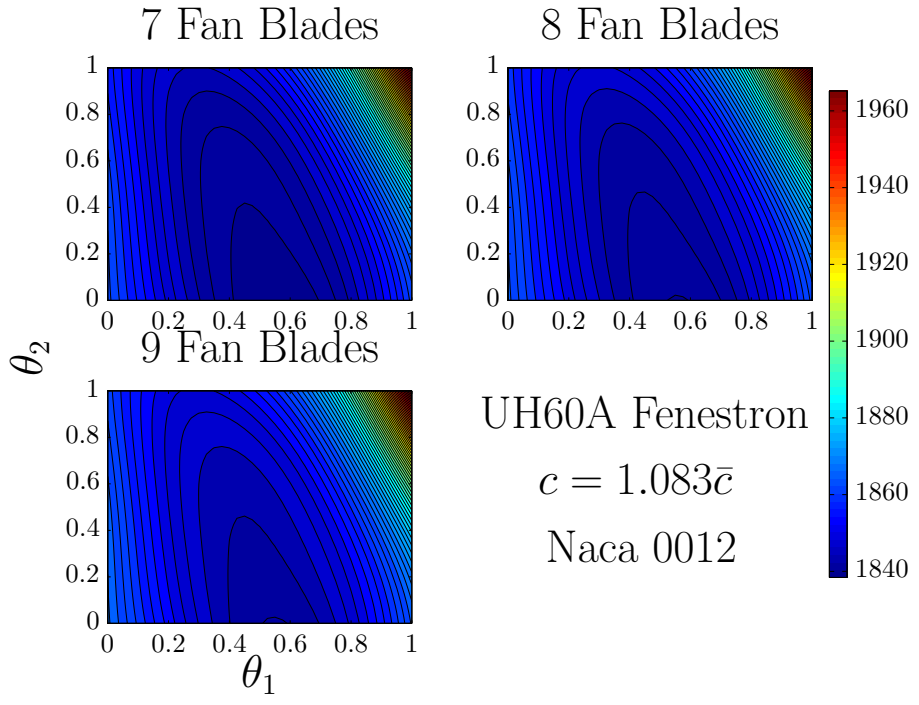


Figure 50: UH60A with Fenestron Total shp_{hov}^{fen} versus θ_1 and θ_2. $c = 1.083\bar{c}$

The landscape of shp_{hov}^{conv} were shown in Figures 26, 27, 28, and 29. One can see that the power consumptions of the two concepts experience similar trends along their

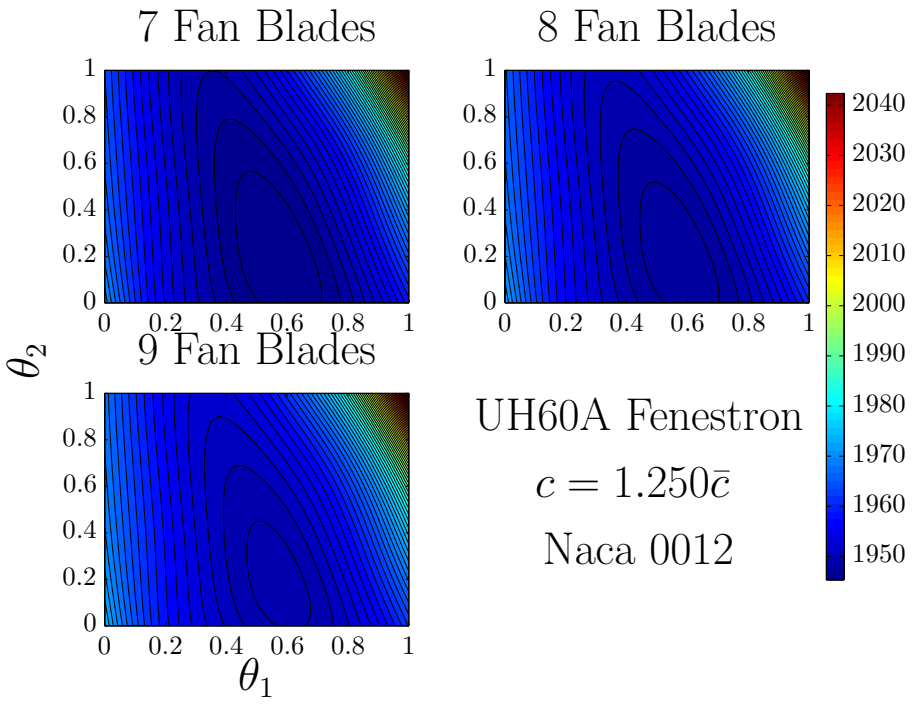


Figure 51: UH60A with Fenestron Total shp_{hov}^{fen} versus θ_1 and θ_2. $c = 1.250\bar{c}$

common design variables.

7.2.1 Influence of New Concept Training Size on Performance

The goal of this Subsection is to see the efficiency of the ECMF surrogate compared with that of the mono-fidelity surrogate in a range of the new concept training set size, specifically from 24 to 288 points. Since both meta-models are probabilistic, two performance indicators are used for the comparison: rms of the concept 2 surrogate "standardized validation error", and rms of concept 2 surrogate "standardized validation residual" given by Equations 84, and 85, respectively. A set of 768 validation points are evaluated to assess the performance indicators. The number of observations from the previous concept is taken to be 20. Figures 52 and 53 show the rms of the concept 2 surrogate "standardized validation error", and rms of concept 2 surrogate "standardized validation residual", respectively.

Again curves are fitted to try to understand the tendency of the results for each

surrogate. For the rms of the concept 2 "standardized validation error" and "standardized validation residual", the fitting curves chosen are Equation 105 and 106, respectively, see Section 5.1.

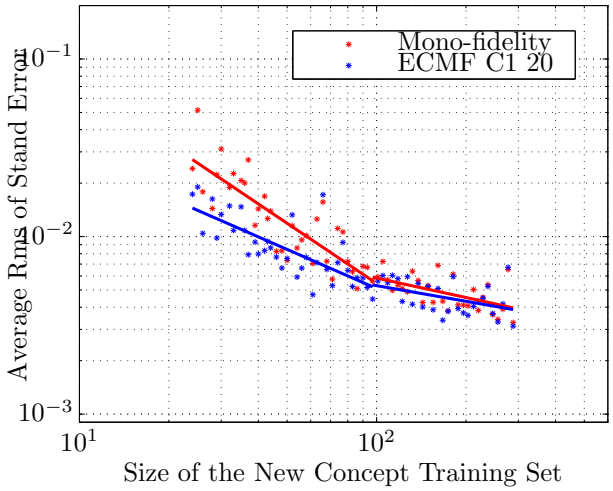


Figure 52: Concept 2 Surrogate "Standardized Validation Error". ECMF vs Mono-Fidelity Surrogate. UH60A with Fenestron shp_{hov}^{fen}

Figure 52 shows that the fenestron concept ECMF surrogate (using 20 previous concept observations) produces a more accurate surrogate than the mono-fidelity surrogate for the small range of new concept sampling sizes.

The reason is that, for low number of new concept points, observations from the previous concept complement the incomplete information from the new concept sampling plan. The ECMF surrogate successfully reuses observations from the previous concept, UH60A with conventional tail, enabling more accurate surrogates.

However, as the new concept training set size becomes larger, the mono-fidelity surrogate gains in performance until it produces a similar accuracy surrogate as the ECMF one; this occurs with new concept training sets of approximately 100 points. It appears to reflect that, when the new concept training set is large, the re-use of observations matters less. It is because the new concept sampling plan provides the necessary features to fit a good surrogate, i.e., previous concept observations could

Table 12: Values of Constant α for the Fitting Curves of the Rms of the "Standardized Validation Error". Mono vs ECMF Surrogate. UH60A with Fenestron shp_{hov}^{fen} Screened Domain

	Low $N_{2,tr.set}$	Large $N_{2,tr.set}$
Mono-fidelity	1.1256	0.3647
ECMF 20	0.7308	0.2927

be useless when the new concept training set is large enough to properly capture the new concept landscape.

The gain in performance of the mono-fidelity surrogate with respect to the ECMF surrogate is again quantified by the values of the constant α in the fitting curves of the rms of the error , see Equation 105. Table 12 contains the values of αs. The mono-fidelity surrogate built with a few new concept observations converges more quickly than the ECMF one. For large $N_{2,tr.set}$, both surrogates converge at a similar speed.

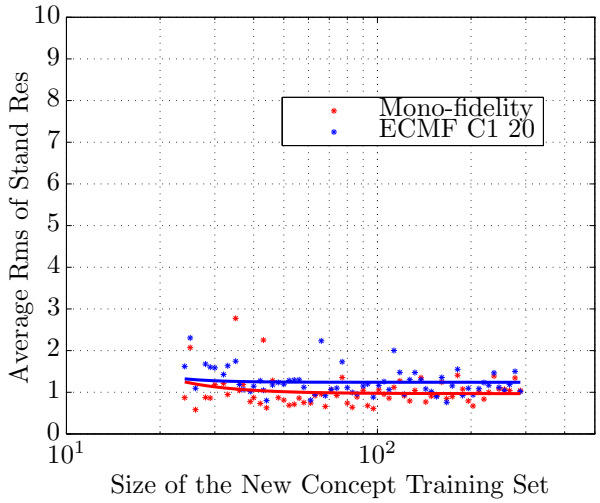


Figure 53: Concept 2 Surrogate "Standardized Validation Residual". UH60A with Fenestron shp_{hov}^{fen}

Figure 53 exhibits the rms of the "standardized validation residual" of the UH60A shp_{hov}^{fen} for the mono-fidelity and the ECMF surrogate. For both surrogates, the fitted curve is close to one. The residual of the ECMF surrogate built with 20 conventional tail UH60A observations is slightly off the ideal value of one.

The "standardized validation error" indicator demonstrates that the ECMF surrogate outperforms the state-of-the-art for small new concept training set sizes when applied on the noisy shp_{hov}^{fen} of the UH60A with fenestron tail.

7.2.2 Influence of Previous Concept Training Size on Performance

The influence of the number of previous concept observations, $N_{1,tr.set}$, in the performance of the ECMF surrogate is studied in this Subsection. A similar study was done in Subsection 7.1.2 for the noise-free function "Michalewicz". For the noisy shp_{hov}^{fen}, the UH60A with conventional tail training sizes tested, $N_{1,tr.set}$, are 12 to 20, 28, and 36.

Figure 54 exhibits shp_{hov}^{fen} rms of the "standardized validation error" for ECMF surrogates built with several values of $N_{1,tr.set}$. This figure exposes similar results as the ones obtained in the free-noise Michalewicz function studied in Section 7.1.

For small values of $N_{2,tr.set}$, the rms of the error decreases as $N_{1,tr.set}$ gets larger. However, this tendency reverses for the high range of $N_{2,tr.set}$, i.e., there exists a tendency that, as $N_{1,tr.set}$ increases, the performance of the surrogate according to the rms of the error is poorer. The ECMF surrogate with $N_{1,tr.set} = 36$ has a poorer performance than the one of the ECMF surrogate with $N_{1,tr.set} = 28$. Same behavior was seen in Subsection 7.1.2.

The reason could be that, as the number of previous concept observations, $N_{1,tr.set}$, increases, the low-fidelity surrogate captures features beyond common trends. This high frequency features can mislead the new concept ECMF surrogate, specially for high values of $N_{2,tr.set}$ where new concept high frequency information is already captured. It results in a decrease in the ECMF performance as old concept observations are added to the ECMF (it occurs for a large $N_{1,tr.set}$). When there are many old concept observations at the designer's disposal, a possible remedy is to build the ECMF surrogate only with a subset of the old concept observations that captures

only similar trends.

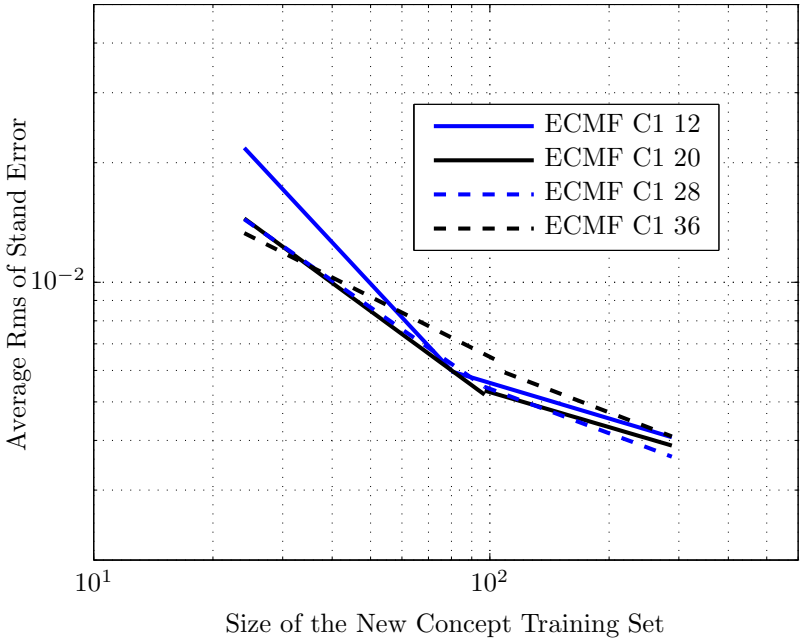


Figure 54: Concept 2 Surrogate "Standardized Validation Error". ECMF Several $N_{1,tr.set}$. UH60A with Fenestron shp^{fen}_{hov}

Table 13 contains the values of αs for all the ECMF surrogates. All the ECMF surrogates gain performance at the same speed for large $N_{2,tr.set}$. It is worth mentioning that for ECMF surrogates with $N_{1,tr.set} = 36$ (the largest previous concept training set) the gain in performance is slower than for other ECMF surrogates with lower $N_{1,tr.set}$ as shown in Table 13, and Figure 54. It reflects that at low $N_{2,tr.set}$ ECMF surrogates with high $N_{1,tr.set}$ are better than their counterpart ECMF with low $N_{1,tr.set}$; however, this tendency reverses when $N_{2,tr.set}$ becomes larger.

Table 13: Values of Constant α for the Fitting Curves of the Rms of the "Standardized Validation Error". All ECMF Surrogates. UH60A with Fenestron shp_{hov}^{fen} Screened Domain

	Low $N_{2,tr.set}$	Large $N_{2,tr.set}$
ECMF $N_{1,tr.set} = 12$	1.0718	0.2975
ECMF $N_{1,tr.set} = 20$	0.7308	0.2927
ECMF $N_{1,tr.set} = 28$	0.6957	0.3719
ECMF $N_{1,tr.set} = 36$	0.5025	0.3813

CHAPTER VIII

DEMONSTRATING MIC ADAPTIVE SAMPLING ON ROTOR-CRAFT PRACTICAL SCENARIO

In this Chapter, the multi-objective adaptive sampling of MIC surrogates is carried out. The purpose is to understand the characteristics of adaptive sampling techniques applied on surrogates that leverage similar trends across categories. The adaptive sampling algorithm is applied to optimize the UH60A power consumption at two flight conditions. The optimization is performed in two design spaces with different dimensionalities.

8.1 Multi-objective Adaptive Sampling of the UH60A Shaft Power. Screened Domain

Before testing the UH60A model on a large design space, some tests are done in a minor design space to better understand the characteristics of the adaptive sampling algorithm on MIC surrogates. The example in these Sections aims to reduce the UH60A power consumption at hover and forward flight with respect to the baseline (for further details of the model, see Section 4.2). The baseline case is presented in Tables 2, and 3, its objective function values are

$$shp_{hov} = 1901.7hp \qquad shp_{fwd} = 1673.9hp$$

.

The design variables are the ones obtained in the screening process, see Subsection 5.2.1: the inner blade twist (continuous variable), the outer blade twist (continuous variable), the chord length (discrete-quantitative variable), and the type of airfoil (categorical variable). Four possible airfoils are available: "NACA 0012", "SC 2110",

"NACA 23012", and "SC 1095". The objective function is summarized as follows:

$$\begin{gathered} shp_{hov}\left(\theta_1, \theta_2, c, \text{airfoil}\right) \\ shp_{fwd}\left(\theta_1, \theta_2, c, \text{airfoil}\right) \end{gathered} \tag{117}$$

subjected to

$$\begin{gathered} \theta_1 \in [-3.4^\circ/m, -0.5^\circ/m] \qquad \theta_1 \in \mathbb{R} \\ \theta_2 \in [-1.5^\circ/m, 0.5^\circ/m] \qquad \theta_2 \in \mathbb{R} \\ \frac{c}{\bar{c}} \in \{0.9 + \left[0, \frac{1}{3}, \frac{2}{3}, 1\right] \cdot 0.85\} \\ \text{airfoil} \in \{\text{NACA 0012}, ..., \text{SC 1095}\} \end{gathered} \tag{118}$$

The remaining parameters are $W = 16994lb$ ($C_T = 5.99 \cdot 10^{-3}$), $R = 26.83ft$, $\Omega = 27.0063\frac{rad}{sec}$, and $\frac{r_{\text{cut-off}}}{R} = 0.047$. Figures 26, 27, 28, and 29 show a regressive meta-model of the hover power consumption for the feasible chords. A meta-model is shown instead of the real function because of the computational expense of the model. The original function is noisy due to the numerical and iterative nature of the software FLIGHTLAB.

In order to find the optimal solutions for both flight conditions, multi-objective optimization is applied. Optimal solutions are given by Pareto optimal points, a set of points that include designs which are so optimized that in order to improve one goal of any of the Pareto members, its performance in at least one of the other goals has to diminish.

First, a good approximation of the Pareto front is calculated by the well-known evolutionary multi-objective algorithm NSGA-II developed by Deb [43]. This non-dominated set is considered herein as the "real" Pareto front. In order to obtain it, the expensive helicopter models have been evaluated 7280 times. The computational intensity of the objective functions limits the number of evaluations and, consequently, the quality of the assessment of the real Pareto front.

The Pareto front obtained with the NSGA-II algorithm is plotted in Figure 55, and 56. The Pareto front is formed of two disconnected fronts as Figure 55 shows. The larger front corresponds to the airfoil "SC 1095", whereas the small one belongs to the airfoil "NACA 0012". Also, the obtained Pareto front from this optimization study is better performing in the two objectives than the baseline, as shown in Figure 55.

The EGO algorithm is applied to two surrogates to determine the non-dominated set that approaches the Pareto front. These surrogates are the following:

MIC Surrogates Surrogates that are developed to leverage similar trends across categories, see Section 3.4.

Independent Surrogates It is the current state-of-the-art method. A surrogate is fitted for each category. The adaptive sampling algorithm is applied simultaneously to these independent surrogates.

The resulting Pareto fronts from both methods are compared with the "real" Pareto front obtained from the evolutionary multi-objective genetic algorithm NSGA-II.

The Pareto fronts after ten updates of the EGO algorithm on surrogates with warm-up training sets of 38 expensive observations are plotted in Figure 55. It shows that after ten updates, the EGO algorithm applied on the MIC meta-model estimates a better approximation to the "real" UH60A Pareto front than the EGO algorithm applied to independent surrogates simultaneously.

After 25 updates, see Figure 56, both EGO processes approach better the "real" Pareto front, but there are twelve non-dominated points in the adaptive sampling on the MIC surrogate over the six in the case of simultaneous adaptive sampling on independent surrogates.

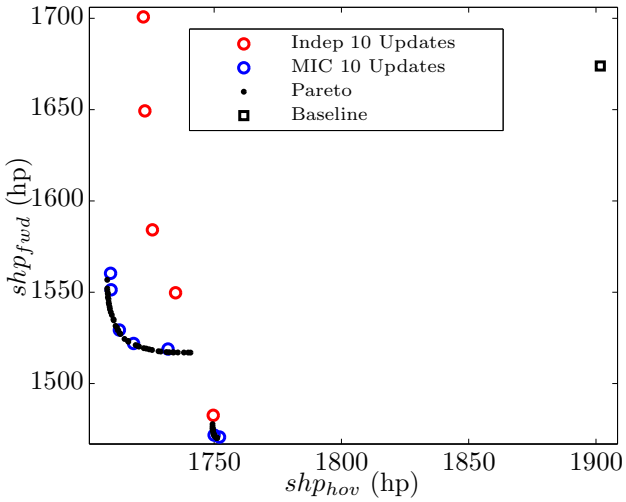


Figure 55: Non-Dominated Set Obtained from EGO Algorithm Applied on MIC and Independent Surrogates. UH60A *shp*. Warm-Up Size 38

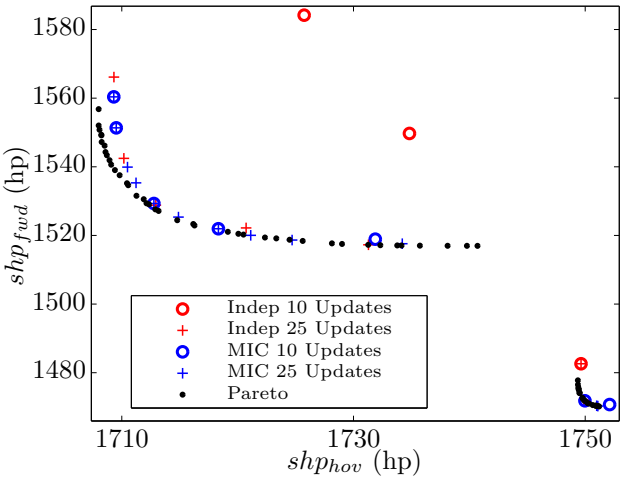


Figure 56: Evolution Non-Dominated Set Obtained from EGO Algorithm Applied on MIC and Independent Surrogates. UH60A *shp*. Warm-Up Size 38

In order to more formally quantify the quality of the non-dominated set for each adaptive sampling method, two performance indicators are employed: the number of points in the non-dominated set, and the average distance between the non-dominated set and "real" Pareto front, see Equation 88. For more details, see Subsection 3.6.1. These two performance indicators are assessed for warm-up datasets of 36, 48, and 66. Figures 57, 58, 59, 60, 61, and 62 show the results.

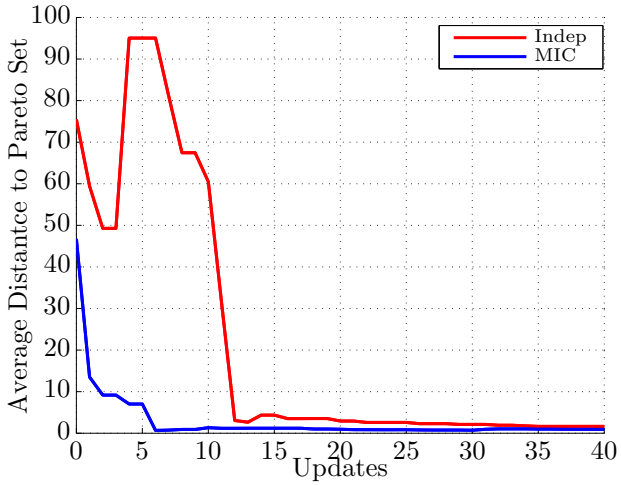


Figure 57: Average Distance to Pareto Set UH60A *shp*. EGO Algorithm Applied on MIC and Independent Surrogates. Warm-Up Size 36. Updates 40

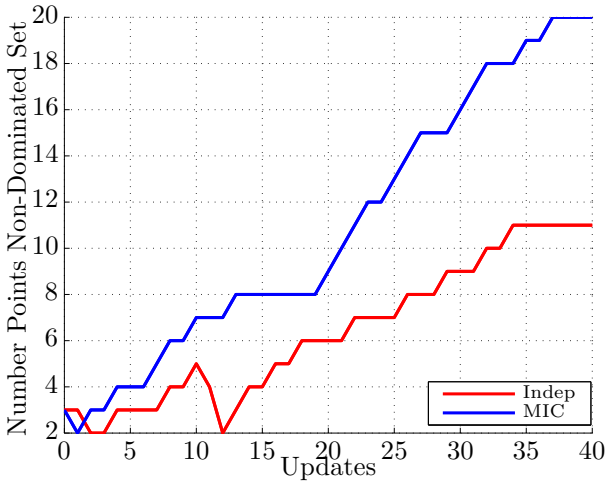


Figure 58: Number Points Non-Dominated Set UH60A *shp*. EGO Algorithm Applied on MIC and Independent Surrogates. Warm-Up Size 36. Updates 40

The EGO runs on a starting surrogate with 36 warm-up observations, results are in Figures 57 and 58, show evidence that the sampling on MIC surrogates provide a much better Pareto approximation for both indicators: the average distance to the "real" Pareto front for MIC surrogates is ten times smaller than in the independent case at five updates; and the number of points in the non-dominated set is around two times larger in the MIC case throughout the whole updating process.

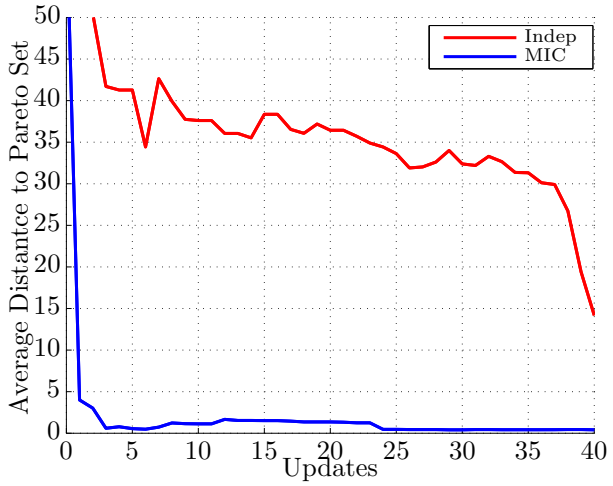


Figure 59: Average Distance to Pareto Set UH60A *shp*. EGO Algorithm Applied on MIC and Independent Surrogates. Warm-Up Size 48. Updates 40

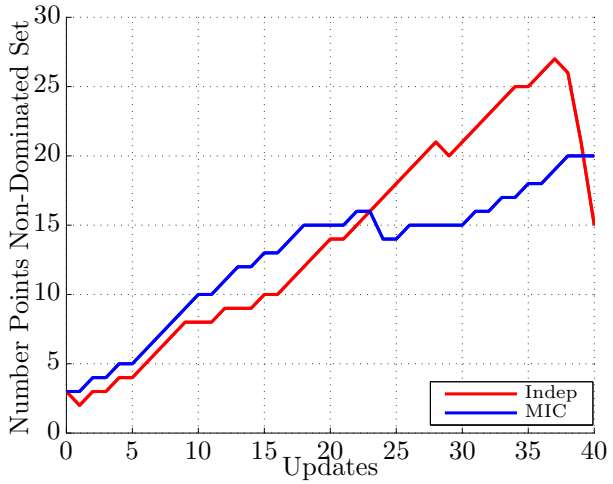


Figure 60: Number Points Non-Dominated Set UH60A *shp*. EGO Algorithm Applied on MIC and Independent Surrogates. Warm-Up Size 48. Updates 40

When the warm-up size is increased to 48, see Figures 59 and 60, results are similar to the ones obtained in the smaller warm-up dataset case. Even though there are more non-dominated points in the independent surrogate case after 23 updates, the average distance to the "real" Pareto set for the independent surrogate case is much larger than for the MIC case throughout the whole updating process (around 30 times larger above 23 updates).

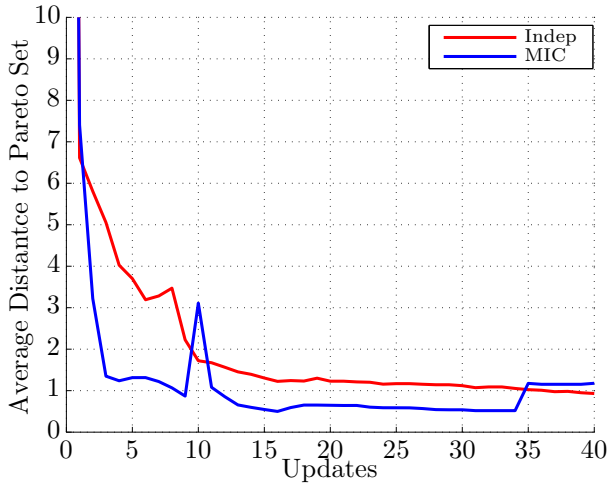


Figure 61: Average Distance to Pareto Set UH60A *shp*. EGO Algorithm Applied on MIC and Independent Surrogates. Warm-Up Size 66. Updates 40

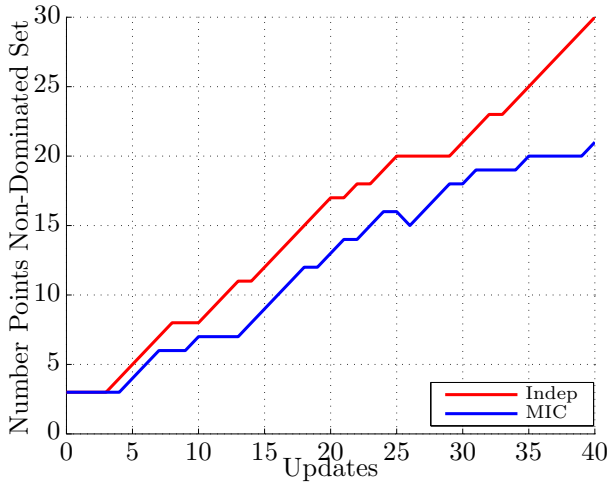


Figure 62: Number Points Non-Dominated Set UH60A *shp*. EGO Algorithm Applied on MIC and Independent Surrogates. Warm-Up Size 66. Updates 40

The final warm-up sets have 66 points. Figure 61 illustrates that the average distance is slightly smaller for the MIC case in the large part of the studied range. However, when it comes to the number of non-dominated points (see Figure 62) the simultaneous EGO on independent surrogates outperforms the EGO on the MIC surrogate beyond 4 updates.

When calculating UH60A Pareto front for hover and forward flight, results demonstrate that EGO algorithm applied on MIC surrogates are more efficient than on independent surrogates for small warm-up training sets. As happened with the surrogate efficiency, the performance of EGO algorithms on MIC surrogates degrades with respect to that of the EGO applied simultaneously on independent surrogates when the training set becomes large.

8.2 Multi-objective Adaptive Sampling of the UH60A Shaft Power. Full Domain

The example in this Section aims to reduce the UH60A power consumption in hover and forward flight even further than in Section 8.1. In order to find even more optimal solutions for both flight conditions, multi-objective adaptive sampling is applied in a larger design space. The MIC surrogate is used. The new variables are: the main rotor radial position where the twist changes r_{tw}, the tail rotor radius R_{tr}, and the tail rotor chord c_{tr}.

The objective functions to optimize are

$$\begin{aligned} &shp_{hov}\left(\theta_1, \theta_2, r_{tw}, R_{tr}, c_{tr}, c, \text{airfoil}\right) \\ &shp_{fwd}\left(\theta_1, \theta_2, r_{tw}, R_{tr}, c_{tr}, c, \text{airfoil}\right) \end{aligned} \tag{119}$$

subjected to the full domain constraints given by

$$\begin{aligned} \theta_1 \in [-3.4^\circ/m, -0.5^\circ/m] \qquad & \theta_1 \in \mathbb{R} \\ \theta_2 \in [-1.5^\circ/m, 0^\circ/m] \qquad & \theta_2 \in \mathbb{R} \\ \frac{r_{tw}}{R} \in [0.72, 0.85] \qquad & \frac{r_{tw}}{R} \in \mathbb{R} \\ \frac{R_{tr}}{\bar{R_{tr}}} \in [0.98, 1.18] \qquad & R_{tr} \in \mathbb{R} \\ \frac{c_{tr}}{\bar{c_{tr}}} \in [0.93, 1.11] \qquad & c_{tr} \in \mathbb{R} \\ \frac{c}{\bar{c}} \in \{0.9 + \left[0, \frac{1}{3}, \frac{2}{3}, 1\right] \cdot 0.85\} & \\ \text{airfoil} \in \{\text{NACA 0012}, ..., \text{SC 1095}\} & \end{aligned} \tag{120}$$

where $\bar{R_{tr}}$, $\bar{c_{tr}}$, $\bar{c}$ are the baseline values, see Table 3.

The starting MIC surrogate warm-up dataset has 1000 observations. Figure 63 contains the Pareto front of the full domain (seven variables) after 600 updates, Pareto front of the reduced domain (four variables), and the baseline values. The

Pareto front of the full domain dominates the Pareto front of the reduced domain as expected. The number of points in the non-dominated set is 200, one third of the number of updates. A close-up of the Pareto front is shown in Figure 64. It is noticeable that the Pareto front corresponding to the airfoil "NACA 0012" is much larger in the full domain than in the domain with lower dimensionality.

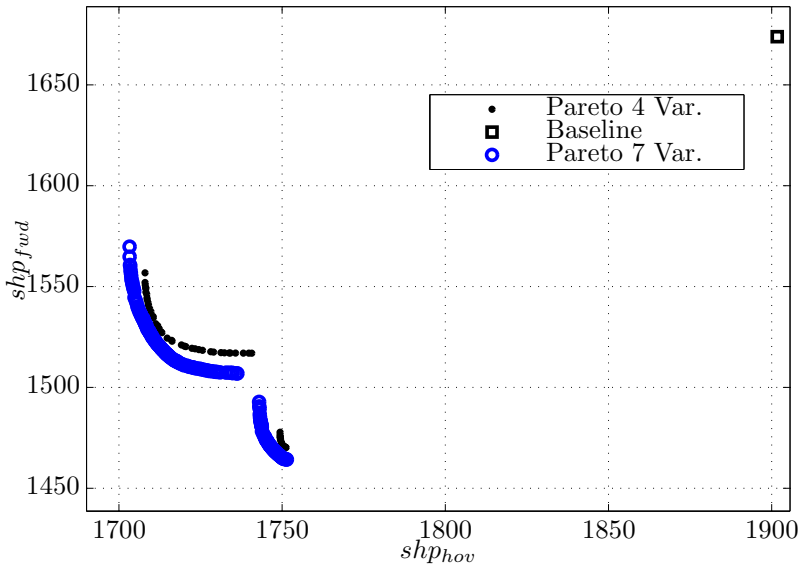


Figure 63: Pareto Fronts of the UH60A with Conventional Tail for the Large and Screened Domain. EGO Algorithm on MIC Surrogates

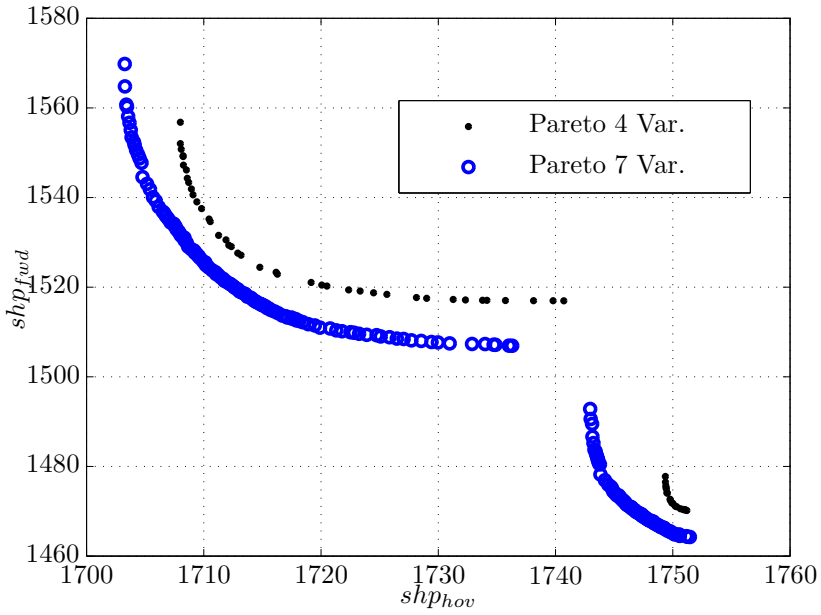


Figure 64: Close-Up of the Pareto Fronts of the UH60A with Conventional Tail for the Large and Screened Domain. EGO Algorithm on MIC Surrogates.

CHAPTER IX

DEMONSTRATING ECMF ADAPTIVE SAMPLING ON ROTOR-CRAFT PRACTICAL SCENARIO

In this Chapter, the multi-objective adaptive sampling of ECMF surrogates is carried out. The one-objective optimization of the ECMF surrogate is skipped because it is not a technical challenge. This is because the ECMF is a stochastic surrogate whose domain can be made of continuous, integer and/or categorical. The most demanding scenario is the adaptive sampling of a meta-model with a mixed-integer-categorical domain that was already carried out in Chapter 6. Thus, this work moves directly to practical conceptual design applications without looking at canonical problems.

The research purpose of this Chapter is to understand the characteristics of the adaptive sampling process on the ECMF surrogates; thus, for simplicity's sake the design domain is made of just continuous and discrete-quantitative variables but not categorical ones. The idea is to avoid mixing the two different methodologies that are present in this work: the cross-use of intense observations across categorical alternatives and the reuse of intense observations from a previous concept.

The adaptive sampling algorithm is applied to optimize the UH60A with fenestron tail power consumption at two flight conditions. The optimization is performed in two design spaces with different dimensionalities.

9.1 Multi-objective Adaptive Sampling of the UH60A with Fenestron Shaft Power. Screened Domain

The example in these Sections aims to find optimal UH60A power consumption in hover and forward flight for the fenestron configuration case. Optimal designs of the UH60A fenestron configuration are compared with baseline values and the optimal

designs of the UH60A with regular tail configuration (explored in Chapter 8). The baseline case is presented in Tables 2, and 3. As proceeded in previous sections, the optimal solution is given in terms of Pareto fronts. The baseline objective function values are

$$shp_{hov} = 1901.7hp \qquad shp_{fwd} = 1673.9hp$$

The objective functions to optimize are given by Equations 121

$$\begin{aligned} &shp_{hov}^{fen}\left(\theta_1, \theta_2, N_{b,tr}, c\right) \\ &shp_{fwd}^{fen}\left(\theta_1, \theta_2, N_{b,tr}, c\right) \end{aligned} \tag{121}$$

subjected to constraints given by Equations 122.

$$\begin{aligned} \theta_1 \in &[-3.4^\circ/m, -0.5^\circ/m] \in \mathbb{R} \\ \theta_2 \in &[-1.5^\circ/m, 0.5^\circ/m] \in \mathbb{R} \\ N_{b,tr} \in &7, 8, 9 \\ \frac{c}{\bar{c}} \in &0.9 + \left[0, \frac{1}{3}, \frac{2}{3}, 1\right] 0.85 \end{aligned} \tag{122}$$

The remaining parameters are $W = 16994lb$ ($C_T = 5.99 \cdot 10^{-3}$), $R = 26.83ft$, $\Omega = 27.0063\frac{rad}{sec}$, and $\frac{r_{\text{cut-off}}}{R} = 0.047$. The main rotor blade section is "SC 1095".

The previous concept that will help to capture the functions shp_{hov}^{fen} and shp_{fwd}^{fen} is the UH60A with conventional tail. Its corresponding objectives shp_{hov}^{conv} and shp_{fwd}^{conv} are defined in Equations 117; and subjected to Equations 118 with the only difference that the airfoil "SC 1095" is the only one available for the designer.

Notice that the design space of the two concepts, $\left(shp_{hov}^{fen}, shp_{fwd}^{fen}\right)$ and $\left(shp_{hov}^{conv}, shp_{fwd}^{conv}\right)$, are different. The UH60A with the conventional tail has one less design variable than the UH60A with fenestron tail. When building the ECMF surrogate, the previous concept behavior along the variable $N_{b,tr}$ is assumed to be constant, as explained in Section 3.5.

As done for multi-objective adaptive sampling on MIC surrogates, see Section 8.1, a good approximation of the Pareto front is calculated by the well-known evolutionary multi-objective algorithm NSGA-II developed by Deb [43]. This approximation set is considered herein as the "real" Pareto front. In order to obtain it, the computationally expensive helicopter model has been evaluated 3380 times. The computational intensity of the objective functions limits the number of evaluations and, consequently, the quality of the assessment of the real Pareto front.

The Pareto front obtained with the NSGA-II algorithm is plotted in Figure 65. It is seen that the obtained Pareto front from this optimization study is better performing in the two objectives than the baseline. Also, the Pareto front for the new concept, UH60A with fenestron tail, dominates the one obtained for the conventional tail configuration.

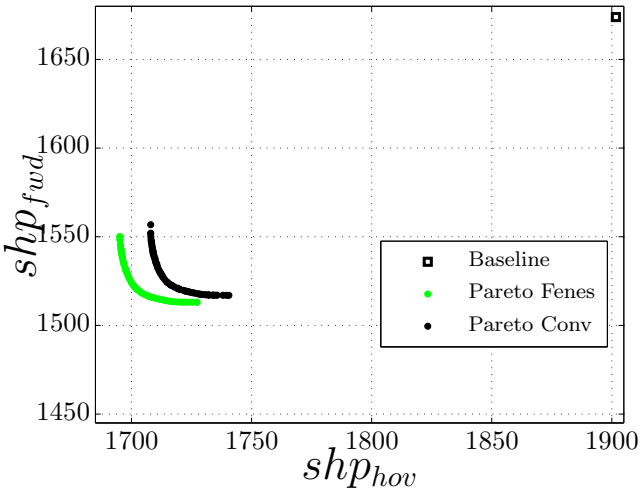


Figure 65: "Real" Pareto Front Obtained from NSGA-II. UH60A with Fenestron *shp*. Airfoil "SC 1095"

The EGO algorithm is applied to two surrogates to determine the non-dominated set that approaches the Pareto front. These two surrogates are the following:

ECMF Surrogates Surrogates that are developed to leverage similar trends of incrementally evolved concepts, see Section 3.5.

Mono-fidelity Surrogates It is the current state-of-the-art method. A surrogate is fitted for the new concept with no reuse of observations from previous concepts.

The resulting Pareto fronts from both methods are compared with the "real" Pareto front obtained from the evolutionary multi-objective genetic algorithm NSGA-II.

The Pareto fronts after 15 updates of the EGO on warm-up training sets of 66 new concept observations and 20 old concept observations are plotted in Figure 66. It shows that, after 15 updates, the adaptive sampling algorithm on the ECMF meta-model estimates a better approximation to the "real" fenestron UH60A Pareto front than the adaptive sampling algorithm applied on the mono-fidelity surrogate.

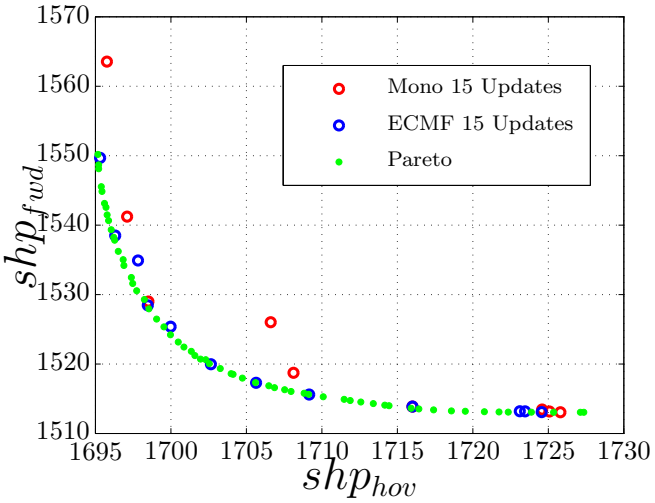


Figure 66: Non-Dominated Set Obtained from EGO Algorithm Applied on ECMF and Mono-Fidelity Surrogates. UH60A with Fenestron *shp*. New Concept Warm-Up Size 66. Reuse of 20 Old Concept Observations

After 30 updates, see Figure 67, both EGO processes approach better the "real" Pareto front, but there are still more points and closer to the "real" Pareto front in the ECMF case than in the state-of-the-art case.

Again, the number of points in the non-dominated set, and the average distance between the non-dominated set and the "real" Pareto front, see Equation 88, are

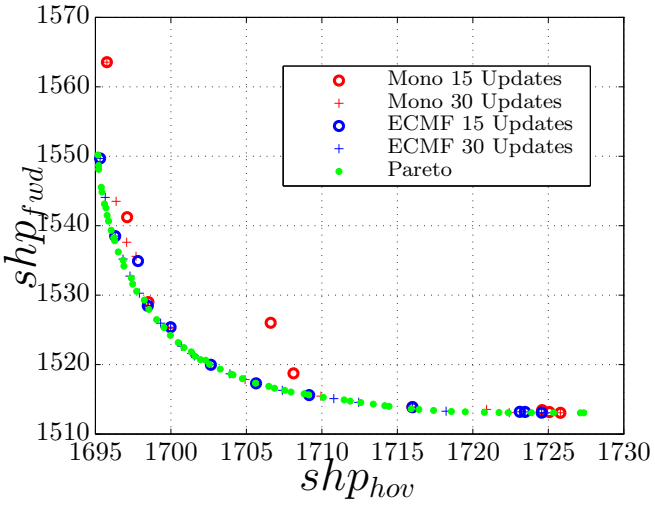


Figure 67: Evolution Non-Dominated Set Obtained from EGO Algorithm Applied on ECMF and Mono-Fidelity Surrogates. UH60A with Fenestron *shp*. New Concept Warm-Up Size 66. Reuse of 20 Old Concept Observations

brought to quantify the quality of the non-dominated set for each method, as done in Section 8.1. These two performance indicators are assessed for warm-up sets of 66, 120, and 180 observations; they are shown in Figures 68, 69, 70, 71, 72, and 73.

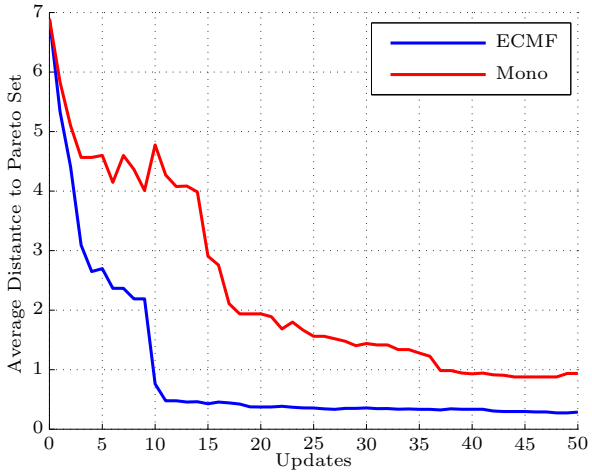


Figure 68: Average Distance to Pareto set UH60A with Fenestron *shp*. EGO Algorithm Applied on ECMF and Mono-Fidelity Surrogates. New Concept Warm-Up Size 66. Reuse of 20 Old Concept Observations. Updates 40

The EGO runs on starting surrogates with 66 warm-up points, Figures 68 and 69,

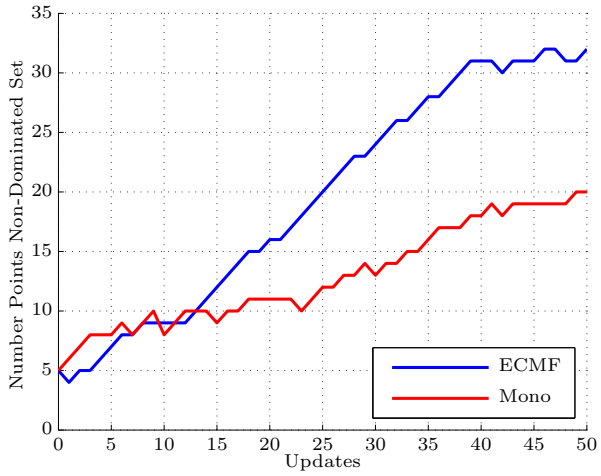


Figure 69: Number Points Non-Dominated Set UH60A with Fenestron *shp*. EGO Algorithm Applied on ECMF and Mono-Fidelity Surrogates. New Concept Warm-Up Size 66. Reuse of 20 Old Concept Observations. Updates 40

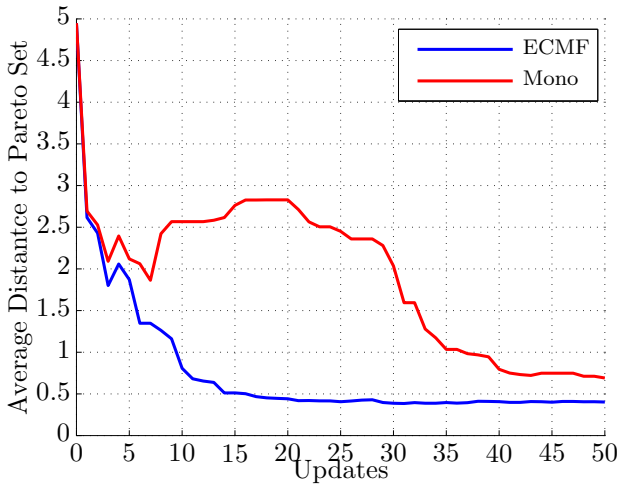


Figure 70: Average Distance to Pareto set UH60A with Fenestron *shp*. EGO Algorithm Applied on ECMF and Mono-Fidelity Surrogates. New Concept Warm-Up Size 120. Reuse of 20 Old Concept Observations. Updates 40

show evidence that ECMF surrogates provide a much better Pareto front approximation for both indicators: the average distance is up to eight times smaller for ECMF surrogates at 11 updates, and the number of points in the non-dominated is close to twice as large above the 25th update.

When the warm-up size is increased to 120, see Figures 70 and 71, results are

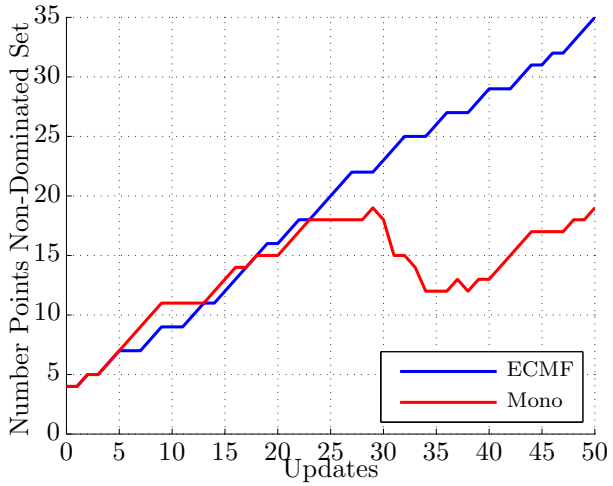


Figure 71: Number Points Non-Dominated Set UH60A with Fenestron *shp*. EGO Algorithm Applied on ECMF and Mono-Fidelity Surrogates. New Concept Warm-Up Size 120. Reuse of 20 Old Concept Observations. Updates 40

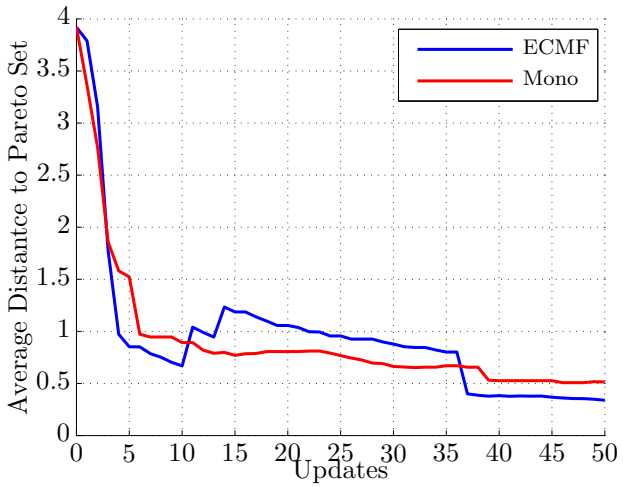


Figure 72: Average Distance to Pareto set UH60A with Fenestron *shp*. EGO Algorithm Applied on ECMF and Mono-Fidelity Surrogates. New Concept Warm-Up Size 180. Reuse of 20 Old Concept Observations. Updates 40

similar to the ones obtained in the warm-up training sets with 66 observations. However, one can see that the quality of the Pareto front obtained from the mono-fidelity surrogate is competitive with the ECMF one during the first updates.

Results for the largest warm-up training sets, 180 observations, are shown in Figures 72 and 73. They demonstrate that the two Pareto fronts have similar quality.

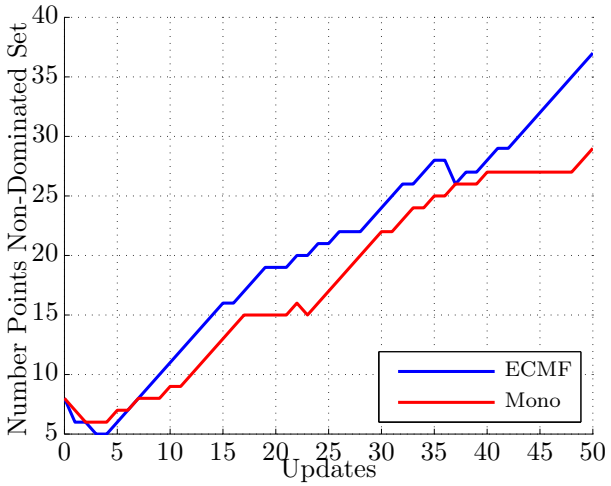


Figure 73: Number Points Non-Dominated Set UH60A with Fenestron *shp*. EGO Algorithm Applied on ECMF and Mono-Fidelity Surrogates. New Concept Warm-Up Size 180. Reuse of 20 Old Concept Observations. Updates 40

It shows evidence that the performance of EGO algorithms on ECMF surrogates degrades with respect to that of EGO algorithm applied on mono-fidelity surrogates when the new concept training set becomes large, as happened with the surrogate efficiency (see Chapter 7).

When calculating the UH60A Pareto front for hover and forward flight, results show evidence that EGO algorithms applied on ECMF surrogates are more efficient than on mono-fidelity surrogates for the small range of new concept warm-up training sets.

9.2 Multi-objective Adaptive Sampling of the UH60A with Fenestron Shaft Power. Full Domain

The example in this Section aims to reduce the power consumption in hover and forward flight of the UH60A with fenestron tail even further than in Section 9.1. The ECMF surrogate is used. A large design space is explored where the new variables are: the radial position of the twist change, the fenestron radius, and the fenestron chord.

The objective functions to optimize are

$$\begin{aligned} &shp_{hov}^{fen}\left(\theta_1, \theta_2, r_{tw}, R_{tr}, c_{tr}, N_{b,tr}, c\right) \\ &shp_{fwd}^{fen}\left(\theta_1, \theta_2, r_{tw}, R_{tr}, c_{tr}, N_{b,tr}, c\right) \end{aligned} \tag{123}$$

subjected to the full domain constraints given by Equations 126.

$$\begin{aligned} \theta_1 \in [-3.4^\circ/m, -0.5^\circ/m] \qquad & \theta_1 \in \mathbb{R} \\ \theta_2 \in [-1.5^\circ/m, 0^\circ/m] \qquad & \theta_2 \in \mathbb{R} \\ \frac{r_{tw}}{R} \in [0.72, 0.85] \qquad & \frac{r_{tw}}{R} \in \mathbb{R} \\ R_{tr} \in [0.8m, 1m] \qquad & R_{tr} \in \mathbb{R} \\ c_{tr} \in [0.14m, 0.2m] \qquad & c_{tr} \in \mathbb{R} \\ N_{b,tr} \in 7, 8, 9 & \\ \frac{c}{\bar{c}} \in \{0.9 + \left[0, \frac{1}{3}, \frac{2}{3}, 1\right] \cdot 0.85\} & \end{aligned} \tag{124}$$

The remaining parameters are $W = 16994lb$ ($C_T = 5.99 \cdot 10^{-3}$), $R = 26.83ft$, $\Omega = 27.0063\frac{rad}{sec}$, and $\frac{r_{\text{cut-off}}}{R} = \frac{1}{3}$. The main rotor blade section is again "SC 1095".

The previous concept that helps to capture the functions shp_{hov}^{fen} and shp_{fwd}^{fen} is the UH60A with conventional tail. Its corresponding objectives shp_{hov}^{conv} and shp_{fwd}^{conv} are defined as

$$\begin{aligned} & shp_{hov}^{conv}\left(\theta_1, \theta_2, r_{tw}, c\right) \\ & shp_{fwd}^{conv}\left(\theta_1, \theta_2, r_{tw}, c\right) \end{aligned} \tag{125}$$

subjected to

$$\begin{aligned} \theta_1 \in [-3.4^\circ/m, -0.5^\circ/m] \qquad & \theta_1 \in \mathbb{R} \\ \theta_2 \in [-1.5^\circ/m, 0^\circ/m] \qquad & \theta_2 \in \mathbb{R} \\ \frac{r_{tw}}{R} \in [0.72, 0.85] \qquad & \frac{r_{tw}}{R} \in \mathbb{R} \\ \frac{c}{\bar{c}} \in \{0.9 + \left[0, \frac{1}{3}, \frac{2}{3}, 1\right] \cdot 0.85\} & \end{aligned} \tag{126}$$

The same blade section, "SC 1095", is used for the previous concept. The remaining previous concept parameters are the same as for the UH60A with fenestron tail: $W = 16994lb$ ($C_T = 5.99 \cdot 10^{-3}$), $R = 26.83ft$, $\Omega = 27.0063\frac{rad}{sec}$, and $\frac{r_{\text{cut-off}}}{R} = 0.047$.

Notice that the design space of the two concepts, $\left(shp_{hov}^{fen}, shp_{fwd}^{fen}\right)$ and $\left(shp_{hov}^{conv}, shp_{fwd}^{conv}\right)$, are different. The UH60A with fenestron tail has three more design variables than the UH60A with conventional tail. These variables are R_{tr}, c_{tr}, and $N_{b,tr}$. Therefore, when building the ECMF surrogate, the previous concept behavior along these three variables is assumed to be constant, as discussed in Section 3.5.

The starting ECMF surrogate warm-up size has 1000 observations and 40 observations of the UH60A with conventional tail are reused when building the ECMF surrogate. Figure 74 contains the Pareto front of the full domain (seven variables) after 400 updates, the Pareto front of the screened domain (four variables), and the Pareto front of the UH60A with conventional tail (seven variables) calculated in Section 8.2.

The seven variable domain Pareto front of the UH60A with fenestron tail dominates the Pareto front of the reduced domain as expected. The number of points in the non-dominated set is 60, around one seventh of the 400 updates. The seven variable domain Pareto front of the UH60A with fenestron tail also dominates the

counterpart Pareto front of the UH60A with conventional tail. It means that the fenestron tail provides a better design than that of the conventional tail according to the power consumption at hover and cruise speed. In other words, the benefits obtained from the increase in efficiency of the fenestron tail are more important than the weight penalty of the fenestron tail with respect to the conventional tail.

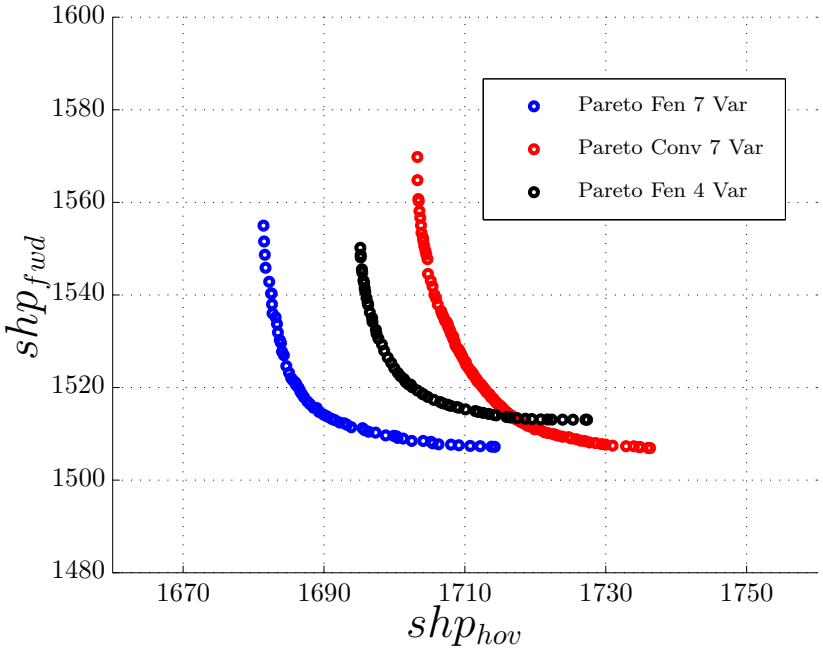


Figure 74: Pareto Front of the UH60A with Fenestron for the Large and Screened Domain. EGO Algorithm Applied on ECMF Surrogates

CHAPTER X

SUMMARY, CONTRIBUTIONS, AND RECOMMENDATIONS

This final Chapter first summarizes the methodologies and experiments carried out to test the proposed methods. Then, a review of the contributions is made, comparing them with the current state-of-the art. It is followed by a list of recommendations to use the surrogates MIC and ECMF, and the presented adaptive sampling algorithms on them. Finally, the possible future work is proposed with the purpose of widening the scenarios where the developed methodologies can be used and making them more suitable for real engineering design situations.

10.1 Summary

Meta-models that identify and apply similar trends across categories and concepts have been constructed. They are intended to accomplish the purpose of the present research: developing new conceptual design tools for the more efficient use of computationally expensive codes with the purpose of improving the quality of initial designs and concept/alternative selection. Results on noise-free canonical problems and noisy UH60A models indicate that, for surrogate modeling and surrogate based-sampling, the surrogates MIC and ECMF are more efficient than traditional independent surrogates in the low range of training set sizes. The principal conclusions of this work are:

1) New meta-models, called MIC, are proposed to approximate computationally expensive models when there is a choice of categories. MIC surrogates aim to leverage similar trends across categories and apply them in the early design

phase. These meta-models contain not only continuous and integer variables but also non-numeric ones. They use a nominal distance to include the categories as a new variable in their domain. MIC surrogates can be constructed with two nominal distances: Hamming distance and intrinsic distance when its definition is possible. These meta-models are an alternative to the current state-of-the-art that uses independent surrogates, one for each category.

2) When modeling the UH60A hover power consumption, MIC surrogates are shown to be more "efficient" for small training sets than independent modeling of the rotor-craft for each airfoil. Same results are obtained for the canonical disturbed Branin function. "Efficiency" means fewer observations are needed for a given accuracy or more accuracy for the same number of intense observations. The second interpretation leads to the conclusion that MIC surrogates are a major step in order to better use computationally intense codes in early design of problems with a categorical choice. In other words, MIC surrogate is a conceptual design tool that, for a given intense function call budget, provides better accuracy than the current state-of-the-art surrogates. MIC meta-modeling is shown to be a tool capable of identifying and leveraging the similarities across categories, leading to better initial designs and categorical selection.

3) In the UH60A hover power consumption example, MIC efficiency degrades with respect to that of the independent UH60A surrogate for each airfoil as the training set becomes large. This behavior can be explained by the fact that as the number of expensive observations increases, high frequency information for each category is available. These high frequency features are different for each airfoil, so the cross-use of observations across airfoils, done by MIC, is no longer efficient. Similar results are found in the test on the canonical disturbed Branin function. The small range of the training set where the MIC surrogate is

more efficient than the independent surrogate depends on how similar the trends are between categories, the dimension of the design space, and the number of categories. For the specific tested problems, these ranges are plotted in Figures 21, and 30.

4) A MICGA is developed to search a domain with continuous, integer, and categorical variables. This stochastic algorithm is used to drive the adaptive sampling process on the MIC surrogate.

5) An EGO algorithm is run on the MIC surrogate via MICGA to adaptively sample the UH60A hover performance. Several EGO processes are performed and it is observed that in all cases a large subset of the updates are in the globally high-performing areas (best designs of best airfoils). It is a characteristic of EGO optimizers. All the EGO runs show an increase in hover performance of around 10% respect the baseline case.

6) New meta-models, called ECMF, are proposed to approximate computationally expensive models when there are sequential concepts. ECMF surrogates aim to leverage similar trends from previously sampled concepts and apply them in the early design phase. "Evolutionary incremental concepts" are assumed; they result from the small changes in concepts done by designers while trying to meet requirements in conceptual design. Multi-fidelity techniques are brought to feed observations from one concept to another in an innovative way. These meta-models are an alternative to the current state-of-the-art that uses independent surrogates for each concept.

7) When modeling the hover power consumption of the UH60A with fenestron tail, ECMF surrogates are shown to be more "efficient" than the mono-fidelity surrogates for small new concept training sets, where no reuse of expensive observations is done. Same results are obtained for the canonical Michalewicz

function. ECMF surrogates are a major step in order to better use the computationally intense codes in early design of problems where there is a sequence of concepts. ECMF meta-modeling is shown to be a tool capable of identifying and leveraging the similarities from previous concepts, leading to better initial designs and concept selection.

8) In the example of the hover power consumption of the UH60A with fenestron tail, ECMF loses efficiency compared to the mono-fidelity surrogate as the new concept training set becomes large. The reason could be that previous concept observations could be useless or misleading when the new concept training set is large enough to properly capture the new concept landscape.

   Also, ECMF efficiency saturates or even decreases as the old concept training set becomes large. This behavior could be explained by the fact that, as the number of old concept expensive observations increases, high frequency information from the previous concept is available. These high frequency features could be different from one concept to another. So, the reuse of observations across concepts, done by ECMF, is no longer efficient. Similar results are found in the test on the canonical Michalewicz function.

   The small range of the new concept training set where the ECMF surrogate is more efficient than the mono-fidelity surrogate depends on how similar the trends between concepts are, the dimension of the design space, and the number of old concept observations reused. For the specific tested problems, these ranges are plotted in Figures 52, and 45.

9) A multi-objective EGO algorithm is run on the MIC surrogate via MICGA to optimize the UH60A hover and forward flight power consumption. The resulting Pareto front is compared with that obtained from the simultaneous

multi-objective EGO algorithm on independent surrogates, one for each airfoil section. The indicators demonstrate again that EGO algorithms on MIC meta-models are more "efficient" to assess the Pareto front of the UH60A in the low training set range. The updates in the MIC case are located in the high-performing areas of the best airfoils earlier than in the current state-of-the-art case. Multi-objective adaptive sampling on MIC surrogates allows a better use of the computationally intense tools in rotor-craft conceptual design because they provide more accurate Pareto fronts given a intense function call budget, which again leads to better initial designs and airfoil selection.

10) A multi-objective EGO algorithm is run on the ECMF surrogate to optimize the hover and forward flight power consumption of the UH60A with a fenestron tail. The resulting Pareto front is compared with that obtained from the multi-objective EGO algorithm on mono-fidelity surrogates. The indicators demonstrate again that EGO algorithms on ECMF meta-models are more "efficient" to assess the Pareto front of the UH60A with fenestron tail in the low training set range. The updates in the ECMF case are located in the high-performing designs earlier than in the current state-of-the-art case. Multi-objective adaptive sampling on ECMF surrogates allows a better use of computationally intense tools in rotor-craft conceptual design because they provide more accurate Pareto fronts given a intense function call budget, which again leads to better initial designs and concept selection.

10.2 Contributions

The MIC and ECMF surrogates and their adaptive sampling algorithms provide new scholarly contributions. Firstly, a new efficient treatment is proposed for building surrogates of concepts when observations show similar trends across several categories. In the literature one can only find surrogates for mixed-integer variables [153, 92].

The use of these proposed meta-models results in more accurate surrogates for a given number of function calls. It represents an alternative to the current independent surrogates, one for each category. This alternative is more efficient for small training sets. It is a major step in order to better use computationally burdensome codes in early design of problems with categorical choices.

The second novelty is the adaptive sampling of MIC surrogates, the ones of the first contribution. A mixed-variable genetic algorithm, MICGA, that searches a mixed-integer-categorical design space at one algorithm call is developed to perform this task. Adaptive Kriging on mixed-integer domains is done [113]; however, no adaptive sampling on MIC domains is found in the literature. This contribution allows adaptive sampling on the whole design space of a surrogate that cross-uses observations across categories, whereas currently, all the independent surrogates (with no cross-use of observations) are adaptively sampled simultaneously that results in a waste of resources when sampling in poor performing categories occurs. Also, the MICGA could be considered as a minor contribution: it successfully optimizes the expected improvement in a mixed-integer-categorical domain. However, one can find in the literature evolutionary strategies capable of searching MIC domains in the field of medical image analysis [127].

The third contribution is a methodology that leverages computationally expensive observations from a previously sampled concept in the construction of a new concept surrogate. In the literature one can find Gaussian meta-models of concepts using variable-fidelities that lay on the same design space [79]. Also, trust-region model management for variable parameterization design spaces has been implemented but only for the same concept [155]. However, the proposed ECMF surrogate deals with two different concepts with different design space. With this third contribution, observations from each concept are employed to build, not only their own concept surrogate, but also other concept surrogates that experience similar trends. It opposes

to the current state-of-the-art that fits independent surrogates for each concept with no reuse of the previous concept intense observations. Again, ECMF meta-models for a given number of function calls are more accurate than mono-fidelity surrogates for small training sets. It is due to the useful information brought from previously sampled concepts. It is a major step in order to better use computationally burdensome codes in early design of problems with sequential concepts.

10.3 Recommendations

In this Section recommendations of the proposed methodologies are discussed.

10.3.1 MIC Surrogate

MIC surrogates are a great tool to make a better use of computationally expensive tools in conceptual design. The recommendations for their use are based on the assumptions and the results obtained in the computer experiments carried out in Chapters 5, 6, and 8. The recommendations are presented in the following enumeration:

1) Designers must choose between several categorical alternatives for the same concept.

2) There are similar trends across the categories along the design variables.

3) Desire of a parametric study of each category design space. It is pursued the exploration of the each category design space and comparison of optimal designs across categories.

4) MIC surrogates must be used in early design when the designers' focus is to reveal important trends, interactions, and sensitivities.

5) MIC surrogates must be employed when the function budget is limited due to: the use of computationally expensive tools, presence of many categorical

alternatives, or existence of large design spaces to explore.

10.3.2 ECMF Surrogate

ECMF surrogates are a great tool to better use of computationally expensive tools in conceptual design. The recommendations for their use are based on the assumptions and the results obtained in the computer experiments carried out in Chapters 7, and 9. The recommendations are presented in the following enumeration:

1) Sequential concepts where observations from previous concepts are available.

2) The new concept is an incremental change of the previous one. So similar trends are expected for the previous and new concept.

3) Desire of a parametric study of each concept. It is pursued by the exploration of the each concept design space and comparison with other concept optimal designs.

4) ECMF surrogates must be used in early design when the designers' focus is to reveal important, trends, interactions and sensitivities.

5) ECMF surrogates must be employed when the function budget is limited due to: the use of computationally expensive tools, presence of many concepts, or existence of large design spaces to explore.

6) Do not use many previous concept observations because it can capture high frequency information that usually differs across concepts.

10.4 Future Work

The future work aims to extend the developed techniques to scenarios broader than the research motivation. Also, the proposed methodologies can spin-off new scenarios that are not directly related to the main research motivation, but where the methodologies can be applied. The next enumeration proposes possible future paths:

1) Take the idea of the cross-use and reuse of computationally expensive observations to deterministic surrogates. See how effective the cross-use and reuse of information are in the deterministic case.

2) Investigate on possible methods to select the number and location of the samples to cross-use and reuse. It would help to increase the performance of the MIC and ECMF surrogates.

3) Build surrogates that cross-use computationally expensive observations across concepts (the proposed ECMF surrogate only reuses observations from previously sampled concepts).

4) Extend ECMF surrogates to less stiff design space assumptions. Space mapping could be a perfect candidate to build ECMF surrogates where the concept 1 design space is not necessarily included in the concept 2 design space.

5) Extend ECMF surrogates to scenarios where observations from two or more previous concepts are available.

6) Do a sensitivity study to understand the influence of the number of categories on the MIC surrogate performance (scalability).

7) Combine ECMF and MIC surrogates and see how the cross-use of observations across categories and the reuse of observations from previous concepts affect each other.

8) Formally develop the MICGA optimizer. Perform tests on canonical problems of different modality.

APPENDIX A

DISTURBED BRANIN FUNCTION

The Branin function is a two-variable function given by Equation 127

$$f(X_1, X_2) = \left(X_2 - \frac{5.1}{4\pi^2}X_2 + \frac{5}{\pi}X_1 - 6\right)^2 + \left[\left(1 - \frac{1}{8\pi}\right) cos(X_1) + 1\right] + 5X_1 \quad (127)$$

where the domain is $X_1 \in [-5, 10]$ and $X_2 \in [0, 15]$.

First of all, the original function is made dimensionless so that non-bias is produced when the Kriging surrogate is fitted. The new independent variables are

$$x_1 = \frac{X_1 + 5}{15} \qquad x_2 = \frac{X_2}{15} \quad (128)$$

The intention is to transform it to a test function on a mixed-integer-categorical design space, called disturbed Branin function. The purpose of this function is to test MIC surrogates. While including a new categorical variable to the Branin function some requirements must be present. MIC surrogates are intended for functions with similar trends across categories, so the modality of the all disturbed Branin function categories must be similar to the original Branin function.

The method proposed to obtain disturbed Branin function consists in calculating the Fourier series of the original Branin function and alter randomly the 9 first Fourier coefficients. Dym and McKean [49], among many others, provide information about the Fourier transform.

The amplitudes and phases of the 9 highest modes of the Fourier transform are randomly modified by two normal distributions according to the following formula:

$$\hat{f\prime}(n_1, n_2) = \hat{f}(n_1, n_2) \exp\{i \cdot rnd_1\} (1 + 0.05 \cdot rnd_2) \quad (129)$$

where $\hat{f}$ and $\hat{f}\prime$ are the Fourier coefficient of the original Branin function and a category of the disturbed Branin function, respectively; n_i is the Fourier frequency in the i coordinate dimension; and rnd_i is a standardized normally distributed random number $\mathcal{N}(\mu = 0, \sigma^2 = 1)$. The second factor of the right hand side, $\exp\{i \cdot rnd_1\}$, is a change of phase of $\frac{rnd_1}{2\pi}$ radians in the Fourier coefficient, whereas the last factor $(1 + 0.05 \cdot rnd_2)$ represents a change of $0.05 \cdot rnd_2$ in the Fourier coefficient amplitude. Notice that to avoid large changes in the trends of the Branin function, the amplitude of the Fourier coefficient is limited to a typical change of 5%, and that of the phase is 9.11 degrees.

Several disturbed Branin function categories could be built by generating several sets of random variables (rnd_1, rnd_2). Each pair of (rnd_1, rnd_2) is considered a category.

Five pairs of random numbers are produced and their corresponding disturbed Branin function categories are compared with the original Branin function in Figure 75. Contours in Figure 75 show that there are similar trends in the disturbed Branin function categories, but the depth and location of the valleys varies.

Once the six categories are created (the first category is the original Branin function, i.e., $rnd_1 = 0$, and $rnd_2 = 0$), a further modification is made to finally convert the design space into a mixed-integer-categorical one. The x_2 coordinate is divided into 11 equally spaced points $\left[0, \frac{1}{10}, \frac{2}{10}, ...1\right]$. It makes the x_2 a discrete-quantitative design input. x_1 is kept as a continuous variable.

Table 14 compares the value of the global minimum and its location for the disturbed Branin function categories. Notice that the original Branin function is considered category 1 of the disturbed Branin function.

Location and value of categorical minimum changes across disturbed Branin function categories; however, they experience similar trends as depicted in Figure 75.

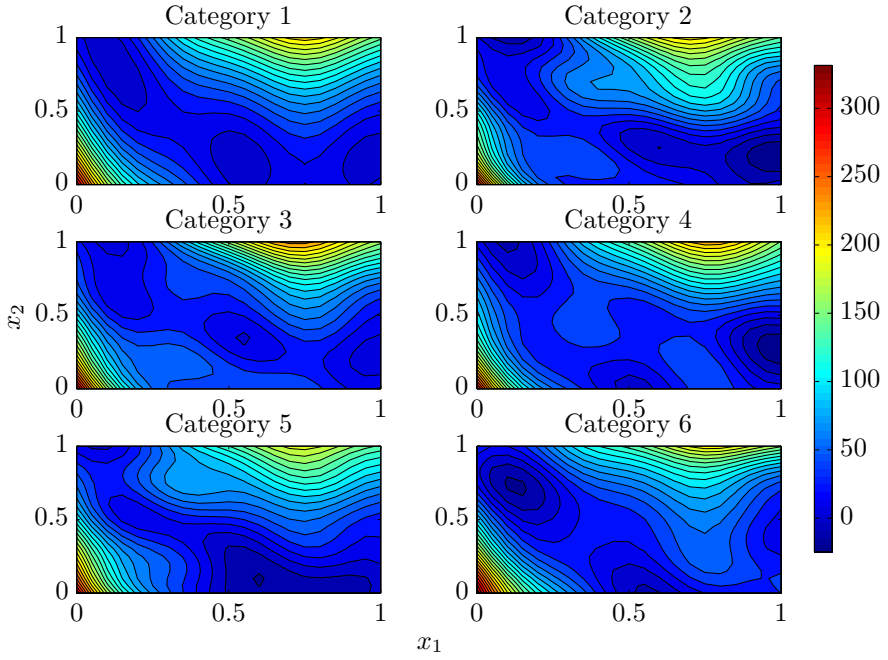


Figure 75: Example of Contours of Disturbed Branin Function. 6 Categories

Table 14: Categorical Minimum Values and their Location for the Disturbed Branin Function

Problem	x_1	x_2	Min Value
Dist Branin 1 (original)	0.1036	0.9	1.5892
Dist Branin 2	0.9921	0.2	-24.8723
Dist Branin 3	0.9893	0.2	4.0047
Dist Branin 4	0.9847	0.3	-21.0302
Dist Branin 5	0.5952	0.1	-13.4278
Dist Branin 6	0.1338	0.7	-15.1981

APPENDIX B

CANONICAL CONCEPTS WITH SIMILAR TRENDS

A canonical set of functions is needed to test ECMF surrogates. It is obtained out of the two-dimensional function proposed by Michalewicz [133]. Concept 2 is the original two-dimensional Michalewicz function whereas concept 1 is a modified spline of a coordinate plane slice of the original two dimensional Michalewicz's function. A quadratic function has been added to the coordinate plane slice of the original two-dimensional Michalewicz's function to build the modified spline. Figure 76 shows the two concepts of the Michalewicz canonical set.

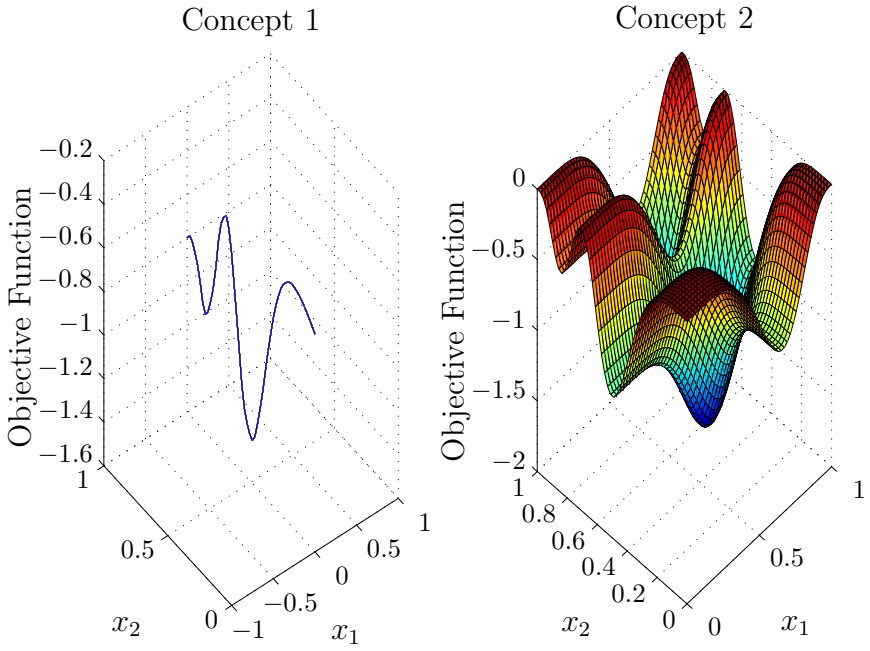


Figure 76: The Two Concepts of the Michalewicz Function

APPENDIX C

MIXED-INTEGER CANONICAL PROBLEMS

1) Test Problem 1. Taken from Floudas, Pardalos et al. [62]. It is the example 12.2.2.

$$\min_{\mathbf{x},y} -0.7y + 5\left(x_1 - 0.5\right)^2 + 0.8 \tag{130}$$

$$\begin{aligned} -\exp\left(x_1 - 0.2\right) - x_2 &\leq 0 && (131)\\ x_2 + 1.1y &\leq -1 && (132)\\ x_1 - 1.2y &\leq 0.2 && (133)\\ 0.2 \leq x_1 &\leq 1 && (134)\\ -2.22554 \leq x_2 &\leq -1 && (135)\\ y &\in \{0,1\} && (136) \end{aligned}$$

The known global optimum is $(x_1, x_2, y, f) = (0.9419, -2.1, 1, 1.0765)$.

2) Test Problem 2. Taken from Floudas, Pardalos et al. [62]. It is the example 12.2.6.

$$\min_{x,y} 3y - 5x \tag{137}$$

$$2y^2 - 2y^{0.5} - 2x^{0.5}y^2 + 11y + 8x \leq 39 \quad (138)$$

$$-y + x \leq 3 \quad (139)$$

$$2y + 3x \leq 24 \quad (140)$$

$$1 \leq x \leq 10 \quad (141)$$

$$y \in [1, 6] \cap \mathbb{N} \quad (142)$$

The known global optimum is $(x, y, f) = (4, 1, -17)$.

3) Test Problem 3. Taken from Appendix A of Ref. [87]. It is the example 21.

$$\min_{\mathbf{x},y} \sum_{i=1}^{9} \left[\exp\left(-\frac{(u_i - y_2)^x}{y_1} - 0.01i \right) \right] \quad (143)$$

where $u_i = 25 + (-50 \log(0.01i))^{\frac{2}{3}}$

$$0 \leq x \leq 5 \quad (144)$$

$$y_1 \in [1, 100] \cap \mathbb{N} \quad (145)$$

$$y_2 \in [0, 25] \cap \mathbb{Z} \quad (146)$$

The known global optimum is $(x, y_1, y_2, f) = (1.5, 50, 25, 0)$.

4) Test Problem 4. Constrained Branin function with the second variable discrete.

$$\min_{x,y} \left(y - \frac{5.1}{4\pi^2}y + \frac{5}{\pi}x - 6 \right)^2 + 10 \left[\left(1 - \frac{1}{8\pi} \right) \cos x + 1 \right] + 5x \quad (147)$$

$$-xy < -0.2 \quad (148)$$

$$-5 \leq x \leq 10 \quad (149)$$

$$y \in \left\{ 0, \frac{3}{2}, 3, \frac{9}{2}, 6, ...15 \right\} \quad (150)$$

The known global optimum is $\left(9.7315, \frac{9}{2}, 8.833\right)$.

REFERENCES

[1] "FLIGHTLAB model editor manual," manual, Advanced Rotorcraft Technologies, May 2006.

[2] Aha, D. W., Kibler, D., and Albert, M. K., "Instance-Based learning algorithms," *Machine Learning*, vol. 6, pp. 37–66, Jan. 1991.

[3] Aha, D. W., Kibler, D., and Albert, M. K., "Instance-Based learning algorithms," *Machine Learning*, vol. 6, p. 3766, 1991.

[4] Alexandrov, N. M., Lewis, R. M., Gumbert, C. R., Green, L. L., and Newman, P. A., "Approximation and model management in aerodynamic optimization with Variable-Fidelity models," *Journal of Aircraft*, vol. 38, pp. 1093–1101, Dec. 2001.

[5] Alexandrov, N. M., Nielsen, E. J., Lewis, R. M., and Anderson, W. K., "First-Order model management with Variable-Fidelity physics applied to Multi-Element airfoil optimization," in *8th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis & Optimization*, (Long Beach, CA), Sept. 2000.

[6] Ash, R. B., *Information Theory*. New York, NY: Interscience Publishers, 1965.

[7] Avigad, G. and Moshaiov, A., "Simultaneous Concept-Based evolutionary Multi-Objective optimization," *Applied Soft Computing Journal*, vol. 11, pp. 193–207, Jan. 2011.

[8] Bäck, T., *Evolutionary Algorithms in Theory and Practice: Evolution Strategies, Evolutionary Programming, Genetic Algorithms*. New York, USA: Oxford University Press, Jan. 1996.

[9] Bader, J. and Zitzler, E., "Hype: An algorithm for fast Hypervolume-Based Many-Objective optimization," *Evolutionary Computation*, vol. 19, no. 1, pp. 45–76, 2011.

[10] Balakrishnan, A. V., *Applied Functional Analysis*. New York, NY: Springer-Verlag, 2nd ed., 1981.

[11] Balas, E., Glover, F., and Zionts, S., "An additive algorithm for solving linear programs with Zero-One variables," *Operations Research*, vol. 13, pp. 517–549, July 1965.

[12] Bandler, J. W., Cheng, Q., Dakroury, S., Mohamed, A., Bakr, M., Madsen, K., and Sndergaard, J., "Space mapping: The state of the art," *Electromagnetics-Based Optimization of Microwave Components and Circuits*, vol. 52, pp. 337–361, Jan. 2004.

[13] Batchelor, B. G., *Pattern Recognition : Ideas in Practice*. New York, NY: Plenum Press, 1978.

[14] Berman, O. and Ashrafi, N., "Optimization models for reliability of modular software systems," *IEEE Transactions on Software Engineering*, vol. 19, pp. 1119–1123, Nov. 1993.

[15] Berry, B., Bowen-Davies, G., Gluesenkamp, K., Kaler, Z., Schmaus, J., Staruk, W., Weiner, E., and Woods, B. K., "Design optimization of gamera II: a human powered helicopter," in *Proceedings of the 68th American Helicopter Society International Annual Forum*, vol. 1, (Fort Worth, TX), pp. 646–664, May 2012.

[16] Beume, N., Naujoks, B., and Emmerich, M., "SMS-EMOA: multiobjective selection based on dominated hypervolume," *European Journal of Operational Research*, vol. 181, pp. 1653–1669, Sept. 2007.

[17] Bishop, C. M., *Neural Networks for Pattern Recognition*. New York, NY: Oxford University Press, 1st ed., 1996.

[18] Bishop, C. M., *Pattern Recognition and Machine Learning*. New York, NY: Springer, 1st ed., 2007.

[19] Booker, A. J., Dennis, J. E., Frank, P. D., Serafini, D. B., and Torczon, V., "Optimization using surrogate objectives on a helicopter test example," in *Computational Methods in Optimal Design and Control* (Borggaard, J., Burns, J., Cliff, E., and Schreek, S., eds.), pp. 49–58, Boston, MA: Birkhauser, 1998.

[20] Borchers, B. and Mitchell, J. E., "An improved branch and bound algorithm for mixed integer nonlinear programs," *Computers & Operations Research*, vol. 21, pp. 359–367, Apr. 1994.

[21] Boukouvala, F., Dubey, A., Vanarase, A., Ramachandran, R., Muzzio, F. J., and Ierapetritou, M., "Computational approaches for studying the granular dynamics of continuous blending processes, 2 - population balance and Data-Based methods," *Macromolecular Materials and Engineering*, vol. 297, pp. 9–19, Jan. 2012.

[22] Boukouvala, F., Muzzio, F. J., and Ierapetritou, M. G., "Design space of pharmaceutical processes using Data-Driven-Based methods," *Journal of Pharmaceutical Innovation*, vol. 5, pp. 119–137, Oct. 2010.

[23] Bousman, W. G. and Kufeld, R. M., "UH-60A airloads catalog," Technical Memorandum 212827, National Aeronautics and Space Administration, Ames Research Center, Moffett Field, CA, 2005.

[24] Box, G. E. P., *Empirical Model-Building and Response Surfaces*. New York, NY: John Wiley & Sons, 1987.

[25] Brentner, K. S., Bres, G. A., Perez, G., and Jones, H. E., "Maneuvering rotorcraft noise prediction," *Journal of Sound and Vibration*, vol. 275, pp. 719–738, Aug. 2004.

[26] Cao, Y. J. and Wu, Q. H., "A mixed variable evolutionary programming for optimisation of mechanical design," *International Journal of Engineering Intelligent Systems for Electrical Engineering and Communications*, vol. 7, no. 2, pp. 77–82, 1999.

[27] Capitanescu, F. and Wehenkel, L., "A new heuristic approach to deal with discrete variables in optimal power flow computations," in *2009 IEEE Bucharest PowerTech: Innovative Ideas Toward the Electrical Grid of the Future*, (Bucharest, Romania), June 2009.

[28] Choi, S., Alonso, J. J., Kroo, I. M., and Wintzer, M., "Multifidelity design optimization of Low-Boom supersonic jets," *Journal of Aircraft*, vol. 45, no. 1, pp. 106–118, 2008.

[29] Choi, S. C., Park, J. S., and Kim, J. H., "Vibration control of Pre-Twisted rotating composite Thin-Walled beams with piezoelectric fiber composites," *Journal of Sound and Vibration*, vol. 300, pp. 176–196, Feb. 2007.

[30] Chu, W. and Ghahramani, Z., "Gaussian processes for ordinal regression," *Journal of Machine Learning Research*, vol. 6, pp. 1019–1041, July 2005.

[31] Coello, C. A., van Veldhuizen, D. A., and Lamont, G. B., *Evolutionary Algorithms for Solving Multi-Objective Problems*. New York, NY: Springer, 2002.

[32] Cover, T. and Hart, P., "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, pp. 21–27, Jan. 1967.

[33] Cox, D. D. and John, S., "A statistical method for global optimization," in *IEEE International Conference on Systems, Man and Cybernetics*, vol. 2, (Chicago, IL), pp. 1241 –1246, Oct. 1992.

[34] Crossley, W. and Laananen, D. H., "The genetic algorithm as an automated methodology for helicopter conceptual design," *Journal of Engineering Design*, vol. 8, pp. 231–250, Sept. 1997.

[35] Crossley, W. A., Welss, V. L., and Laananen, D. H., "The potential of genetic algorithms for conceptual design of rotor systems," *Engineering Optimization*, vol. 24, no. 3, pp. 221–238, 1995.

[36] Dakin, R. J., "A Tree-Search algorithm for mixed integer programming problems," *The Computer Journal*, vol. 8, pp. 250 –255, Jan. 1965.

[37] Davis, E. and Ierapetritou, M., "A kriging based method for the solution of mixed-integer nonlinear programs containing Black-Box functions," *Journal of Global Optimization*, vol. 43, pp. 191–205, Mar. 2009.

[38] Davis, J., *Design Methodology For Developing Concept Independent Rotorcraft Analysis And Design Software*. PhD thesis, Georgia Institute of Technology, Atlanta, GA, Dec. 2007.

[39] de Jong, K. A., *Analysis of the Behavior of a Class of Genetic Adaptive Systems*. PhD thesis, University of Michigan, Ann Arbor, MI, 1975.

[40] Deb, K., *Multi-Objective Optimization Using Evolutionary Algorithms*. Chichester, NY: John Wiley and Sons, 2001.

[41] Deb, K., Agrawal, S., Pratap, A., and Meyarivan, T., "A fast elitist non-dominated sorting genetic algorithm for Multi-Objective optimization: NSGA-II," in *Proceedings of 6th International Conference on Parallel Problem Solving from Nature*, (Berlin, Germany), pp. 849–58, Sept. 2000.

[42] Deb, K., Mohan, M., and Mishra, S., "Evaluating the ϵ-Domination based Multi-Objective evolutionary algorithm for a quick computation of Pareto-Optimal solutions," *Evolutionary Computation*, vol. 13, no. 4, pp. 501–525, 2005.

[43] Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T., "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 6, pp. 182–97, Apr. 2002.

[44] Deep, K., Singh, K. P., Kansal, M. L., and Mohan, C., "A real coded genetic algorithm for solving integer and mixed integer optimization problems," *Applied Mathematics and Computation*, vol. 212, pp. 505–518, June 2009.

[45] Deep, K. and Thakur, M., "A new crossover operator for real coded genetic algorithms," *Applied Mathematics and Computation*, vol. 188, pp. 895–911, May 2007.

[46] Duran, M. A. and Grossmann, I. E., "A mixed integer nonlinear programming algorithm for process systems synthesis," *AIChE Journal*, vol. 32, no. 4, pp. 592–606, 1986.

[47] Duran, M. A. and Grossmann, I. E., "An Outer-Approximation algorithm for a class of Mixed-Integer nonlinear programs," *Mathematical Programming*, vol. 36, no. 3, pp. 307–339, 1986.

[48] Dym, C. L., Wood, W. H., and Scott, M. J., "Rank ordering engineering designs: Pairwise comparison charts and borda counts," *Research in Engineering Design*, vol. 13, pp. 236–242, Sept. 2002.

[49] Dym, H. and McKean, H. P., *Fourier Series and Integrals*. New York, NY: Academic Press, 1972.

[50] Edwards, B., Andrews, J., and Rahnke, C., "Ducted tail rotor designs for rotorcraft and their low noise features," in *Proceedings of the FVP Symposium on 'Advances in Rotorcraft Technology'*, (Ottawa, Canada), 1996.

[51] El-Beltagy, M. A., "An evolutionary risk adjusting model fusion framework for optimizing models with variable fidelity," in *Collection of Technical Papers - 10th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, vol. 5, (Albany, NY), pp. 3042–3050, Aug. 2004.

[52] El-Samanoudy, M., Ghorab, A. A. E., and Youssef, S. Z., "Effect of some design parameters on the performance of a giromill vertical axis wind turbine," *Ain Shams Engineering Journal*, vol. 1, pp. 85–95, Sept. 2010.

[53] Eldred, M. S., Giunta, A. A., and Collis, S. S., "Second-Order corrections for Surrogate-Based optimization with model hierarchies," in *Collection of Technical Papers - 10th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, vol. 3, (Albany, NY), pp. 1754–1768, Aug. 2004.

[54] Emmerich, M., Deutz, A. H., and Klinkenberg, J. W., "Hypervolume-Based expected improvement: Monotonicity properties and exact computation," in *Proceedings of the 2011 IEEE Congress of Evolutionary Computation*, (New Orleans, LA), pp. 2147–2154, June 2011.

[55] Emmerich, M., Giannakoglou, K. C., and Naujoks, B., "Single- and multiobjective evolutionary optimization assisted by gaussian random field metamodels," *IEEE Transactions on Evolutionary Computation*, vol. 10, pp. 421–439, Aug. 2006.

[56] Emmerich, M., Grötzner, M., Groβ, B., and Schütz, M., "Mixed-Integer evolution strategy for chemical plant optimization with simulators," in *Evolutionary Design and Manufacture: Selected papers ACDM 2000*, London, UK: Springer, 2000.

[57] Emmerich, M. T. M., Li, R., Zhang, A., Flesch, I., Lucas, P. J. F., and Nijholt, A., "Mixed-integer bayesian optimization utilizing a-priori knowledge on parameter dependences," in *The 20th Belgian-Netherlands Conference on Artificial Intelligence*, (Enschede, Netherlands), Oct. 2008.

[58] Fleischer, M. and Fleischer, M., "The measure of pareto optima. applications to multi-objective metaheuristics," in *Proceedings of the Evolutionary Multi-Criterion Optimization: Second International Conference*, (Faro, Portugal), Apr. 2003.

[59] Fletcher, R. and Leyffer, S., "Solving mixed integer nonlinear programs by outer approximation," *Mathematical Programming, Series B*, vol. 66, pp. 327–349, Sept. 1994.

[60] Floudas, C. A., *Deterministic Global Optimization: Theory, Methods, and Applications*. Dordrecht, The Netherlands: Kluwer Academic Publishers, 2000.

[61] Floudas, C. A., Ciric, A. R., and Grossmann, I. E., "Automatic synthesis of optimum heat exchanger network configurations," *AIChE Journal*, vol. 32, pp. 276–290, Feb. 1986.

[62] Floudas, C. A., Pardalos, P. M., Adjiman, C., Esposito, W. R., Gümüs, Z. H., Harding, S. T., Klepeis, J. L., Meyer, C. A., and Schweiger, C. A., *Handbook of Test Problems in Local and Global Optimization: Nonconvex Optimization and Its Applications*. Dordrecht, The Netherlands: Kluwer Academic Publishers, 1st ed., June 1999.

[63] Forrester, A., Bressloff, N. W., and Keane, A., "Optimization using surrogate models and partially converged computational fluid dynamics simulations," *Proceedings of the Royal Society of London, Series A*, vol. 462, pp. 2177–2204, July 2006.

[64] Forrester, A., Sobester, A., and Keane, A., "Multi-fidelity optimization via surrogate modelling," *Proceedings of the Royal Society of London, Series A*, vol. 463, pp. 3251–69, Dec. 2007.

[65] Forrester, A., Sobester, A., and Keane, A., *Engineering Design Via Surrogate Modelling: A Practical Guide*. Chichester, UK: John Wiley & Sons, Sept. 2008.

[66] Forrester, A. I. J., Keane, A. J., and Bressloff, N. W., "Design and analysis of "Noisy" computer experiments," *AIAA Journal*, vol. 44, pp. 2331–2339, Oct. 2006.

[67] Friedrich, T., Horoba, C., and Neumann, F., "Multiplicative approximations and the hypervolume indicator," in *11th Annual Genetic and Evolutionary Computation Conference*, (Montreal, Canada), pp. 571–578, July 2009.

[68] Ganguli, R., "Survey of recent developments in rotorcraft design optimization," *Journal of Aircraft*, vol. 41, pp. 493–510, June 2004.

[69] Gao, Y., Ren, Z., and Gao, Y., "Modified differential evolution algorithm of constrained nonlinear mixed integer programming problems," *Information Technology Journal*, vol. 10, no. 11, pp. 2068–2075, 2011.

[70] Gelfand, I. M., Fomin, S. V., and Silverman, R. A., *Calculus of Variations*. Mineola, NY: Dover Publications, 2000.

[71] Geoffrion, A. M., "Generalized benders decomposition," *Journal of Optimization Theory and Applications*, vol. 10, no. 4, pp. 237–260, 1972.

[72] Giesing, J. and Barthelemy, J. F., "A summary of industry MDO applications and needs," in *AIAA/USAF/NASA/ISSMO Symposium of Multidisciplinary Analysis and Optimization*, (St. Louis, MO), Sept. 1998.

[73] Glaz, B., Friedmann, P. P., and Liu, L., "Helicopter vibration reduction throughout the entire flight envelope using surrogate-based optimization," *Journal of the American Helicopter Society*, vol. 54, pp. 1–15, Jan. 2009.

[74] Glaz, B., Liu, L., and Friedmann, P. P., "Reduced-Order nonlinear unsteady aerodynamic modeling using a Surrogate-Based recurrence framework," *AIAA Journal*, vol. 48, pp. 2418–2429, Oct. 2010.

[75] Glaz, B., Liu, L., Friedmann, P. P., Bain, J., and Sankar, L. N., "A Surrogate-Based approach to Reduced-Order dynamic stall modeling," *Journal of the American Helicopter Society*, vol. 57, Apr. 2012.

[76] Gupta, O. K. and Ravindran, A., "Branch and bound experiments in convex nonlinear integer programming," *Management Science*, vol. 31, pp. 1533–1546, Dec. 1985.

[77] Haftka, R. T., "Combining global and local approximations," *AIAA Journal*, vol. 29, pp. 1523–1525, Sept. 1991.

[78] Hamming, R. W., "Error detecting and error correcting codes," *Bell System Technical Journal*, vol. 29, pp. 147–160, Apr. 1950.

[79] Han, Z. H., Zimmermann, R., and Gortz, S., "A new cokriging method for Variable-Fidelity surrogate modeling of aerodynamic data," in *48th AIAA Aerospace Sciences Meeting Including the New Horizons Forum and Aerospace Exposition*, (Orlando, FL), Jan. 2010. Compendex.

[80] Hao, W., Ying, Y., Wei, Y., and Baohua, L., "Adaptive Approximation-Based optimization of composite advanced grid-stiffened cylinder," *Chinese Journal of Aeronautics*, vol. 23, pp. 423–429, Aug. 2010.

[81] Harjunkoski, I., *Application Of MINLP Methods On A Scheduling Problem In The Paper Converting Industry*. Ph.D. thesis, Abŏ Akademi University, Abŏ, Finland, 1997.

[82] He, S., Prempain, E., and Wu, Q. H., "An improved particle swarm optimizer for mechanical design optimization problems," *Engineering Optimization*, vol. 36, pp. 585–605, Oct. 2004.

[83] Hein, B. R. and Chopra, I., "Hover performance of a micro air vehicle: Rotors at low reynolds number," *Journal of the American Helicopter Society*, vol. 52, pp. 254–262, July 2007.

[84] Hepperle, H., "JAVAFOIL user's guide," Tech. Rep. Manual, Dec. 2011.

[85] Hess, R. A., Zeyada, Y., and Heffley, R., "Modeling and simulation for helicopter task analysis," *Journal of the American Helicopter Society*, vol. 47, pp. 243–252, Oct. 2002.

[86] Hevesi, J. A., Flint, A. L., and Istok, J. D., "Precipitation estimation in mountainous terrain using multivariate geostatistics, part II: isohyetal maps," *Journal of Applied Meteorology*, vol. 31, pp. 677–688, July 1992.

[87] Himmelblau, D. M., *Applied Nonlinear Programming*. New York, NY: McGraw-Hill, 1972.

[88] Holland, J. H., *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. Ann Arbor, MI: University of Michigan Press, 1975.

[89] Huband, S., Hingston, P., While, L., and Barone, L., "An evolution strategy with probabilistic mutation for multi-objective optimisation," in *2003 Congress on Evolutionary Computation*, vol. 4, (Piscataway, NJ), pp. 2284–91, Dec. 2003.

[90] Igel, C., Hansen, N., and Roth, S., "Covariance matrix adaptation for multi-objective optimization," *Evolutionary Computation*, vol. 15, no. 1, pp. 1–28, 2007.

[91] Imiela, M., "High-Fidelity optimization framework for helicopter rotors," *Aerospace Science and Technology*, vol. 23, pp. 2–16, Dec. 2012.

[92] Jansson, N., Wakeman, W. D., and Månson, J. A., "Optimization of hybrid thermoplastic composite structures using surrogate models and genetic algorithms," *Composite Structures*, vol. 80, pp. 21–31, Sept. 2007.

[93] Jaynes, E. T., "Prior probabilities," *IEEE Transactions on Systems Science and Cybernetics*, vol. SSC-4, no. 3, pp. 227–241, 1968.

[94] Jin, R., Chen, W., and Simpson, T. W., "Comparative studies of metamodeling techniques under multiple modeling criteria," *Structural and Multidisciplinary Optimization*, vol. 23, pp. 1–13, Dec. 2001.

[95] Johnson, C. and Barakos, G., "Development of a framework for optimising aspects of rotor blades," in *Proceedings of the 66th American Helicopter Society International Annual Forum*, (Phoenix, AZ), pp. 2119–2134, May 2010.

[96] Johnson, M. E., Moore, L. M., and Ylvisaker, D., "Minimax and maximin distance designs," *Journal of Statistical Planning and Inference*, vol. 26, pp. 131–148, 1990.

[97] Johnson, W., "NDARC NASA design and analysis of rotorcraft," Tech. Rep. TP-2009-215402, NASA Ames Research Center, Moffett Field, CA, 2009.

[98] Jones, D. R., "A taxonomy of global optimization methods based on response surfaces," *Journal of Global Optimization*, vol. 21, pp. 345–383, Dec. 2001.

[99] Jones, D. R., Schonlau, M., and Welch, W. J., "Efficient global optimization of expensive Black-Box functions," *Journal of Global Optimization*, vol. 13, pp. 455–492, Dec. 1998.

[100] Jones, D. R. and Welch, W. J., "Global optimization using response surfaces," in *Fifth SIAM conference on optimization*, (Victoria, Canada), May 1996.

[101] Journel, A. G. and Huijbregts, C. J., *Mining Geostatistics*. London, UK: Academic Press, 1978.

[102] Kalra, T. S., Amiraux, M., Nagarai, V. T., Chopra, I., and Baeder, J. D., "A comparative study of different weight formulations affecting preliminary sizing of rotorcraft," in *68th American Helicopter Society International Annual Forum*, vol. 1, (Fort Worth, TX), pp. 617–631, May 2012.

[103] Karoui, K., Platbrood, L., Crisciu, H., and Waltz, R. A., "New Large-Scale security constrained optimal power flow program using a new interior point algorithm," in *2008 5th International Conference on the European Electricity Market*, (Lisboa, Portugal), May 2008.

[104] Keane, A. and Nair, P., *Computational Approaches for Aerospace Design: The Pursuit of Excellence*. Chichester, UK: John Wiley & Sons, Aug. 2005.

[105] Keane, A. J., "Wing optimization using design of experiment, response surface, and data fusion methods," *Journal of Aircraft*, vol. 40, pp. 741–750, Aug. 2003.

[106] Keane, A. J., "Statistical improvement criteria for use in multiobjective design optimization," *AIAA Journal*, vol. 44, pp. 879–891, Apr. 2006.

[107] Keane, A. J. and Petruzzelli, N., "Aircraft wing design using Ga-Based multi-level strategies," in *Eighth AIAA/USAF/NASA/ISSMO Symp. on Multidisciplinary Analysis and Optimization*, (Long Beach, Ca), Sept. 2000.

[108] Kelley, J. L., *General Topology*. New York, NY: Springer-Verlag, 1975.

[109] Kennedy, M. C. and O'Hagan, A., "Predicting the output from a complex computer code when fast approximations are available," *Biometrika*, vol. 87, no. 1, pp. 1–13, 2000.

[110] Kenney, J. F., *Mathematics of Statistics*. Princeton, NJ: Van Nostrand Company, 1962.

[111] Kesavan, P. and Barton, P. I., "Generalized Branch-and-Cut framework for Mixed-Integer nonlinear optimization problems," *Computers & Chemical Engineering*, vol. 24, no. 2-7, pp. 1361–1366, 2000.

[112] Keung, J., "Software development cost estimation using analogy: A review," in *Proceedings of the Australian Software Engineering Conference ASWEC*, (Gold Coast, Australia), pp. 327–336, Apr. 2009.

[113] Kleijnen, J. P., van Beers, W., and van Nieuwenhuyse, I., "Constrained optimization in expensive simulation: Novel approach," *European Journal of Operational Research*, vol. 202, pp. 164–174, Apr. 2010.

[114] Knill, D. L., Giunta, A. A., Baker, C. A., Grossman, B., Mason, W. H., Haftka, R. T., and Watson, L. T., "Response surface models combining linear and euler aerodynamics for supersonic transport design," *Journal of Aircraft*, vol. 36, no. 1, pp. 75–86, 1999.

[115] Knowles, J. and Corne, D., "On metrics for comparing nondominated sets," in *Proceedings of the 2002 Congress on Evolutionary Computation*, vol. 1, (Hawaii, HI), pp. 711–716, May 2002.

[116] Koch, P. N., Simpson, T. W., Allen, J. K., and Mistree, F., "Statistical approximations for multidisciplinary design optimization: The problem of size," *Journal of Aircraft*, vol. 36, no. 1, pp. 275–286, 1999.

[117] Krige, D. G., "Statistical approach to some basic mine valuation problems on witwatersrand," *Journal of Institute of Mine Surveyors of South Africa*, vol. 7, no. 4, pp. 145–154, 1953.

[118] Kumar, D., Cesnik, C. E., Rohl, P. J., and Sutton, M., "Optimization framework for the dynamic analysis and design of active twist rotors," in *Proceedings of the 68th American Helicopter Society International Annual Forum*, (Fort Worth, TX), pp. 1280–1292, May 2012.

[119] Kumar, D., Glaz, B., Mok, J., Friedmann, P., and Cesnik, C., "Determination of optimum camber distribution in rotating wings with deformable airfoils for vibration reduction and performance enhancement using surrogate modeling," in *Proceedings of 10th European Rotorcraft Forum*, (The Hague, Holland), Aug. 1984.

[120] Kushner, H. J., "A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise," *Journal of Basic Engineering*, vol. 86, pp. 97–106, Mar. 1964.

[121] Lawrence, T. H., Corning, S., and Wharburton, F., "Helicopter maneuverability and agility design sensitivity and air combat maneuver data correlation study," Technical Report 160550, US Army Aviation System Command, 1991.

[122] Leary, S. J., Bhaskar, A., and Keane, A. J., "A Knowledge-Based approach to response surface modelling in multifidelity optimization," *Journal of Global Optimization*, vol. 26, pp. 297–319, July 2003.

[123] Leishman, J. G., *Principles of Helicopter Aerodynamics*. New York, NY: Cambridge University Press, 2006.

[124] Li, C. and Li, H., "A survey of distance metrics for nominal attributes," *Journal of Software*, vol. 5, pp. 1262–1269, Nov. 2010.

[125] Li, R., Eggermont, J., Emmerich, M., Bovenkamp, E., Bäck, T., Dijkstra, J., and Reiber, J., "Mixed-Integer optimization of coronary vessel image analysis using evolution strategies," in *8th Annual Genetic and Evolutionary Computation Conference*, vol. 2, (Seattle, WA), July 2006.

[126] Li, R., Eggermont, J., Shir, O. M., Emmerich, M., Bäck, T., Dijkstra, J., and Reiber, J., "Mixed-Integer evolution strategies with dynamic niching," in *Parallel Problem Solving from Nature*, vol. 5199, pp. 246–255, Berlin, Germany: Springer Berlin Heidelberg, 2008.

[127] Li, R., Emmerich, M., Eggermont, J., Bovenkamp, E., Bäck, T., Dijkstra, J., and Reiber, J., "Optimizing a medical image analysis system using Mixed-Integer evolution strategies," in *Evolutionary Image Analysis and Signal Processing*, vol. 213, pp. 91–112, Berlin, Germany: Springer Berlin Heidelberg, 2009.

[128] Liu, W., Zhang, Q., Tsang, E., Liu, C., and Virginas, B., "On the performance of metamodel assisted MOEA/D," in *2nd International Symposium on Intelligence Computation and Applications*, (Wuhan, China), Sept. 2007.

[129] Lyle, K. H., Jackson, K. E., and Fasanella, E. L., "Development of an ACAP helicopter finite element impact model," *Journal of the American Helicopter Society*, vol. 45, pp. 137–142, Apr. 2000.

[130] MacKay, D. J. C., *Information Theory, Inference and Learning Algorithms*. Cambridge, UK: Cambridge University Press, 1st ed., 2003.

[131] Mair, C. and Shepperd, M., "The consistency of empirical comparisons of regression and analogy-based software project cost prediction," in *International Symposium on Empirical Software Engineering*, pp. 509–518, Nov. 2005.

[132] Mason, W. H., Knill, D. L., Giunta, A. A., Grossman, B., Watson, L. T., and Haftka, R. T., "Getting the full benefits of CFD in conceptual

design," in *Sixteenth AIAA Applied Aerodynamics Conference*, vol. 98, (Alburquerque, NM), p. 2513, June 1998.

[133] Michalewicz, Z., *Genetic Algorithms + Data Structures = Evolution Programs*. Berlin, Germany: Springer-Verlag, 2nd ed., 1994.

[134] Michalski, R. S., Stepp, R. E., and Diday, E., "A recent advance in data analysis: Clustering objects into classes characterized by conjunctive concepts," in *Progress in Pattern Recognition*, vol. 1, pp. 33–55, Amsterdam, The Netherlands: North-Holland Publishing, 1981.

[135] Mitchell, T. M., *Machine Learning*. New York, NY: McGraw-Hill Science/Engineering/Math, 1 ed., 1997.

[136] Mockus, J., Tiesis, V., and Zilinskas, A., "The application of bayesian methods for seeking the extremum," in *Towards global optimisation. II*, pp. 117–29, Amsterdam, Netherlands: North-Holland Publishing, 1978.

[137] Moore, D. S., *The Basic Practice of Statistics*. New York, NY: W.H. Freeman, 2nd ed. ed., 2000.

[138] Mouille, R., "The fenestron, shrouded tail rotor of the SA. 341 gazelle," *Journal of the American Helicopter Society*, vol. 15, pp. 31–37, Oct. 1970.

[139] Mouille, R. and d'Ambra, F., "The 'Fenestron' a shrouded tail rotor concept for helicopters," in *Proceedings of the 42nd American Helicopter Society International Annual Forum*, (Alexandria, VA), June 1986.

[140] Murugan, M., Ganguli, R., and Harursampath, D., "Surrogate based design optimisation of composite aerofoil cross-section for helicopter vibration reduction," *Aeronautical Journal*, vol. 116, pp. 709–725, July 2012.

[141] Murugan, S. and Ganguli, R., "Aeroelastic stability enhancement and vibration suppression in a composite helicopter rotor," *Journal of Aircraft*, vol. 42, pp. 1013–1024, July 2005.

[142] Nadler, M. and Smith, E. P., *Pattern Recognition Engineering*. New York, NY: Wiley-Interscience, 1st ed., 1993.

[143] of Defense, D., *Systems Engineering Fundamentals*. Fort Belvoir, VA: Defense Acquisition University, 2001.

[144] Pareto, V., *Cours d'Èconomie Politique Professé à l'Université de Lausanne*. 1896.

[145] Ponweiser, W., Wagner, T., Biermann, D., and Vincze, M., "Multiobjective optimization on a limited budget of evaluations using Model-Assisted S-Metric selection," in *10th International Conference on Parallel Problem Solving from Nature*, vol. 5199 LNCS, (Dortmund, Germany), pp. 784–794, Sept. 2008.

[146] Prouty, R. W., *Helicopter Performance, Stability, and Control*. Malabar, FL: Krieger Publishing Company, 1995.

[147] Pugh, S., *Total Design: Integrated Methods for Successful Product Engineering*. Wokingham, England: Addison-Wesley Pub., 1991.

[148] Quesada, I. and Grossman, I. E., "An LP/NLP based branch and bound algorithm for convex MINLP optimization problems," *Computers & Chemical Engineering*, vol. 16, pp. 937–47, Oct. 1992.

[149] Radcliffe, N. J., "Forma analysis and random respectful recombination," in *Proceedings of the 4th International Conference on Genetic Algorithms*, (San Mateo, CA), pp. 222–229, Morgan-Kaufmann Publ, Inc., July 1991.

[150] Rand, O. and Khromov, V., "Helicopter sizing by statistics," *Journal of the American Helicopter Society*, vol. 49, pp. 300–317, July 2004.

[151] Raymer, D. P., *Aircraft Design: A Conceptual Approach*. Washington D.C., USA: American Institute of Aeronautics and Astronautics, 4th ed., 2006.

[152] Regulwar, D. G., "Differential evolution algorithm with application to optimal operation of multipurpose reservoir," *Journal of Water Resource and Protection*, vol. 2, pp. 560–568, June 2010.

[153] Rikards, R., Abramovich, H., Kalnins, K., and Auzins, J., "Surrogate modeling in design optimization of stiffened composite shells," *Composite Structures*, vol. 73, pp. 244–251, May 2006.

[154] Robinson, F., "Increasing tail rotor thrust and comments on other yaw control devices," *Journal of the American Helicopter Society*, vol. 15, pp. 46–52, Oct. 1970.

[155] Robinson, T. D., Eldred, M. S., Willcox, K. E., and Haimes, R., "Surrogate-Based optimization using multifidelity models with variable parameterization and corrected space mapping," *AIAA Journal*, vol. 46, pp. 2814–2822, Nov. 2008.

[156] Rohl, P. J., Kumar, D., Dorman, P., Cesnik, C. E., and Sutton, M., "A composite rotor blade structural design environment for aeromechanical assessments in conceptual and preliminary design," in *Proceedings of the 68th American Helicopter Society International Annual Forum*, (Fort Worth, TX), pp. 632–645, May 2012.

[157] Roskam, J., *Airplane Design Part I: Preliminary Sizing of Airplanes*. Lawerence, KS: DARcorporation, 3rd ed., 2003.

[158] Rousis, D., *A Pareto Frontier Intersection-based Approach for Efficient Multiobjective Optimization of Competing Concept Alternatives*. PhD thesis, Georgia Institute of Technology, Atlanta, GA, Aug. 2011.

[159] Rudolph, G., "An evolutionary algorithm for integer programming," in *Parallel Problem Solving from Nature-PPSN III*, vol. 866 of *Lecture Notes in Computer Science*, pp. 139–148, Berlin, Germany: Springer-Verlag, 1994.

[160] Sacks, J., "Design and analysis of computer experiments," *Statistical Science*, vol. 4, pp. 409–423, Nov. 1989.

[161] Sagan, H., *Introduction to the Calculus of Variations*. Mineola, NY: Courier Dover Publications, 1992.

[162] Saijal, K., Ganguli, R., and Viswamurthy, S., "Optimization of helicopter rotor using polynomial and neural network metamodels," *Journal of Aircraft*, vol. 48, pp. 553–66, Mar. 2011.

[163] Salonen, M. and Perttula, M., "Utilization of concept selection methods: A survey of finnish industry," in *ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, (Long Beach, CA), pp. 527–535, Sept. 2005.

[164] Salzberg, S., "A nearest hyperrectangle learning method," *Machine Learning*, vol. 6, pp. 251–276, May 1991.

[165] Schleicher, D. R., "Advanced civil tiltrotor design optimization and issues," in *Proceedings of the 49th American Helicopter Society International Annual Forum*, (St. Louis, MO), pp. 885–901, May 1993.

[166] Schölkopf, B. and Smola, A., *Learning with Kernels Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA: MIT Press, 2002.

[167] Schütz, M. and Sprave, J., "Application of parallel Mixed-Integer evolution strategies with mutation rate pooling," in *Proceedings of the Fifth Annual Conference on Evolutionary Programming*, pp. 345–354, Cambridge, MA: MIT Press, 1996.

[168] Shinoda, P. M., Yeo, H., and Norman, T. R., "Rotor performance of a UH-60 rotor system in the NASA ames 80- by 120-Foot wind tunnel," *Journal of the American Helicopter Society*, vol. 49, pp. 401–413, Oct. 2004.

[169] Shishko, R., "Developing analogy cost estimates for space missions," in *A Collection of Technical Papers - AIAA Space 2004 Conference and Exposition*, vol. 2 of *A Collection of Technical Papers - AIAA Space 2004 Conference and Exposition*, pp. 1530–1537, American Institute of Aeronautics and Astronautics Inc., Sept. 2004.

[170] Simpson, T. W., Booker, A. J., Ghosh, D., Giunta, A. A., Koch, P. N., and Yang, R. J., "Approximation methods in multidisciplinary analysis and optimization: A panel discussion," *Structural and Multidisciplinary Optimization*, vol. 27, pp. 302–313, June 2004.

[171] Simpson, T. W., Poplinski, J. D., Koch, P. N., and Allen, J. K., "Metamodels for Computer-Based engineering design: Survey and recommendations," *Engineering With Computers*, vol. 17, pp. 129–150, July 2001.

[172] Singpurwalla, N. D. and Booker, J. M., "Membership functions and probability measures of fuzzy sets," *Journal of the American Statistical Association*, vol. 99, pp. 867–877, Sept. 2004.

[173] Smith, E. M. B. and Pantelides, C. C., "A symbolic Reformulation/Spatial Branch-and-Bound algorithm for the global optimisation of nonconvex MILNPs," *Computers & Chemical Engineering*, vol. 23, pp. 457–478, May 1999.

[174] Sobek II, D. K., Ward, A. C., and Liker, J. K., "Toyota's principles of Set-Based concurrent engineering," *Sloan Management Review*, vol. 40, pp. 67–83, Jan. 1999.

[175] Sobester, A., Leary, S., and Keane, A., "On the design of optimization strategies based on global response surface approximation models," *Journal of Global Optimization*, vol. 33, pp. 31–59, Sept. 2005.

[176] Sobester, A., Leary, S. J., and Keane, A., "A parallel updating scheme for approximating and optimizing high fidelity computer simulations," *Structural and Multidisciplinary Optimization*, vol. 27, pp. 371–383, July 2004.

[177] Sobieszczanski-Sobieski, J. and Haftka, R. T., "Multidisciplinary aerospace design optimization: Survey of recent developments," *Structural and Multidisciplinary Optimization*, vol. 14, pp. 1–23, Aug. 1997.

[178] Spiegel, M. R., *Theory And Problems Of Probability And Statistics (Schaum S Outline Series)*. New York, NY: McGraw-Hill Education, 2003.

[179] Stanfill, C. and Waltz, D., "Towards Memory-Based reasoning," *Communications of the ACM*, vol. 29, pp. 1213–1228, Dec. 1986.

[180] Sun, H., Kim, Y., Lee, S., and Lee, D., "Aerodynamic design of helicopter rotor blade in forward flight using response surface methodology," *Journal of the American Helicopter Society*, vol. 48, pp. 300–304, Oct. 2003.

[181] Sutherland, W. A., *Introduction to Metric and Topological Spaces*. Oxford, NY: Oxford Science Publications, 2009.

[182] Tawarmalani, M. and Sahinidis, N. V., *Convexification and Global Optimization in Continuous and Mixed-Integer Nonlinear Programming: Theory, Algorithms, Software, and Applications*. Boston, MA: Kluwer Academic Publishers, Oct. 2002.

[183] Tian, D. and Deng, N., "Support vector classification with nominal attributes," in *Proceedings of the Computational Intelligence and Security - International Conference*, vol. 3801, (Xian, China), pp. 586–591, Dec. 2005.

[184] Tipping, M. E., "Bayesian inference: An introduction to principles and practice in machine learning," in *Advanced Lectures on Machine Learning*, vol. 3176, pp. 41–62, Berlin, Germany: Springer Berlin, 2004.

[185] Tishchenko, M. N., Nagaraj, V. T., and Chopra, I., "Preliminary design of transport helicopters," *Journal of the American Helicopter Society*, vol. 48, pp. 71–79, Apr. 2003.

[186] Ulrich, K. T. and Eppinger, S. D., *Product Design and Development*. New York, NY: McGraw-Hill, 2nd ed., 2000.

[187] Vapnik, V. N., *Statistical Learning Theory*. New York, NY: John Wiley & Sons, 1998.

[188] Viana, F. A., Haftka, R. T., Hamman, R., and Venter, G., "Efficient global optimization with experimental data: Revisiting the paper helicopter design," in *Proceedings of the 52nd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference*, (Denver, CO), Apr. 2011.

[189] Vitali, R., Haftka, R. T., and Sankar, B. V., "Multifidelity design of stiffened composite panel with a crack," *Structural and Multidisciplinary Optimization*, vol. 23, pp. 347–356, May 2002.

[190] Vos, J., Rizzi, A., Darracq, D., and Hirschel, E., "Navier-Stokes solvers in european aircraft design," *Progress in Aerospace Sciences*, vol. 38, Nov. 2002.

[191] Vu, N. A., Lee, J. W., and Shu, J. I., "Aerodynamic design optimization of helicopter rotor blades including airfoil shape for hover performance," *Chinese Journal of Aeronautics*, vol. 26, pp. 1–8, Jan. 2013.

[192] Wagner, T., Emmerich, N., Deutz, A., and Ponweiser, W., "On Expected-Improvement criteria for Model-Based Multi-Objective optimization," in *11th International Conference on Parallel Problem Solving from Nature - PPSN XI*, (Krakow, Poland), pp. 718–727, Sept. 2010.

[193] Walsh, J. L., Bingham, G. J., and Riley, M. F., "Optimization methods applied to the aerodynamic design of helicopter rotor blades," *Journal of the American Helicopter Society*, vol. 32, pp. 39–44, Oct. 1987.

[194] Wang, G. G. and Shan, S., "Review of metamodeling techniques in support of engineering design optimization," *Journal of Mechanical Design*, vol. 129, pp. 370–380, Apr. 2007.

[195] Wang, J. and Yin, Z., "A ranking Selection-Based particle swarm optimizer for engineering design optimization problems," *Structural and Multidisciplinary Optimization*, vol. 37, pp. 131–147, Dec. 2008.

[196] Wells, V. L. and Han, A. Y., "Acoustic design of rotor blades using a genetic algorithm," in *AGARD Symposium on Aerodynamics and Aeroacoustics of Rotorcraft*, (Berlin, Germany), pp. 1–10, Oct. 1994.

[197] Westerlund, T. and Pettersson, F., "An extended cutting plane method for solving convex MINLP problems," *Computers & Chemical Engineering*, vol. 19, Supplement 1, pp. 131–136, 1995.

[198] Westerlund, T., Pettersson, F., and Grossmann, I., "Optimization of pump configurations as a MINLP problem," *Computers and Chemical Engineering*, vol. 18, pp. 845–858, Sept. 1994.

[199] While, L., Bradstreet, L., and Barone, L., "A fast way of calculating exact hypervolumes," *IEEE Transactions on Evolutionary Computation*, vol. 16, pp. 86–95, Feb. 2012.

[200] While, L., Hingston, P., Barone, L., and Huband, S., "A faster algorithm for calculating hypervolume," *IEEE Transactions on Evolutionary Computation*, vol. 10, pp. 29–38, Feb. 2006.

[201] Wilson, D. R. and Martinez, T. R., "Improved heterogeneous distance functions," *Journal of Artificial Intelligence Research*, vol. 6, pp. 1–34, Jan. 1997.

[202] Xie, W., Nelson, B., and Staum, J., "The influence of correlation functions on stochastic kriging metamodels," in *Proceedings of the 2010 Winter Simulation Conference*, (Baltimore, MD), pp. 1067–1078, Dec. 2010.

[203] Yan, L., Shen, K., and Hu, S., "Solving mixed integer nonlinear programming problems with Line-Up competition algorithm," *Computers & Chemical Engineering*, vol. 28, pp. 2647–2657, Nov. 2004.

[204] Yang, Q. and Ding, S., "Novel algorithm to calculate hypervolume indicator of pareto approximation set," in *Advanced Intelligent Computing Theories and Applications: With Aspects of Contemporary Intelligent Computing Techniques*, vol. 2, pp. 235–244, Berlin, Germany: Springer, 2007.

[205] Yeo, H., Bousman, W. G., and Johnson, W., "Performance analysis of a utility helicopter with standard and advanced rotors," *Journal of the American Helicopter Society*, vol. 49, pp. 250–270, July 2004.

[206] Yiqing, L., Xigang, Y., and Yongjian, L., "An improved PSO algorithm for solving Non-Convex NLP/MINLP problems with equality constraints," *Computers & Chemical Engineering*, vol. 31, pp. 153–162, Jan. 2007.

[207] Yokota, T., Gen, M., and Li, Y. X., "Genetic algorithm for Non-Linear mixed integer programming problems and its applications," *Computers and Industrial Engineering*, vol. 30, pp. 905–917, Sept. 1996.

[208] Young, P., Parkinson, S., and Lees, M., "Simplicity out of complexity in environmental modelling: Occam's razor revisited," *Journal of Applied Statistics*, vol. 23, no. 2-3, pp. 165–210, 1996.

[209] Yuan, K. A. and Friedmann, P. P., "Structural optimization for vibratory loads reduction of composite helicopter rotor blades with advanced geometry tips," *Journal of the American Helicopter Society*, vol. 43, pp. 246–256, July 1998.

[210] Zadeh, L. A., "Fuzzy sets," *Information and Control*, vol. 8, pp. 338–535, 1965.

[211] Zahedi, F., "The analytic hierarchy process - a survey of the method and its applications," *Interfaces*, vol. 16, pp. 96–108, July 1986.

[212] Zeidler, E., *Nonlinear Functional Analysis and its Applications*. New York, NY: Springer-Verlag, 1985.

[213] Zitzler, E., *Evolutionary Algorithms for Multiobjective Optimization: Methods and Applications*. PhD thesis, Swiss Federal Institute of Technology Zurich, Zurich, Switzerland, 1999.

[214] Zitzler, E., Brockhoff, D., and Thiele, L., "The hypervolume indicator revisited: On the design of Pareto-Compliant indicators via weighted integration," vol. 4403 LNCS of *Lecture Notes in Computer Science*, (Matsushima, Japan), pp. 862–876, Mar. 2007.

[215] Zitzler, E., Thiele, L., Laumanns, M., Fonseca, C. M., and da Fonseca, V. G., "Performance assessment of multiobjective optimizers: An analysis and review," *IEEE Transactions on Evolutionary Computation*, vol. 7, pp. 117–32, Apr. 2003.

[216] Zwicky, F., *Morphological Analysis and Construction*. New York, NY: Wiley-Interscience, 1948.