



Research paper

SMT 2.0: A Surrogate Modeling Toolbox with a focus on hierarchical and mixed variables Gaussian processes

Paul Saves^{a,b,*}, Rémi Lafage^{a,1}, Nathalie Bartoli^{a,1}, Youssef Diouane^{c,1}, Jasper Bussemaker^{d,1}, Thierry Lefebvre^{a,1}, John T. Hwang^{e,1}, Joseph Morlier^{f,1}, Joaquim R.R.A. Martins^{g,1}

^a ONERA/DTIS, Université de Toulouse, Toulouse, France

^b ISAE-SUPAERO, Université de Toulouse, Toulouse, France

^c Polytechnique Montréal, Montréal, QC, Canada

^d German Aerospace Center (DLR), Institute of System Architectures in Aeronautics, Hamburg, Germany

^e University of California San Diego, Department of Mechanical and Aerospace Engineering, La Jolla, CA, USA

^f ICA, Université de Toulouse, ISAE-SUPAERO, INSA, CNRS, MINES ALBI, UPS, Toulouse, France

^g University of Michigan, Department of Aerospace Engineering, Ann Arbor, MI, USA

ARTICLE INFO

Dataset link: https://colab.research.google.com/github/SMTorg/smt/blob/master/tutorial/NotebookRunTestCases_Paper_SMT_v2.ipynb

Keywords:

Surrogate modeling

Gaussian process

Kriging

Hierarchical problems

Hierarchical and mixed-categorical inputs

Meta variables

ABSTRACT

The Surrogate Modeling Toolbox (SMT) is an open-source Python package that offers a collection of surrogate modeling methods, sampling techniques, and a set of sample problems. This paper presents SMT 2.0, a major new release of SMT that introduces significant upgrades and new features to the toolbox. This release adds the capability to handle mixed-variable surrogate models and hierarchical variables. These types of variables are becoming increasingly important in several surrogate modeling applications. SMT 2.0 also improves SMT by extending sampling methods, adding new surrogate models, and computing variance and kernel derivatives for Kriging. This release also includes new functions to handle noisy and use multi-fidelity data. To the best of our knowledge, SMT 2.0 is the first open-source surrogate library to propose surrogate models for hierarchical and mixed inputs. This open-source software is distributed under the New BSD license.²

1. Motivation and significance

With the increasing complexity and accuracy of numerical models, it has become more challenging to run complex simulations and computer codes [1,2]. As a consequence, surrogate models have been recognized as a key tool for engineering tasks such as design space exploration, uncertainty quantification, and optimization [3]. In practice, surrogate models are used to reduce the computational effort of these tasks by replacing expensive numerical simulations with closed-form approximations [4, Ch. 10]. To build such a model, we start by evaluating the original expensive simulation at a set of points through a Design of Experiments (DoE). Then, the corresponding evaluations are used to build the surrogate model according to the chosen approximation, such as Kriging, quadratic interpolation, or least squares regression.

The Surrogate Modeling Toolbox (SMT) is an open-source framework that provides functions to efficiently build surrogate models [5].

Kriging models (also known as Gaussian processes) that take advantage of derivative information are one of SMT's key features [6]. Numerical experiments have shown that SMT achieved lower prediction error and computational cost than Scikit-learn [7] and UQLab [8] for a fixed number of points [9]. SMT has been applied to rocket engine coaxial-injector optimization [10], aircraft engine consumption modeling [11], numerical integration [12], multi-fidelity sensitivity analysis [13], high-order robust finite elements methods [14,15], planning for photovoltaic solar energy [16], wind turbines design optimization [17], porous material optimization for a high pressure turbine vane [18], chemical process design [19] and many other applications.

In systems engineering, architecture-level choices significantly influence the final system performance, and therefore, it is desirable to consider such choices in the early design phases [20]. Architectural choices are parameterized with discrete design variables; examples include the selection of technologies, materials, component connections,

* Corresponding author at: ONERA/DTIS, Université de Toulouse, Toulouse, France.

E-mail addresses: paul.saves@onera.fr (P. Saves), remi.lafage@onera.fr (R. Lafage), nathalie.bartoli@onera.fr (N. Bartoli), youssef.diouane@polymtl.ca (Y. Diouane), jasper.bussemaker@dlr.de (J. Bussemaker), thierry.lefebvre@onera.fr (T. Lefebvre), jhwang@eng.ucsd.edu (J.T. Hwang), joseph.morlier@isae-supaero.fr (J. Morlier), jrram@umich.edu (J.R.R.A. Martins).

¹ All authors contributed to this work, research and manuscript.

² <https://github.com/SMTorg/SMT>

Table 1

Comparison of software packages for hierarchical and mixed Kriging models. ✓ = implemented. * = user-defined.

Package	BoTorch	Dakota	DiceKriging	KerGP	LVGP	Parmoo	Spearmint	SMT 2.0
Reference	[25]	[26]	[27]	[32]	[28]	[29]	[30]	This paper
License	MIT	EPL	GPL	GPL	GPL	BSD	GNU	BSD
Language	Python	C	R	R	R	Python	Python	Python
Mixed var.	✓	✓	✓	✓	✓	✓	✓	✓
GD kernel	✓	✓	✓	*				✓
CR kernel					✓	✓	✓	✓
HH kernel				✓				✓
EHH kernel				*				✓
Hierarchical var.								✓

and number of instantiated elements. When design problems include both discrete variables and continuous variables, they are said to have *mixed variables*.

When architectural choices lead to different sets of design variables, we have *hierarchical variables* [21,22]. For example, consider different aircraft propulsion architectures [23]. A conventional gas turbine would not require a variable to represent a choice in the electrical power source, while hybrid or pure electric propulsion would require such a variable. The relationship between the choices and the sets of variables can be represented by a hierarchy.

Handling hierarchical and mixed variables requires specialized surrogate modeling techniques [24]. To address these needs, SMT 2.0 is offering researchers and practitioners a collection of cutting-edge tools to build surrogate models with continuous, mixed and hierarchical variables. The main objective of this paper is to detail the new enhancements that have been added in this release compared to the original SMT 0.2 release [5].

There are two new major capabilities in SMT 2.0: the ability to build surrogate models involving mixed variables and the support for hierarchical variables within Kriging models. To handle mixed variables in Kriging models, existing libraries such as BoTorch [25], Dakota [26], DiceKriging [27], LVGP [28], Parmoo [29], and Spearmint [30] implement simple mixed models by using either continuous relaxation (CR), also known as *one-hot encoding* [30], or a Gower distance (GD) based correlation kernel [31]. KerGP [32] (developed in R) implements more general kernels but there is no Python open-source toolbox that implements more general kernels to deal with mixed variables, such as the homoscedastic hypersphere (HH) [33] and exponential homoscedastic hypersphere (EHH) [34] kernels. Such kernels require the tuning of a large number of hyperparameters but lead to more accurate Kriging surrogates than simpler mixed kernels [34,35]. SMT 2.0 implements all these kernels (CR, GD, HH, and EHH) through a unified framework and implementation. To handle hierarchical variables, no library in the literature can build peculiar surrogate models except SMT 2.0, which implements two Kriging methods for these variables. Notwithstanding, most softwares are compatible with a naïve strategy called the imputation method [24] but this method lacks depth and depends on arbitrary choices. This is why Hutter and Osborne [21] proposed a first kernel, called Arc-Kernel which in turn was generalized by Horn et al. [36] with a new kernel called the Wedge-Kernel [37]. None of these kernels are available in any open-source modeling software. Furthermore, thanks to the framework introduced in Audet et al. [38], our proposed kernels are sufficiently general so that all existing hierarchical kernels are included within it. Section 4 describes the two kernels implemented in SMT 2.0 that are referred as SMT Arc-Kernel and SMT Alg-Kernel. In particular, Alg-Kernel is a novel hierarchical kernel introduced in this paper. Table 1 outlines the main features of the state-of-the-art modeling software that can handle hierarchical and mixed variables.

SMT 2.0 introduces other enhancements, such as additional sampling procedures, new surrogate models, new Kriging kernels (and their derivatives), Kriging variance derivatives, and an adaptive criterion for high-dimensional problems. SMT 2.0 adds applications of Bayesian

optimization (BO) with hierarchical and mixed variables or noisy co-Kriging that have been successfully applied to aircraft design [39], data fusion [40], and structural design [41]. The SMT 2.0 interface is more user-friendly and offers an improved and more detailed documentation for users and developers.³ SMT 2.0 is hosted publicly⁴ and can be directly imported within Python scripts. It is released under the New BSD License and runs on Linux, MacOS, and Windows operating systems. Regression tests are run automatically for each operating system whenever a change is committed to the repository. In short, SMT 2.0 builds on the strengths of the original SMT package while adding new features. On one hand, the emphasis on derivatives (including prediction, training and output derivatives) is maintained and improved in SMT 2.0. On the other hand, this new release includes support for hierarchical and mixed variables Kriging based models. For the sake of reproducibility, an open-source notebook is available that gathers all the methods and results presented on this paper.⁵

The remainder of the paper is organized as follows. First, we introduce the organization and the main implemented features of the release in Section 2. Then, we describe the mixed-variable Kriging model with an example in Section 3. Similarly, we describe and provide an example for a hierarchical-variable Kriging model in Section 4. The Bayesian optimization models and applications are described in Section 5. Finally, we describe the other relevant contributions in Section 6 and conclude in Section 7.

2. SMT 2.0 : an improved surrogate modeling toolbox

From a software point of view, SMT 2.0 maintains and improves the modularity and generality of the original SMT version [5]. In this section, we describe the software as follows. Section 2.1 describes the legacy of SMT 0.2. Then, Section 2.2 describes the organization of the repository. Finally, Section 2.3 shows the new capabilities implemented in the SMT 2.0 update.

2.1. Background on SMT former version: SMT 0.2

SMT [5] is an open-source collaborative work originally developed by ONERA, NASA Glenn, ISAE-SUPAERO/ICA and the University of Michigan. Now, both Polytechnique Montréal and the University of California San Diego are also contributors. SMT 2.0 updates and extends the original SMT repository capabilities among which the original publication [5] focuses on different types of derivatives for surrogate models detailed hereafter.

³ <http://smt.readthedocs.io/en/latest>

⁴ <https://github.com/SMTorg/smt>

⁵ https://github.com/SMTorg/smt/tree/master/tutorial/NotebookRunTestCases_Paper_SMT_v2.ipynb

Table 2

Impact of using Numba on training time of the hierarchical Goldstein problem. Speedup is calculated excluding the JIT compilation time, as this step is only needed once after SMT installation.

Training set	Without numba	Numba	Speedup	JIT overhead
15 points	1.3 s	1.1 s	15%	24 s
150 points	38 s	7.4 s	80%	23 s

A Python surrogate modeling framework with derivatives. One of the original main motivations for SMT was derivative support. In fact, none of the existing packages for surrogate modeling such as Scikit-learn in Python [7], SUMO in Matlab [42] or GPML in Matlab and Octave [43] focuses on derivatives. Three types of derivatives are distinguished: prediction derivatives, training derivatives, and output derivatives. SMT also includes new models with derivatives such as Kriging with Partial Least Squares (KPLS) [44] and Regularized Minimal-energy Tensor-product Spline (RMTS) [3]. These developed derivatives were even used in a novel algorithm called Gradient-Enhanced Kriging with Partial Least Squares (GEKPLS) [6] to use with adjoint methods, for example [45].

Software architecture, documentation, and automatic testing. SMT is organized along three main sub-modules that implement a set of sampling techniques (**sampling methods**), benchmarking functions (**problems**), and surrogate modeling techniques (**surrogate models**). The toolbox documentation⁶ is created using reStructuredText and Sphinx, a documentation generation package for Python, with custom extensions. Code snippets in the documentation pages are taken directly from actual tests in the source code and are automatically updated. The output from these code snippets and tables of options are generated dynamically by custom Sphinx extensions. This leads to high-quality documentation with minimal effort. Along with user documentation, developer documentation is also provided to explain how to contribute to SMT. This includes a list of API methods for the **SurrogateModel**, **SamplingMethod**, and **Problem** classes, that must be implemented to create a new surrogate modeling method, sampling technique, or benchmarking problem. When a developer submits a pull request, it is merged only after passing the automated tests and receiving approval from at least one reviewer. The repository on GitHub⁷ is linked to continuous integration tests (GitHub Actions) for Windows, Linux and MacOS, to a coverage test on coveralls.io and to a dependency version check for Python with Dependabot. Various parts of the source code have been accelerated using Numba [46], an LLVM-based just-in-time (JIT) compiler for numpy-heavy Python code. Numba is applied to conventional Python code using function decorators, thereby minimizing its impact on the development process and not requiring an additional build step. For a mixed Kriging surrogate with 150 training points, a speedup of up to 80% is observed, see Table 2. The JIT compilation step only needs to be done once when installing or upgrading SMT and adds an overhead of approximately 24 s on a typical workstation. In this paper, all results are obtained using an Intel® Xeon® CPU E5-2650 v4 @ 2.20 GHz core and 128 GB of memory with a Broadwell-generation processor front-end and a compute node of a peak power of 844 GFlops.

2.2. Organization of SMT 2.0

The main features of the open-source repository SMT 2.0 are described in Fig. 1. More precisely, Sampling Methods, Problems and Surrogate models are kept from SMT 0.2 and two new sections Models applications and Interactive notebooks have been added to the architecture of the code. These sections are

highlighted in blue and detailed on Fig. 1. The new major features implemented in SMT 2.0 are highlighted in lavender whereas the legacy features that were already in present in the original publication for SMT 0.2 [5] are in black.

2.3. New features within SMT 2.0

The main objective of this new release is to enable Kriging surrogate models for use with both hierarchical and mixed variables. Moreover, for each of these five sub-modules described in Section 2.2, several improvements have been made between the original version and the SMT 2.0 release.

Hierarchical and mixed design space. A new design space definition class `DesignSpace` has been added that implements hierarchical and mixed functionalities. Design variables can either be continuous (`FloatVariable`), ordered (`OrdinalVariable`) or categorical (`CategoricalVariable`). The integer type (`IntegerVariable`) represents a special case of the ordered variable, specified by bounds (inclusive) rather than a list of possible values. The hierarchical structure of the design space can be defined using `declare_decreed_var`: this function declares that a variable is a decreed variable that is activated when the associated meta variable takes one of a set of specified values, see Section 4 for background. The `DesignSpace` class also implements mechanisms for sampling valid design vectors (i.e. design vectors that adhere to the hierarchical structure of the design space) using any of the below-mentioned samplers, for correcting and imputing design vectors, and for requesting which design variables are acting in a given design vector. Correction ensures that variables have valid values (e.g. integers for discrete variables) [24], and imputation replaces non-acting variables by some default value (0 for discrete variables, mid-way between the bounds for continuous variables in SMT 2.0) [47].

Sampling. SMT implements three methods for sampling. The first one is a naïve approach, called `Random` that draws uniformly points along every dimension. The second sampling method is called `Full Factorial` and draws a point for every cross combination of variables, to have an “exhaustive” design of experiments. The last one is the `Latin Hypercube Sampling (LHS)` [48] that draws a point in every Latin square parameterized by a certain criterion. For LHS, a new criterion to manage the randomness has been implemented and the sampling method was adapted for multi-fidelity and mixed or hierarchical variables. More details about the new sampling techniques are given in Section 6.1.

Problems. SMT implements two new engineering problems: a mixed variant of a cantilever beam described in Section 3 and a hierarchical neural network described in Section 4.

Surrogate models. In order to keep up with state-of-art, several releases done from the original version developed new options for the already existing surrogates. In particular, compared to the original publication [5], SMT 2.0 adds gradient-enhanced neural networks [45] and marginal Gaussian process [49] models to the list of available surrogates. More details about the new models are given in Section 6.2.

Applications. Several applications have been added to the toolbox to demonstrate the surrogate models capabilities. The most relevant application is efficient global optimization (EGO), a Bayesian optimization algorithm [50,51]. EGO optimizes expensive-to-evaluate black-box problems with a chosen surrogate model and a chosen optimization criterion [52]. The usage of EGO with hierarchical and mixed variables is described in Section 5.

⁶ <https://smt.readthedocs.org>

⁷ <https://github.com/SMTorg/smt>

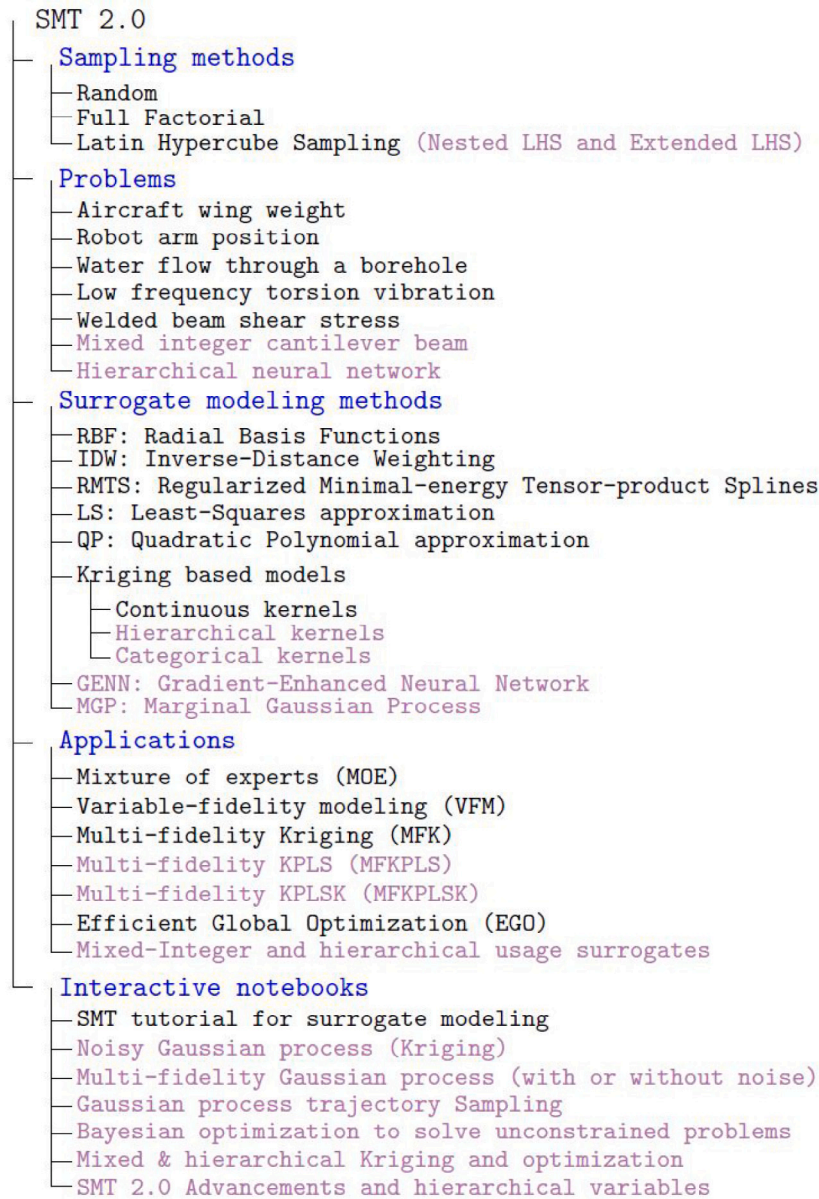


Fig. 1. Functionalities of SMT 2.0. The new major features implemented in SMT 2.0 compared to SMT 0.2 are highlighted with the lavender color.

Interactive notebooks. These tutorials introduce and explain how to use the toolbox for different surrogate models and applications.⁸ Every tutorial is available both as a .ipynb file and directly on Google colab.⁹ In particular, a hierarchical and mixed variables dedicated notebook is available to reproduce the results presented on this paper.¹⁰

In the following, Section 3 details the Kriging based surrogate models for mixed variables, and Section 4 presents our new Kriging surrogate for hierarchical variables. Section 5 details the EGO application and the other new relevant features aforementioned are described succinctly in Section 6.

3. Surrogate models with mixed variables in SMT 2.0

As mentioned in Section 1, design variables can be either of continuous or discrete type, and a problem with both types is a mixed-variable problem. Discrete variables can be ordinal or categorical. A discrete variable is *ordinal* if there is an order relation within the set of possible values. An example of an ordinal design variable is the number of engines in an aircraft. A possible set of values in this case could be 2, 4, 8. A discrete variable is *categorical* if no order relation is known between the possible choices the variable can take. One example of a categorical variable is the color of a surface. A possible example of a set of choices could be blue, red, green. The possible choices are called the *levels* of the variable.

Several methods have been proposed to address the recent increase interest in mixed Kriging based models [30–33,35,39,53,54]. The main difference from a continuous Kriging model is in the estimation of

⁸ <https://github.com/SMTorg/smt/tree/master/tutorial>

⁹ <https://colab.research.google.com/github/SMTorg/smt/>

¹⁰ https://github.com/SMTorg/smt/tree/master/tutorial/NotebookRunTestCases_Paper_SMT_v2.ipynb

Table 3
Categorical kernels implemented in SMT 2.0.

Name	$\kappa(\phi)$	$\Phi(\theta_i)$	# of hyperparam.
SMT GD	$\exp(-\phi)$	$[\Phi(\theta_i)]_{j,j} := \frac{1}{2}\theta_i ; [\Phi(\theta_i)]_{j \neq j'} := 0$	1
SMT CR	$\exp(-\phi)$	$[\Phi(\theta_i)]_{j,j} := [\theta_i]_{j,j} ; [\Phi(\theta_i)]_{j \neq j'} := 0$	L_i
SMT EHH	$\exp(-\phi)$	$[\Phi(\theta_i)]_{j,j} := 0 ; [\Phi(\theta_i)]_{j \neq j'} := \frac{\log \epsilon}{2} ([C(\theta_i)C(\theta_i)^T]_{j,j'} - 1)$	$\frac{1}{2}(L_i)(L_i - 1)$
SMT HH	ϕ	$[\Phi(\theta_i)]_{j,j} := 1 ; [\Phi(\theta_i)]_{j \neq j'} := [C(\theta_i)C(\theta_i)^T]_{j,j'}$	$\frac{1}{2}(L_i)(L_i - 1)$

the categorical correlation matrix, which is critical to determine the mean and variance predictions. As mentioned in Section 1, approaches such as CR [30,39], continuous latent variables [54], and GD [31] use a kernel-based method to estimate the correlation matrix. Other methods estimate the correlation matrix by modeling the correlation entries directly [32,35,53], such as HH [33] and EHH [34]. The HH correlation kernel is of particular interest because it generalizes simpler kernels such as CR and GD [34]. In SMT 2.0, the correlation kernel is an option that can be set to either CR (CONT_RELAX_KERNEL), GD (GOWER_KERNEL), HH (HOMO_HSPHERE_KERNEL) or EHH (EXP_HOMO_HSPHERE_KERNEL).

3.1. Mixed Gaussian processes

The continuous and ordinal variables are both treated similarly in SMT 2.0 with a continuous kernel, where the ordinal values are converted to continuous through relaxation. For categorical variables, four models (GD, CR, EHH and HH) can be used in SMT 2.0 if specified by the API. This is why we developed a unified mathematical formulation that allows a unique implementation for any model.

Denote l the number of categorical variables. For a given $i \in \{1, \dots, l\}$, the i th categorical variable is denoted c_i and its number of levels is denoted L_i . The hyperparameter matrix peculiar to this variable c_i is

$$\theta_i = \begin{bmatrix} [\theta_i]_{1,1} & & & & \\ [\theta_i]_{1,2} & [\theta_i]_{2,2} & & & \\ \vdots & \ddots & \ddots & \ddots & \\ [\theta_i]_{1,L_i} & \dots & [\theta_i]_{L_i-1,L_i} & [\theta_i]_{L_i,L_i} \end{bmatrix}, \quad \text{Sym.}$$

and the categorical parameters are defined as $\theta^{cat} = \{\theta_1, \dots, \theta_l\}$. For two given inputs in the DoE, for example, the r th and s th points, let c_i^r and c_i^s be the associated categorical variables taking respectively the ℓ_r^i and the ℓ_s^i level on the categorical variable c_i . The categorical correlation kernel is defined by

$$k^{cat}(c^r, c^s, \theta^{cat}) = \prod_{i=1}^l \kappa([\Phi(\theta_i)]_{\ell_r^i, \ell_s^i}) \kappa([\Phi(\theta_i)]_{\ell_s^i, \ell_r^i}) \kappa([\Phi(\theta_i)]_{\ell_r^i, \ell_r^i}) \kappa([\Phi(\theta_i)]_{\ell_s^i, \ell_s^i}) \quad (1)$$

where κ is either a positive definite kernel or identity and $\Phi(\cdot)$ is a symmetric positive definite (SPD) function such that the matrix $\Phi(\theta_i)$ is SPD if θ_i is SPD. For an exponential kernel, Table 3 gives the parameterizations of Φ and κ that correspond to GD, CR, HH, and EHH kernels. The complexity of these different kernels depends on the number of hyperparameters that characterizes them. As defined by Saves et al. [34], for every categorical variable $i \in \{1, \dots, l\}$, the matrix $C(\theta_i) \in \mathbb{R}^{L_i \times L_i}$ is lower triangular and built using a hypersphere decomposition [55,56] from the symmetric matrix $\theta_i \in \mathbb{R}^{L_i \times L_i}$ of hyperparameters. The variable ϵ is a small positive constant and the variable θ_i denotes the only positive hyperparameter that is used for the Gower distance kernel.

Another Kriging based model that can use mixed variables is Kriging with partial least squares (KPLS) [57]. KPLS adapts Kriging to high dimensional problems by using a reduced number of hyperparameters thanks to a projection into a smaller space. Also, for a general surrogate, not necessarily Kriging, SMT 2.0 uses continuous relaxation to allow whatever model to handle mixed variables. For example, we can use mixed variables with least squares (LS) or quadratic polynomial (QP) models. We now illustrate the abilities of the toolbox in terms of mixed modeling over an engineering test case.

Table 4

Results of the cantilever beam models [34, Table 4].

Categorical kernel	Displacement error (cm)	Likelihood	# of hyperparam.
SMT GD	1.3861	111.13	3
SMT CR	1.1671	155.32	14
SMT EHH	0.1613	236.25	68
SMT HH	0.2033	235.66	68

3.2. An engineering design test-case

A classic engineering problem commonly used for model validation is the beam bending problem [32,58]. This problem is illustrated on Fig. 2(a) and consists of a cantilever beam in its linear range loaded at its free end with a force F . As in Cheng et al. [58], the Young modulus is $E = 200$ GPa and the chosen load is $F = 50$ kN. Also, as in Roustant et al. [32], 12 possible cross-sections can be used. These 12 sections consist of 4 possible shapes that can be either hollow, thick or full as illustrated in Fig. 2(b).

To compare the mixed Kriging models of SMT 2.0, we draw a 98 point LHS as training set and the validation set is a grid of $12 \times 30 \times 30 = 10800$ points. For the four implemented methods, displacement error (computed with a root-mean-square error criterion), likelihood, number of hyperparameters and computational time for every model are shown in Table 4. For the continuous variables, we use the square exponential kernel. More details are found in [34]. As expected, the complex EHH and HH models lead to a lower displacement error and a higher likelihood value, but use more hyperparameters and increase the computational cost compared to GD and CR. On this test case, the kernel EHH is easier to optimize than HH but in general, they are similar in terms of performance. Also, by default SMT 2.0 uses CR as it is known to be a good trade-off between complexity and performance [59].

4. Surrogate models with hierarchical variables in SMT 2.0

To introduce the newly developed Kriging model for hierarchical variables implemented in SMT 2.0, we present the general mathematical framework for hierarchical and mixed variables established by Audet et al. [38]. In SMT 2.0, two variants of our new method are implemented, namely SMT Alg-Kernel and SMT Arc-Kernel. In particular, the SMT Alg-Kernel is a novel correlation kernel introduced in this paper.

4.1. The hierarchical variables framework

A problem structure is classified as hierarchical when the sets of active variables depend on architectural choices. This occurs frequently in industrial design problems. In hierarchical problems, we can classify variables as neutral, meta (also known as dimensional) or decreed (also known as conditionally active) as detailed in Audet et al. [38]. Neutral variables are the variables that are not affected by the hierarchy whereas the value assigned to meta variables determines which decreed variables are activated. For example, a meta variable could be the number of engines. If the number of engines changes, the number of decreed bypass ratios that every engine should specify also changes.

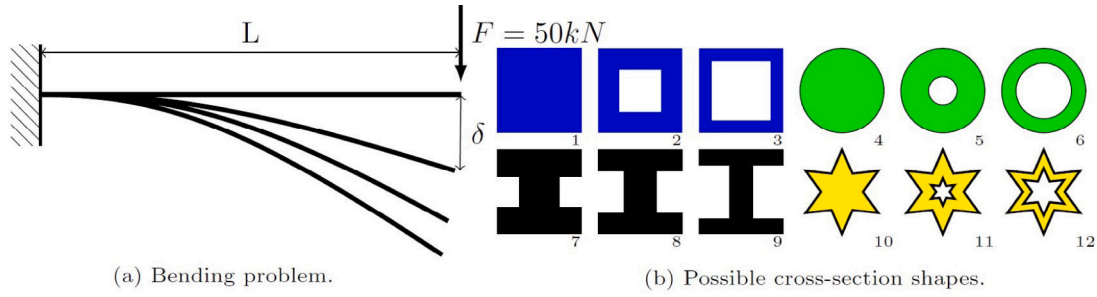


Fig. 2. Cantilever beam problem [34, Figure 6].

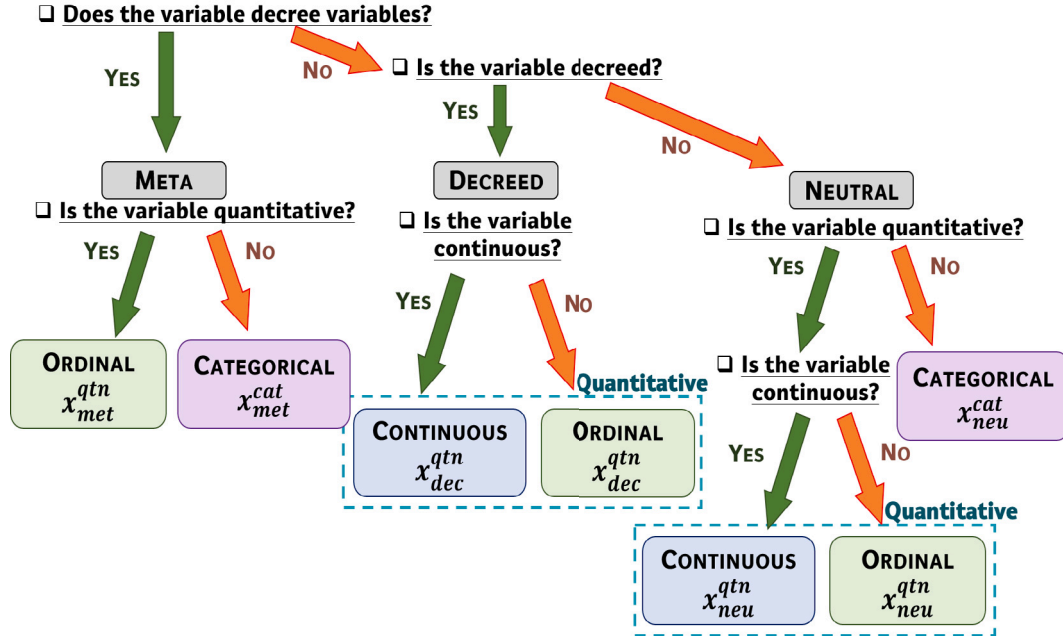


Fig. 3. Variables classification as used in SMT 2.0.

However, the wing aspect ratio being neutral, it is not affected by this hierarchy.

Problems involving hierarchical variables are generally dependant on discrete architectures and as such involve mixed variables. Hence, in addition to their role (neutral, meta or decreed), each variable also has a variable type amongst categorical, ordinal or continuous. For the sake of simplicity and because both continuous and ordinal variables are treated similarly [34], we chose to regroup them as *quantitative variables*. For instance, the neutral variables x_{neu} may be partitioned into different variable types, such that $x_{neu} = (x_{neu}^{cat}, x_{neu}^{qtn})$ where x_{neu}^{cat} represents the categorical variables and x_{neu}^{qtn} are the quantitative ones. The variable classification scheme in SMT 2.0 is detailed in Fig. 3.

To explain the framework and the new Kriging model, we illustrate the inputs variables of the model using a classical machine learning problem related to the hyperparameters optimization of a fully-connected Multi-Layer Perceptron (MLP) [38] on Fig. 4. In Table 5, we detail the input variables of the model related to the MLP problem (i.e., the hyperparameters of the neural network, together with their types and roles). To keep things clear and concise, the chosen problem is a simplification of the original problem developed by Audet et al. [38]. Regarding the MLP problem of Fig. 4 and following the classification scheme of Fig. 3, we start by separating the input variables according to their role. In fact,

1. changing the number of hidden layers modifies the number of inputs variables. Therefore, “# of hidden layers” is a meta variable.

2. The number of neurons in the hidden layer number k is either included or excluded. For example, the “# of neurons in the 3rd layer” would be excluded for an input that only has 2 hidden layers. Therefore, “# of neurons hidden layer k ” are decreed variables.
3. The “Learning rate”, “Momentum”, “Activation function” and “Batch size” are not affected by the hierarchy choice. Therefore, they are neutral variables.

According to their types, the MLP input variables can be classified as follows:

4. The meta variable “# of hidden layers” is an integer and, as such, is represented by the component x_{met}^{qtn} .
5. The decreed variables “# of neurons hidden layer k ” are integers and, as such, are represented by the component x_{dec}^{qtn} .
6. The “Learning rate”, “Momentum”, “Activation function” and “Batch size” are, respectively, continuous, for the first two (every value between two bounds), categorical (qualitative between three choices) and integer (quantitative between 6 choices). Therefore, the “Activation function” and the “Momentum” are represented by the component x_{neu}^{cat} . The “Learning rate” and the “Batch size” are represented by the component x_{neu}^{qtn} .

To model hierarchical variables, as proposed in [38], we separate the input space \mathcal{X} as $(\mathcal{X}_{neu}, \mathcal{X}_{met}, \mathcal{X}_{dec})$ where $\mathcal{X}_{dec} = \bigcup_{x_{met} \in \mathcal{X}_{met}} \mathcal{X}_{inc}(x_{met})$.

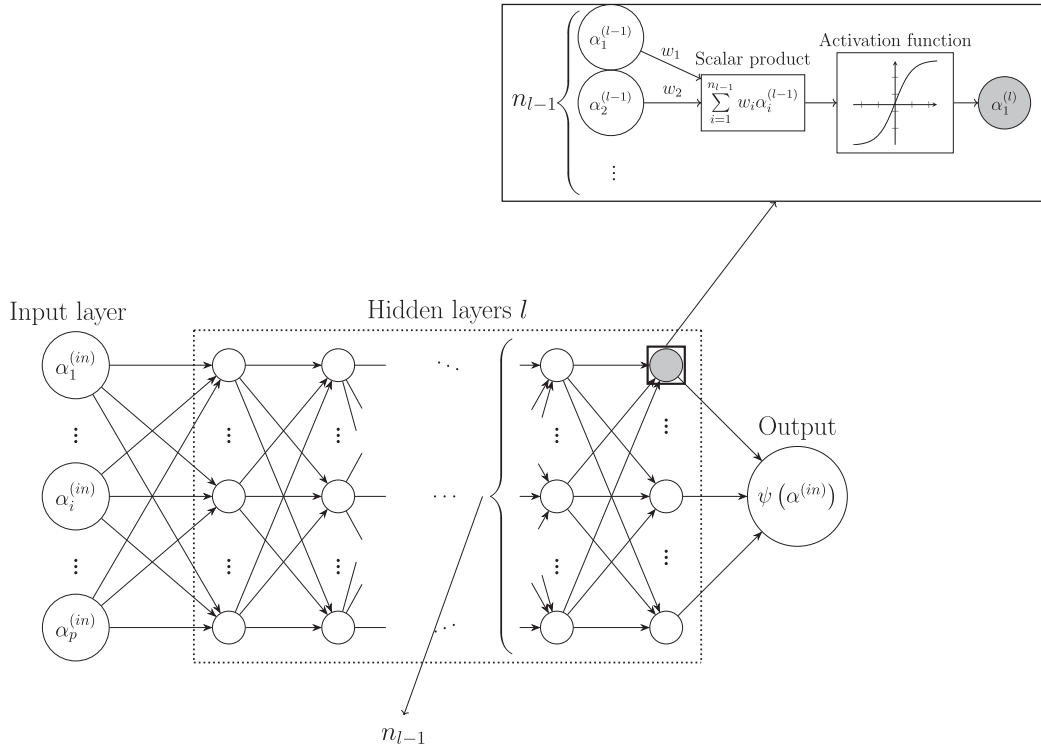


Fig. 4. The Multi-Layer Perceptron (MLP) problem.
Source: Figure adapted from [38, Figure 1].

Table 5

A detailed description of the variables in the MLP problem.

MLP Hyperparameters	Variable	Domain	Type	Role
Learning rate	r	$[10^{-5}, 10^{-2}]$	FLOAT	NEUTRAL
Momentum	α	$[0, 1]$	FLOAT	NEUTRAL
Activation function	a	$\{ReLU, Sigmoid, Tanh\}$	ENUM	NEUTRAL
Batch size	b	$\{8, 16, \dots, 128, 256\}$	ORD	NEUTRAL
# of hidden layers	l	$\{1, 2, 3\}$	ORD	META
# of neurons hidden layer k	n_k	$\{50, 51, \dots, 55\}$	ORD	DECREED

Hence, for a given point $x \in \mathcal{X}$, one has $x = (x_{\text{neu}}, x_{\text{met}}, x_{\text{inc}}(x_{\text{met}}))$, where $x_{\text{neu}} \in \mathcal{X}_{\text{neu}}$, $x_{\text{met}} \in \mathcal{X}_{\text{met}}$ and $x_{\text{inc}}(u_{\text{met}}) \in \mathcal{X}_{\text{inc}}(u_{\text{met}})$ are defined as follows:

- The components $x_{\text{neu}} \in \mathcal{X}_{\text{neu}}$ gather all neutral variables that are not impacted by the meta variables but needed. For example, in the MLP problem, \mathcal{X}_{neu} gathers the possible learning rates, momentum, activation functions and batch sizes. Namely, from Table 5, $\mathcal{X}_{\text{neu}} = [10^{-5}, 10^{-2}] \times [0, 1] \times \{ReLU, Sigmoid, Tanh\} \times \{8, 16, \dots, 256\}$.
- The components x_{met} gather the meta (also known as dimensional) variables that determine the inclusion or exclusion of other variables. For example, in the MLP problem, \mathcal{X}_{met} represents the possible numbers of layers in the MLP. Namely, from Table 5, $\mathcal{X}_{\text{met}} = \{1, 2, 3\}$.
- The components $x_{\text{inc}}(x_{\text{met}})$, contain the decreed variables whose inclusion (decreed-included) or exclusion (decreed-excluded) is determined by the values of the meta components x_{met} . For example, in the MLP problem, \mathcal{X}_{dec} represents the number of neurons in the decreed layers. Namely, from Table 5, $\mathcal{X}_{\text{inc}}(x_{\text{met}} = 3) = [50, 55]^3$, $\mathcal{X}_{\text{inc}}(x_{\text{met}} = 2) = [50, 55]^2$ and $\mathcal{X}_{\text{inc}}(x_{\text{met}} = 1) = [50, 55]$.

4.2. A Kriging model for hierarchical variables

In this section, a new method to build a Kriging model with hierarchical variables is introduced based on the framework aforementioned. The proposed methods are included in SMT 2.0.

4.2.1. Motivation and state-of-the-art

Assuming that the decreed variables are quantitative, Hutter and Osborne [21] proposed several kernels for the hierarchical context. A classic approach, called the imputation method (Imp-Kernel) leads to an efficient paradigm in practice that can be easily integrated into a more general framework as proposed by Bussemaker et al. [24]. However, this kernel lacks depth and depends on arbitrary choices. Therefore, Hutter and Osborne [21] also proposed a more general kernel, called Arc-Kernel and Horn et al. [36] generalized this kernel even more and proposed a new formulation called the Wedge-Kernel [37]. The drawbacks of these two methods are that they add some extra hyperparameters for every decreed dimension (respectively one extra hyperparameter for the Arc-Kernel and two hyperparameters for the Wedge-Kernel) and that they need a normalization according to the bounds. More recently, Pelamatti et al. [60] developed a hierarchical kernel for Bayesian optimization. However, our work is also more general thanks to the framework introduced earlier [38] that considers variable-wise formulation and is more flexible as we are allowing sub-problems to be intersecting.

In the following, we describe our new method to build a correlation kernel for hierarchical variables. In particular, we introduce a new algebraic kernel called Alg-Kernel that behaves like the Arc-Kernel whilst correcting most of its drawbacks. In particular, our kernel does not add any hyperparameters, and the normalization is handled in a natural way.

4.2.2. A new hierarchical correlation kernel

For modeling purposes, we assume that the decreed space is quantitative, i.e., $\mathcal{X}_{\text{dec}} = \mathcal{X}_{\text{dec}}^{\text{qnt}}$. Let $u \in \mathcal{X}$ be an input point partitioned as

$u = (u_{\text{neu}}, u_{\text{met}}, u_{\text{inc}}(u_{\text{met}}))$ and, similarly, $v \in \mathcal{X}$ is another input such that $v = (v_{\text{neu}}, v_{\text{met}}, v_{\text{inc}}(v_{\text{met}}))$. The new kernel k that we propose for hierarchical variables is given by

$$k(u, v) = k_{\text{neu}}(u_{\text{neu}}, v_{\text{neu}}) \times k_{\text{met}}(u_{\text{met}}, v_{\text{met}}) \times k_{\text{met,dec}}([u_{\text{met}}, u_{\text{inc}}(u_{\text{met}})], [v_{\text{met}}, v_{\text{inc}}(v_{\text{met}})]), \quad (2)$$

where k_{neu} , k_{met} and $k_{\text{met,dec}}$ are as follows:

- k_{neu} represents the neutral kernel that encompasses both categorical and quantitative neutral variables, i.e., k_{neu} can be decomposed into two parts $k_{\text{neu}}(u_{\text{neu}}, v_{\text{neu}}) = k^{\text{cat}}(u_{\text{neu}}^{\text{cat}}, v_{\text{neu}}^{\text{cat}})k^{\text{qnt}}(u_{\text{neu}}^{\text{qnt}}, v_{\text{neu}}^{\text{qnt}})$. The categorical kernel, denoted k^{cat} , could be any Symmetric Positive Definite (SPD) [34] mixed kernel (see Section 3). For the quantitative (integer or continuous) variables, a distance-based kernel is used. The chosen quantitative kernel (Exponential, Matérn, ...), always depends on a given distance d . For example, the n -dimensional exponential kernel gives
- $$k^{\text{qnt}}(u^{\text{qnt}}, v^{\text{qnt}}) = \prod_{i=1}^n \exp(-d(u_i^{\text{qnt}}, v_i^{\text{qnt}})). \quad (3)$$
- k_{met} is the meta variables related kernel. It is also separated into two parts: $k_{\text{met}}(u_{\text{met}}, v_{\text{met}}) = k^{\text{cat}}(u_{\text{met}}^{\text{cat}}, v_{\text{met}}^{\text{cat}})k^{\text{qnt}}(u_{\text{met}}^{\text{qnt}}, v_{\text{met}}^{\text{qnt}})$ where the quantitative kernel is ordered and not continuous because meta variables take value in a finite set.
 - $k_{\text{met,dec}}$ is an SPD kernel that models the correlations between the meta levels (all the possible subspaces) and the decreed variables. In what comes next, we detailed this kernel.

4.2.3. Towards an algebraic meta-decreed kernel

Meta-decreed kernels like the imputation kernel or the Arc-Kernel were first proposed in [21,47] and the distance-based kernels such as Arc-Kernel or Wedge-Kernel [37] were proven to be SPD. Nevertheless, to guarantee this SPD property, the same hyperparameters are used to model the correlations between the meta levels and between the decreed variables [47]. For this reason, the Arc-Kernel includes additional continuous hyperparameters which makes the training of the GP models more expensive and introduces more numerical stability issues. In this context, we are proposing a new algebraic meta-decreed kernel (denoted as Alg-Kernel) that enjoys similar properties as Arc-Kernel but without using additional continuous hyperparameters nor rescaling. In the SMT 2.0 release, we included Alg-Kernel and a simpler version of Arc-Kernel that do not relies on additional hyperparameters.

Our proposed Alg-Kernel kernel is given by

$$k_{\text{met,dec}}^{\text{alg}}([u_{\text{met}}, u_{\text{inc}}(u_{\text{met}})], [v_{\text{met}}, v_{\text{inc}}(v_{\text{met}})]) = k_{\text{met}}^{\text{alg}}(u_{\text{met}}, v_{\text{met}}) \times k_{\text{dec}}^{\text{alg}}(u_{\text{inc}}(u_{\text{met}}), v_{\text{inc}}(v_{\text{met}})). \quad (4)$$

Mathematically, we could consider that there is only one meta variable whose levels correspond to every possible included subspace. Let I_{sub} denotes the components indices of possible subspaces, the subspaces parameterized by the meta component u_{met} are defined as $\mathcal{X}_{\text{inc}}(u_{\text{met}} = l)$, $l \in I_{\text{sub}}$. It follows that the fully extended continuous decreed space writes as $\mathcal{X}_{\text{dec}} = \bigcup_{l \in I_{\text{sub}}} \mathcal{X}_{\text{inc}}(u_{\text{met}} = l)$ and I_{dec} is the set of the associated indices. Let $I_{u,v}^{\text{inter}}$ denotes the set of components related to the space $\mathcal{X}_{\text{inc}}(u_{\text{met}}, v_{\text{met}})$ containing the variables decreed-included in both $\mathcal{X}_{\text{inc}}(u_{\text{met}})$ and $\mathcal{X}_{\text{inc}}(v_{\text{met}})$.

Since the decreed variables are quantitative, one has

$$k_{\text{dec}}^{\text{alg}}(u_{\text{inc}}(u_{\text{met}}), v_{\text{inc}}(v_{\text{met}})) = k^{\text{qnt}}(u_{\text{inc}}(u_{\text{met}}), v_{\text{inc}}(v_{\text{met}})) = \prod_{i \in I_{u,v}^{\text{inter}}} k^{\text{qnt}}([u_{\text{inc}}(u_{\text{met}})]_i, [v_{\text{inc}}(v_{\text{met}})]_i) \quad (5)$$

The construction of the quantitative kernel k^{qnt} depends on a given distance denoted d^{alg} . The kernel $k_{\text{met}}^{\text{alg}}$ is an induced meta kernel that depends on the same distance d^{alg} to preserve the SPD property of

$k_{\text{met,dec}}^{\text{alg}}$. For every $i \in I_{\text{dec}}$, if $i \in I_{u,v}^{\text{inter}}$, the new algebraic distance is given by

$$d^{\text{alg}}([u_{\text{inc}}(u_{\text{met}})]_i, [v_{\text{inc}}(v_{\text{met}})]_i) = \left(\frac{2|u_{\text{inc}}(u_{\text{met}})]_i - v_{\text{inc}}(v_{\text{met}})]_i|}{\sqrt{[u_{\text{inc}}(u_{\text{met}})]_i^2 + 1} \sqrt{[v_{\text{inc}}(v_{\text{met}})]_i^2 + 1}} \right) \theta_i, \quad (6)$$

where $\theta_i \in \mathbb{R}^+$ is a continuous hyperparameter. Otherwise, if $i \in I_{\text{dec}}$ but $i \notin I_{u,v}^{\text{inter}}$, there should be a non-zero residual distance between the two different subspaces $\mathcal{X}_{\text{inc}}(u_{\text{met}})$ and $\mathcal{X}_{\text{inc}}(v_{\text{met}})$ to ensure the kernel SPD property. To have a residual not depending on the decreed values, our model considers that there is a unit distance

$$d^{\text{alg}}([u_{\text{inc}}(u_{\text{met}})]_i, [v_{\text{inc}}(v_{\text{met}})]_i) = 1.0 \theta_i, \quad \forall i \in I_{\text{dec}} \setminus I_{u,v}^{\text{inter}}.$$

The induced meta kernel $k_{\text{met}}^{\text{alg}}(u_{\text{met}}, v_{\text{met}})$ to preserve the SPD property of k^{alg} is defined as:

$$k_{\text{met}}^{\text{alg}}(u_{\text{met}}, v_{\text{met}}) = \prod_{i \in I_{\text{met}}} k^{\text{qnt}}(1.0 \theta_i). \quad (7)$$

Not only our kernel of Eq. (2) uses less hyperparameters than the Arc-Kernel (as we cut off its extra parameters) but it is also a more flexible kernel as it allows different kernels for meta and decreed variables. Moreover, another advantage of our kernel is that it is numerically more stable thanks to the new non-stationary [61] algebraic distance defined in Eq. (7) [62]. Our proposed distance also does not need any rescaling by the bounds to have values between 0 and 1.

In what comes next, we will refer to the implementation of the kernels Arc-Kernel and Alg-Kernel by SMT Arc-Kernel and SMT Alg-Kernel. We note also that the implementation of SMT Arc-Kernel differs slightly from the original Arc-Kernel as we fixed some hyperparameters to 1 in order to avoid adding extra hyperparameters and use the formulation of Eq. (2) and rescaling of the data.

4.2.4. Illustration on the MLP problem

In this section, we illustrate the hierarchical Arc-Kernel on the MLP example. For that sake, we consider two design variables u and v such that $u = (2.10^{-4}, 0.9, \text{ReLU}, 16, 2, 55, 51)$ and $v = (5.10^{-3}, 0.8, \text{Sigmoid}, 64, 3, 50, 54, 53)$. Since the value of u_{met} (i.e., the number of hidden layers) differs from one point to another (namely, 2 for u and 3 for v), the associated variables $u_{\text{inc}}(u_{\text{met}})$ have either 2 or 3 variables for the number of neurons in each layer (namely 55 and 51 for u , and 50, 54 and 53 for the point v). In our case, 8 hyperparameters ($[R_1]_{1,2}, \theta_1, \dots, \theta_7$) will have to be optimized where k is given by Eq. (2). These 7 hyperparameters can be described using our proposed framework as follows:

- For the neutral components, we have $u_{\text{neu}} = (2.10^{-4}, 0.9, \text{ReLU}, 16)$ and $v_{\text{neu}} = (5.10^{-3}, 0.8, \text{Sigmoid}, 64)$. Therefore, for a categorical matrix kernel R_1 and a square exponential quantitative kernel,

$$k_{\text{neu}}(u_{\text{neu}}, v_{\text{neu}}) = k^{\text{cat}}(u_{\text{neu}}^{\text{cat}}, v_{\text{neu}}^{\text{cat}})k^{\text{qnt}}(u_{\text{neu}}^{\text{qnt}}, v_{\text{neu}}^{\text{qnt}}) = [R_1]_{1,2} \exp[-\theta_1(2.10^{-4} - 5.10^{-3})^2] \exp[-\theta_2(0.9 - 0.8)^2] \exp[-\theta_3(16 - 64)^2].$$

The values $[R_1]_{1,2}$, θ_1 , θ_2 and θ_3 need to be optimized. Here, $[R_1]_{1,2}$ is the correlation between "ReLU" and "Sigmoid".

- For the meta components, we have $u_{\text{met}} = 2$ and $v_{\text{met}} = 3$. Therefore, for a square exponential quantitative kernel,

$$k_{\text{met}}(u_{\text{met}}, v_{\text{met}}) = k^{\text{cat}}(u_{\text{met}}^{\text{cat}}, v_{\text{met}}^{\text{cat}})k^{\text{qnt}}(u_{\text{met}}^{\text{qnt}}, v_{\text{met}}^{\text{qnt}}) = \exp[-\theta_4(3 - 2)^2].$$

The value θ_4 needs to be optimized.

- For the meta-decreed kernel, we have $[u_{\text{met}}, u_{\text{inc}}(u_{\text{met}})] = [2, (55, 51)]$ and $[v_{\text{met}}, v_{\text{inc}}(v_{\text{met}})] = [3, (50, 54, 53)]$ which gives

$$k_{\text{met,dec}}^{\text{alg}}([u_{\text{met}}, u_{\text{inc}}(u_{\text{met}})], [v_{\text{met}}, v_{\text{inc}}(v_{\text{met}})]) \\ = k_{\text{met}}^{\text{alg}}(2, 3) k_{\text{dec}}^{\text{alg}}((55, 51), (50, 54, 53)).$$

The distance $d^{\text{alg}}(51, 54) = \left(\frac{2 \times |51 - 54|}{\sqrt{51^2 + 1} \sqrt{54^2 + 1}} \right) \theta_6 = 2.178 \cdot 10^{-3} \theta_6$. In general, for surrogate models, and in particular in SMT 2.0, the input data are normalized. With a unit normalization from $[50, 55]$ to $[0, 1]$, we would have $d^{\text{alg}}(0.2, 0.8) = \left(\frac{2 \times 0.6}{\sqrt{0.2^2 + 1} \sqrt{0.8^2 + 1}} \right) \theta_6 = 0.919 \theta_6$. Similarly, we have, between 55 and 50, $d^{\text{alg}}(0, 1) = 1.414 \theta_5$. Hence, for a square exponential quantitative kernel, one gets

$$k_{\text{met,dec}}^{\text{alg}}([u_{\text{met}}, u_{\text{inc}}(u_{\text{met}})], [v_{\text{met}}, v_{\text{inc}}(v_{\text{met}})]) \\ = \exp[-\theta_7] \times \exp[-1.414 \theta_5] \times \exp[-0.919 \theta_6],$$

where the meta induced component is $k_{\text{met}}^{\text{alg}}(u_{\text{met}}, v_{\text{met}}) = \exp[-\theta_7]$ because the decreed value 53 in v has nothing to be compared with in u as in Eq. (7). The values θ_5 , θ_6 and θ_7 need to be optimized which complete the description of the hyperparameters. We note that for the MLP problem, Alg-Kernel models use 10 hyperparameters whereas the Arc-Kernel would require 12 hyperparameters without the meta kernel (θ_4) but with 3 extra decreed hyperparameters and the Wedge-Kernel would require 15 hyperparameters. For deep learning applications, a more complex perceptron with up to 10 hidden layers would require 17 hyperparameters with SMT 2.0 models against 26 for Arc-Kernel and 36 for Wedge-Kernel. The next section illustrates the interest of our method to build a surrogate model for this neural network engineering problem.

4.3. A neural network test-case using SMT 2.0

In this section, we apply our models to the hyperparameters optimization of a MLP problem aforementioned in Fig. 4. Within SMT 2.0 an example illustrates this MLP problem. For the sake of showing the Kriging surrogate abilities, we implemented a dummy function with no significance to replace the real black-box that would require training a whole Neural Network (NN) with big data. This function requires a number of variables that depends on the value of the meta variable, i.e the number of hidden layers. To simplify, we have chosen only 1, 2 or 3 hidden layers and therefore, we have 3 decreed variables but deep neural networks could also be investigated as our model can tackle a few dozen variables. A test case (*test_hierarchical_variables_NN*) shows that our model SMT Alg-Kernel interpolates the data properly, checks that the data dimension is correct and also asserts that the inactive decreed variables have no influence over the prediction. In Fig. 5 we illustrate the usage of Kriging surrogates with hierarchical and mixed variables based on the implementation of SMT 2.0 for *test_hierarchical_variables_NN*.

To compare the hierarchical models of SMT 2.0 (SMT Alg-Kernel and SMT Arc-Kernel) with the state-of-the-art imputation method previously used on industrial application (Imp-Kernel) [24], we draw a 99 point LHS (33 points by meta level) as a training set and the validation set is a LHS of $3 \times 1000 = 3000$ points. For the Imp-Kernel, the decreed-excluded values are replaced by 52 because the mean value 52.5 is not an integer (by default, SMT rounds to the floor value). For the three methods, the precision (computed with a root-mean-square error RMSE criterion), the likelihood and the computational time are shown in Table 6. For this problem, we can see that SMT Alg-kernel gives better performance than the imputation method or SMT Arc-kernel. Also, as all methods use the same number of hyperparameters, they have similar time performances. A direct application of

Table 6

Results on the neural network model.

Hierarchical method	Prediction error (RMSE)	Likelihood	# of hyperparam.
SMT Alg-kernel	3.7610	176.11	10
SMT Arc-kernel	4.9208	162.01	10
Imp-Kernel	4.5455	170.64	10

our modeling method is Bayesian optimization to perform quickly the hyperparameter optimization of a neural network [63].

5. Bayesian optimization within SMT 2.0

Efficient global optimization (EGO) is a sequential Bayesian optimization algorithm designed to find the optimum of a black-box function that may be expensive to evaluate [52]. EGO starts by fitting a Kriging model to an initial DoE, and then uses an acquisition function to select the next point to evaluate. The most used acquisition function is the expected improvement. Once a new point has been evaluated, the Kriging model is updated. Successive updates increase the model accuracy over iterations. This enrichment process repeats until a stopping criterion is met.

Because SMT 2.0 implements Kriging models that handle mixed and hierarchical variables, we can use EGO to solve problems involving such design variables. Other Bayesian optimization algorithms often adopt approaches based on solving subproblems with continuous or non-hierarchical Kriging. This subproblem approach is less efficient and scales poorly, but it can only solve simple problems. Several Bayesian optimization software packages can handle mixed or hierarchical variables with such a subproblem approach. The packages include BoTorch [25], SMAC [65], Trieste [66], HEBO [67], OpenBox [68], and Dragonfly [69].

5.1. A mixed optimization problem

Fig. 6 compares the four EGO methods implemented in SMT 2.0: SMT GD, SMT CR, SMT EHH and SMT HH. The mixed test case that illustrates Bayesian optimization is a toy test case [64] detailed in Appendix A. This test case has two variables, one continuous and one categorical with 10 levels. To assess the performance of our algorithm, we performed 20 runs with different initial DoE sampled by LHS. Every DoE consists of 5 points and we chose a budget of 55 infill points. Fig. 6(a) plots the convergence curves for the four methods. In particular, the median is the solid line, and the first and third quantiles are plotted in dotted lines. To visualize better the data dispersion, the boxplots of the 20 best solutions after 20 evaluations are plotted in Fig. 6(b). As expected, the more a method is complex, the better the optimization. Both SMT HH and SMT EHH converged in around 18 evaluations whereas SMT CR and SMT GD take around 26 iterations as shown on Fig. 6(a). Also, the more complex the model, the closer the optimum is to the real value as shown on Fig. 6(b).

In Fig. 7 we illustrate how to use EGO with mixed variables based on the implementation of SMT 2.0. The illustrated problem is a mixed variant of the Branin function [70].

Note that a dedicated notebook is available to reproduce the results presented in this paper and the mixed integer notebook also includes an extra mechanical application with composite materials [41].¹¹

¹¹ https://colab.research.google.com/github/SMTorg/smt/blob/master/tutorial/SMT_MixedInteger_application.ipynb

```

from smt.sampling_methods import LHS
from smt.surrogate_models import KRG, MixIntKernelType, \
    MixHrcKernelType, DesignSpace, FloatVariable, \
    IntegerVariable, OrdinalVariable, CategoricalVariable
from smt.applications.mixed_integer import \
    MixedIntegerSamplingMethod, MixedIntegerKrigingModel
import f1_NN, f2_NN, f3_NN #dummy example
def test_hierarchical_variables_NN(self):
    def dummy_f(x):
        if x[0] == 1:
            y=(f1_NN(x[1],x[2],x[3],2**x[4],x[5]))
        elif x[0] == 2:
            y=(f2_NN(x[1],x[2],x[3],2**x[4],x[5],x[6]))
        elif x[0] == 3:
            y=(f3_NN(x[1],x[2],x[3],2**x[4],x[5],x[6],x[7]))
        return y
    # Define the mixed hierarchical design space
    ds = DesignSpace([
        IntegerVariable(1, 3), FloatVariable(1e-5, 1e-2),
        FloatVariable(0, 1),
        CategoricalVariable(["ReLU", "Sigmoid","Tanh"]),
        IntegerVariable(3, 8), IntegerVariable(50, 55),
        IntegerVariable(50, 55), IntegerVariable(50, 55),
    ])
    # activate x5 when x0 in [2, 3]; x6 when x0 == 3
    ds.declare_decreed_var(
        decreed_var=6, meta_var=0, meta_value=[2, 3])
    ds.declare_decreed_var(7, meta_var=0, meta_value=3)
    #Perform the mixed integer sampling
    sampling = MixedIntegerSamplingMethod(
        LHS, ds, criterion="ese", random_state=42)
    Xt = sampling(100)
    Yt = dummy_f(Xt)
    #Build the surrogate
    sm = MixedIntegerKrigingModel(
        surrogate=KRG(design_space=ds, corr="abs_exp",
            categorical_kernel=MixIntKernelType.HOMO_HSPHERE,
            hierarchical_kernel=MixHrcKernelType.ALG_KERNEL)
    sm.set_training_values(Xt, Yt)
    sm.train()
    # Check prediction accuracy
    y_s = sm.predict_values(Xt)
    pred_RMSE = np.linalg.norm(y_s - Yt) / len(Yt)
    y_sv = sm.predict_variances(Xt)
    var_RMSE = np.linalg.norm(y_sv) / len(Yt)
    assert pred_RMSE < 1e-7
    assert var_RMSE < 1e-7

```

Fig. 5. Example of usage of Hierarchical and Mixed Kriging surrogate.

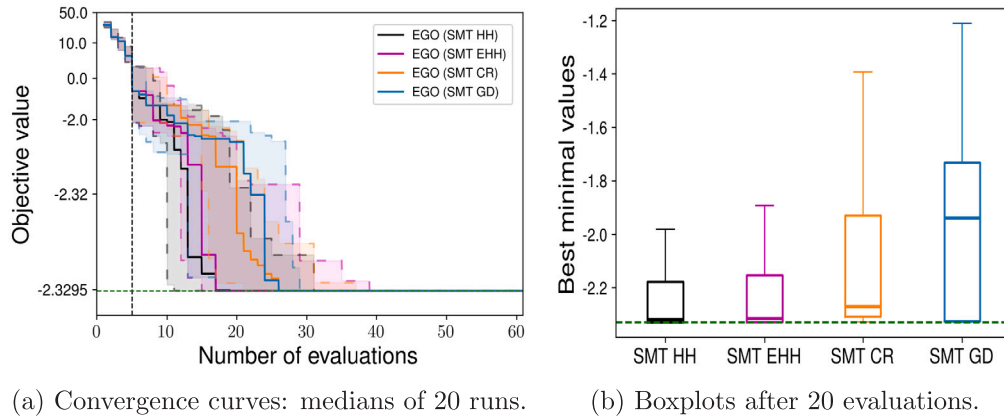


Fig. 6. Optimization results for the Toy function [64].

```
#Import the Mixed Integer API
from smt.surrogate_models import KRG,MixIntKernelType,\
    DesignSpace,FloatVariable,IntegerVariable
from smt.applications.mixed_integer import \
    MixedIntegerSamplingMethod as misamp
#Define the function
from smt.problems import Branin
fun = Branin(ndim=2)
#Define the mixed design space
design_space = DesignSpace([
    IntegerVariable(*fun.xlimits[0]),
    FloatVariable(*fun.xlimits[1]),
])
#Perform a mixed integer sampling with LHS
from smt.sampling_methods import LHS
smp = misamp(LHS,design_space,random_state=42)
xdoe = smp(10)
# Call the Bayesian optimizer
from smt.applications import EGO
criterion = "EI" #'EI' or 'SBO' or 'LCB'
ego = EGO(xdoe=xdoe,
    n_iter=20,
    criterion="EI",
    random_state=42,
    surrogate=KRG(design_space=design_space,
        categorical_kernel=MixIntKernelType.GOWER))
x_opt, y_opt, _, _, _ = ego.optimize(fun=fun)
# Check if the result is correct
self.assertEqual(0.494, float(y_opt), delta=1)
```

Fig. 7. Example of usage of mixed surrogates for Bayesian optimization.

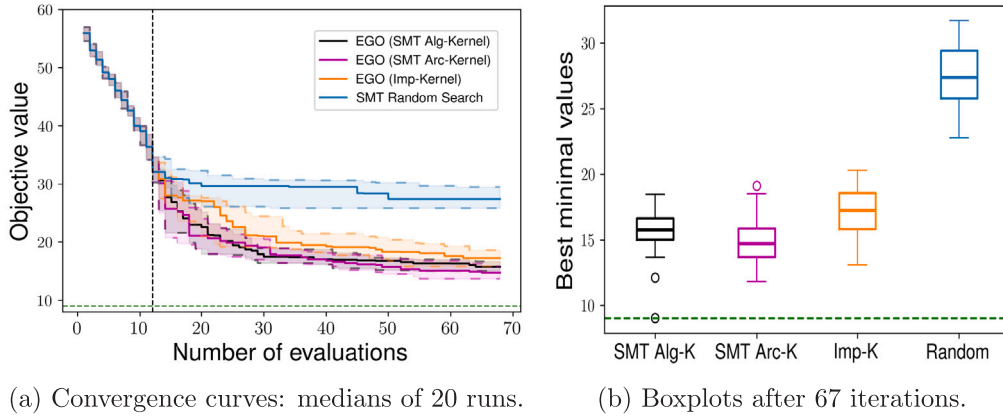


Fig. 8. Optimization results for the hierarchical Goldstein function.

5.2. A hierarchical optimization problem

The hierarchical test case considered in this paper to illustrate Bayesian optimization is a modified Goldstein function [60] detailed in Appendix B. The resulting optimization problem involves 11 variables: 5 are continuous, 4 are integer (ordinal) and 2 are categorical. These variables consist in 6 neutral variables, 1 dimensional (or meta) variable and 4 decreed variables. Depending on the meta variable values, 4 different sub-problems can be identified. The optimization problem is given by:

$$\begin{aligned} \min f(x_{\text{neu}}^{\text{cat}}, x_{\text{neu}}^{\text{qnt}}, x_m^{\text{cat}}, x_{\text{dec}}^{\text{qnt}}) \\ \text{w.r.t. } x_{\text{neu}}^{\text{cat}} = w_2 \in \{0, 1\} \\ x_{\text{neu}}^{\text{qnt}} = (x_1, x_2, x_5, z_3, z_4) \in \{0, 100\}^3 \times \{0, 1, 2\}^2 \\ x_m^{\text{cat}} = w_1 \in \{0, 1, 2, 3\} \\ x_{\text{dec}}^{\text{qnt}} = (x_3, x_4, z_1, z_2) \in \{0, 100\}^2 \times \{0, 1, 2\}^2 \end{aligned} \quad (8)$$

Compared to the model choice of Pelamatti et al. [60], we chose to model x_5 and w_2 as neutral variables even if f does not depend on x_5 when $w_2 = 0$. Other modeling choices are kept; for example, w_2 is a so-called “binary variable” and not a categorical one [71]. Similarly, we also keep the formulation of w_1 as a categorical variable but a better model would be to model it as a “cyclic variable” [72]. The resulting problem is described in Appendix B. To assess the performance of our algorithm, we performed 20 runs with different initial DoE sampled by LHS. Every DoE consists of $n + 1 = 12$ points and we chose a budget of $5n = 55$ infill points. To compare our method with a baseline, we also tested the random search method thanks to the `expand_lhs` new method [40] described in Section 6.1 and we also optimized the Goldstein function using EGO with a classic Kriging model based on imputation method (Imp-Kernel). This method replaces the decreed-excluded variables by their mean values: 50 or 1 respectively for (x_3, x_4) and (z_1, z_2) . Fig. 8(a) plots the convergence curves for the four methods. In particular, the median is the solid line and the first and third quantiles are plotted in dotted lines. To visualize better the corresponding data dispersion, the boxplots of the 20 best solutions are plotted in Fig. 8(b). The results in Fig. 8 show that the hierarchical Kriging models of SMT 2.0 lead to better results than the imputation method or the random search both in terms of final objective value and variance over the 20 runs and in term of convergence rate. More precisely, SMT Arc-Kernel and SMT Alg-Kernel Kriging model gave the best EGO results and managed to converge correctly as shown in Fig. 8(b). More precisely, the 20 sampled DoEs led to similar performance and from one DoE, the method SMT Alg-Kernel managed to find the true minimum. However, this result has not been reproduced in other runs and is therefore not statistically significant. The variance between the runs is of similar magnitude regardless of the considered methods.

6. Other relevant contributions in SMT 2.0

The new release SMT 2.0 introduces several improvements besides Kriging for hierarchical and mixed variables. This section details the most important new contributions. Recall from Section 2.2 that five sub-modules are present in the code: Sampling, Problems, Surrogate Models, Applications and Notebooks.

6.1. Contributions to Sampling

Pseudo-random sampling. The Latin Hypercube Sampling (LHS) is a stochastic sampling technique to generate quasi-random sampling distributions. It is among the most popular sampling method in computer experiments thanks to its simplicity and projection properties with high-dimensional problems. The LHS method uses the `pyDOE` package (Design Of Experiments for Python). Five criteria for the construction of LHS are implemented in SMT. The first four criteria (`center`, `maximin`, `centermaximin`, `correlation`) are the same as in `pyDOE`.¹² The last criterion `ese`, is implemented by the authors of SMT [48]. In SMT 2.0 a new LHS method was developed for the Nested design of experiments (NestedLHS) [73] to use with multi-fidelity surrogates. A new mathematical method (`expand_lhs`) [40] was developed in SMT 2.0 to increase the size of a design of experiments while maintaining the `ese` property. Moreover, we proposed a sampling method for mixed variables, and the aforementioned LHS method was applied to hierarchical variables in Fig. 8.

6.2. Contributions to Surrogate models

New kernels and their derivatives for Kriging. Kriging surrogates are based on hyperparameters and on a correlation kernel. Four correlation kernels are now implemented in SMT 2.0 [74]. In SMT, these correlation functions are absolute exponential (`abs_exp`), Gaussian (`squar_exp`), Matern 5/2 (`matern52`) and Matern 3/2 (`matern32`). In addition, the implementation of gradient and Hessian for each kernel makes it possible to calculate both the first and second derivatives of the GP likelihood with respect to the hyperparameters [5].

Variance derivatives for Kriging. To perform uncertainty quantification for system analysis purposes, it could be interesting to know more about the variance derivatives of a model [75–77]. For that purpose and also to pursue the original publication about derivatives [5], SMT 2.0 extends the derivative support to Kriging variances and kernels.

¹² <https://pythonhosted.org/pyDOE/index.html>

Noisy Kriging. In engineering and in big data contexts with real experiments, surrogate models for noisy data are of significant interest. In particular, there is a growing need for techniques like noisy Kriging and noisy Multi-Fidelity Kriging (MFK) for data fusion [78]. For that purpose, SMT 2.0 has been designed to accommodate Kriging and MFK to noisy data including the option to incorporate heteroscedastic noise (using the `use_het_noise` option) and to account for different noise levels for each data source [40].

Kriging with partial least squares. Beside MGP, for high-dimensional problems, the toolbox implements Kriging with partial least squares (KPLS) [57] and its extension KPLSK [44]. The PLS information is computed by projecting the data into a smaller space spanned by the principal components. By integrating this PLS information into the Kriging correlation matrix, the number of inputs can be scaled down, thereby reducing the number of hyperparameters required. The resulting number of hyperparameters d_e is indeed much smaller than the original problem dimension d . Recently, in SMT 2.0, we extended the KPLS method for multi-fidelity Kriging (MFKPLS and MFKPLSK) [73,79,80]. We also proposed an automatic criterion to choose automatically the reduced dimension d_e based on Wold's R criterion [81]. This criterion has been applied to aircraft optimization with EGO where the number d_e (`n_comp` in the code) for the model is automatically selected at every iteration [39]. Special efforts have been made to accommodate KPLS for multi-fidelity and mixed integer data [79,80].

Marginal Gaussian process. SMT 2.0 implements Marginal Gaussian Process (MGP) surrogate models for high dimensional problems [82]. MGP are Gaussian processes taking into account hyperparameters uncertainty defined as a density probability function. Especially we suppose that the function to model $f : \Omega \mapsto \mathbb{R}$, where $\Omega \subset \mathbb{R}^d$ and d is the number of design variables, lies in a linear embedding \mathcal{A} such as $\mathcal{A} = \{u = Ax, x \in \Omega\}$, $A \in \mathbb{R}^{d \times d_e}$ and $f(x) = f_{\mathcal{A}}(Ax)$ with $f(x) = f_{\mathcal{A}} : \mathcal{A} \mapsto \mathbb{R}$ and $d_e \ll d$. Then, we must use a kernel $k(x, x') = k_{\mathcal{A}}(Ax, Ax')$ whose each component of the transfer matrix A is an hyperparameter. Thus we have $d_e \times d$ hyperparameters to find. Note that d_e is defined as `n_comp` in the code [49].

Gradient-enhanced neural network. The new release SMT 2.0 implements Gradient-Enhanced Neural Network (GENN) models [45]. Gradient-Enhanced Neural Networks (GENN) are fully connected multi-layer perceptrons whose training process was modified to account for gradient information. Specifically, the model is trained to minimize not only the prediction error of the response but also the prediction error of the partial derivatives: the chief benefit of gradient enhancement is better accuracy with fewer training points. Note that GENN applies to regression (single-output or multi-output), but not classification since there is no gradient in that case. The implementation is fully vectorized and uses ADAM optimization, mini-batch, and L2-norm regularization. For example, GENN can be used to learn airfoil geometries from a database. This usage is documented in SMT 2.0.¹³

6.3. Contributions to Applications

Kriging trajectory and sampling. Sampling a GP with high resolution is usually expensive due to the large dimension of the associated covariance matrix. Several methods are proposed to draw samples of a GP on a given set of points. To sample a conditioned GP, the classic method consists in using a Cholesky decomposition (or eigendecomposition) of the conditioned covariance matrix of the process but some numerical computational errors can lead to non SPD matrix. A more recent approach based on Karhunen–Loève decomposition of the covariance kernel with the Nyström method has been proposed in [83]

where the paths can be sampled by generating independent standard Normal distributed samples. The different methods are documented in the tutorial *Gaussian Process Trajectory Sampling* [84].

Parallel Bayesian optimization. Due to the recent progress made in hardware configurations, it has been of high interest to perform parallel optimizations. A parallel criterion called qEI [85] was developed to perform Efficient Global Optimization (EGO): the goal is to be able to run batch optimization. At each iteration of the algorithm, multiple new sampling points are extracted from the known ones. These new sampling points are then evaluated using a parallel computing environment. Five criteria are implemented in SMT 2.0: Kriging Believer (KB), Kriging Believer Upper Bound (KBUB), Kriging Believer Lower Bound (KBLB), Kriging Believer Random Bound (KBRand) and Constant Liar (CLmin) [86].

7. Conclusion

SMT 2.0 introduces significant upgrades to the Surrogate Modeling Toolbox. This new release adds support for hierarchical and mixed variables and improves the surrogate models with a particular focus on Kriging (Gaussian process) models. SMT 2.0 is distributed through an open-source license and is freely available online.¹⁴ We provide documentation that caters to both users and potential developers.¹⁵ SMT 2.0 enables users and developers collaborating on the same project to have a common surrogate modeling tool that facilitates the exchange of methods and reproducibility of results.

SMT has been widely used in aerospace and mechanical modeling applications. Moreover, the toolbox is general and can be useful for anyone who needs to use or develop surrogate modeling techniques, regardless of the targeted applications. SMT is currently the only open-source toolbox that can build hierarchical and mixed surrogate models.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request. Results can be reproduced freely online at https://colab.research.google.com/github/SMTorg/smt/blob/master/tutorial/NotebookRunTestCases_Paper_SMT_v2.ipynb.

Acknowledgments

We want to thank all those who contribute to this release. Namely, M. A. Bouhlel, I. Cardoso, R. Carreira Rufato, R. Charayron, R. Conde Arenzana, S. Dubreuil, A. F. López-Lopera, M. Meliani, M. Menz, N. Moëlle, A. Thouvenot, R. Priem, E. Roux and F. Vergnes. This work is part of the activities of ONERA - ISAE - ENAC joint research group. We also acknowledge the partners institutions: ONERA, NASA Glenn, ISAE-SUPAERO, Institut Clément Ader (ICA), the University of Michigan, Polytechnique Montréal and the University of California San Diego.

The research presented in this paper has been performed in the framework of the AGILE 4.0 project (Towards cyber-physical collaborative aircraft development), funded by the European Union Horizon 2020 research and innovation framework programme under grant agreement n° 815122 and in the COLOSSUS project (Collaborative System of Systems Exploration of Aviation Products, services and

¹³ https://smt.readthedocs.io/en/latest/_src_docs/examples/airfoil_parameters/learning_airfoil_parameters.html

¹⁴ <https://github.com/SMTorg/SMT>

¹⁵ <https://smt.readthedocs.io/en/latest/>

Business Models) funded by the European Union Horizon Europe research and innovation framework programme under grant agreement n° 101097120.

We also are grateful to E. Hallé-Hannan from Polytechnique Montréal for the hierarchical variables framework.

Appendix A. Toy test function

This Appendix gives the detail of the toy function of Section 5.1.¹⁶ First, we recall the optimization problem:

$$\begin{aligned} \min f(x^{\text{cat}}, x^{\text{qnt}}) \\ \text{w.r.t. } x^{\text{cat}} = c_1 \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\} \\ x^{\text{qnt}} = x_1 \in [0, 1] \end{aligned} \quad (\text{A.1})$$

The toy function f is defined as

$$\begin{aligned} f(x, c_1) = & \mathbb{1}_{c_1=0} \cos(3.6\pi(x-2)) + x - 1 \\ & + \mathbb{1}_{c_1=1} 2 \cos(1.1\pi \exp(x)) - \frac{x}{2} + 2 \\ & + \mathbb{1}_{c_1=2} \cos(2\pi x) + \frac{1}{2}x \\ & + \mathbb{1}_{c_1=3} x(\cos(3.4\pi(x-1)) - \frac{x-1}{2}) \\ & + \mathbb{1}_{c_1=4} - \frac{x^2}{2} \\ & + \mathbb{1}_{c_1=5} 2 \cos(0.25\pi \exp(-x^4))^2 - \frac{x}{2} + 1 \\ & + \mathbb{1}_{c_1=6} x \cos(3.4\pi x) - \frac{x}{2} + 1 \\ & + \mathbb{1}_{c_1=7} - x(\cos(3.5\pi x) + \frac{x}{2}) + 2 \\ & + \mathbb{1}_{c_1=8} - \frac{x^5}{2} + 1 \\ & + \mathbb{1}_{c_1=9} - \cos(2.5\pi x)^2 \sqrt{x} - 0.5 \ln(x+0.5) - 1.3 \end{aligned} \quad (\text{A.2})$$

Appendix B. Hierarchical Goldstein test function

This Appendix gives the detail of the hierarchical Goldstein problem of Section 5.2.¹⁷ First, we recall the optimization problem:

$$\begin{aligned} \min f(x^{\text{cat}}_{\text{neu}}, x^{\text{qnt}}_{\text{neu}}, x^{\text{cat}}_m, x^{\text{qnt}}_{\text{dec}}) \\ \text{w.r.t. } x^{\text{cat}}_{\text{neu}} = w_2 \in \{0, 1\} \\ x^{\text{qnt}}_{\text{neu}} = (x_1, x_2, x_5, z_3, z_4) \in [0, 100]^3 \times \{0, 1, 2\}^2 \\ x^{\text{cat}}_m = w_1 \in \{0, 1, 2, 3\} \\ x^{\text{qnt}}_{\text{dec}} = (x_3, x_4, z_1, z_2) \in [0, 100]^2 \times \{0, 1, 2\}^2 \end{aligned} \quad (\text{B.1})$$

The hierarchical and mixed function f is defined as a hierarchical function that depends on f_0 , f_1 , f_2 and $Gold_{\text{cont}}$ as describes in the following.

$$\begin{aligned} f(x_1, x_2, x_3, x_4, z_1, z_2, z_3, z_4, x_5, w_1, w_2) = \\ \mathbb{1}_{w_1=0} f_0(x_1, x_2, z_1, z_2, z_3, z_4, x_5, w_2) \\ + \mathbb{1}_{w_1=1} f_1(x_1, x_2, x_3, z_2, z_3, z_4, x_5, w_2) \\ + \mathbb{1}_{w_1=2} f_2(x_1, x_2, x_4, z_1, z_3, z_4, x_5, w_2) \\ + \mathbb{1}_{w_1=3} Gold_{\text{cont}}(x_1, x_2, x_3, x_4, z_3, z_4, x_5, w_2). \end{aligned} \quad (\text{B.2})$$

Then, the functions f_0 , f_1 and f_2 are defined as mixed variants of $Gold_{\text{cont}}$ as such

$$\begin{aligned} f_0(x_1, x_2, z_1, z_2, z_3, z_4, x_5, w_2) = \\ \mathbb{1}_{z_2=0} (\mathbb{1}_{z_1=0} Gold_{\text{cont}}(x_1, x_2, 20, 20, z_3, z_4, x_5, w_2) \\ + \mathbb{1}_{z_1=1} Gold_{\text{cont}}(x_1, x_2, 50, 20, z_3, z_4, x_5, w_2) \\ + \mathbb{1}_{z_1=2} Gold_{\text{cont}}(x_1, x_2, 80, 20, z_3, z_4, x_5, w_2)) \\ \mathbb{1}_{z_2=1} (\mathbb{1}_{z_1=0} Gold_{\text{cont}}(x_1, x_2, 20, 50, z_3, z_4, x_5, w_2) \\ + \mathbb{1}_{z_1=1} Gold_{\text{cont}}(x_1, x_2, 50, 50, z_3, z_4, x_5, w_2) \\ + \mathbb{1}_{z_1=2} Gold_{\text{cont}}(x_1, x_2, 80, 50, z_3, z_4, x_5, w_2)) \\ \mathbb{1}_{z_2=2} (\mathbb{1}_{z_1=0} Gold_{\text{cont}}(x_1, x_2, 20, 80, z_3, z_4, x_5, w_2) \\ + \mathbb{1}_{z_1=1} Gold_{\text{cont}}(x_1, x_2, 50, 80, z_3, z_4, x_5, w_2) \\ + \mathbb{1}_{z_1=2} Gold_{\text{cont}}(x_1, x_2, 80, 80, z_3, z_4, x_5, w_2)) \end{aligned} \quad (\text{B.3})$$

$$\begin{aligned} f_1(x_1, x_2, x_3, z_2, z_3, z_4, x_5, w_2) = \\ \mathbb{1}_{z_2=0} Gold_{\text{cont}}(x_1, x_2, x_3, 20, z_3, z_4, x_5, w_2) \\ + \mathbb{1}_{z_2=1} Gold_{\text{cont}}(x_1, x_2, x_3, 50, z_3, z_4, x_5, w_2) \\ + \mathbb{1}_{z_2=2} Gold_{\text{cont}}(x_1, x_2, x_3, 80, z_3, z_4, x_5, w_2) \end{aligned}$$

$$\begin{aligned} f_2(x_1, x_2, x_4, z_1, z_3, z_4, x_5, w_2) = \\ \mathbb{1}_{z_1=0} Gold_{\text{cont}}(x_1, x_2, 20, x_4, z_3, z_4, x_5, w_2) \\ + \mathbb{1}_{z_1=1} Gold_{\text{cont}}(x_1, x_2, 50, x_4, z_3, z_4, x_5, w_2) \\ + \mathbb{1}_{z_1=2} Gold_{\text{cont}}(x_1, x_2, 80, x_4, z_3, z_4, x_5, w_2) \end{aligned}$$

To finish with, the function $Gold_{\text{cont}}$ is given by

$$\begin{aligned} Gold_{\text{cont}}(x_1, x_2, x_3, x_4, z_3, z_4, x_5, w_2) = & 53.3108 + 0.184901x_1 \\ & - 5.02914x_1^3 \cdot 10^{-6} + 7.72522x_1^{z_3} \cdot 10^{-8} - 0.0870775x_2 - 0.106959x_3 \\ & + 7.98772x_3^{z_4} \cdot 10^{-6} + 0.00242482x_4 + 1.32851x_4^3 \cdot 10^{-6} - 0.00146393x_1x_2 \\ & - 0.00301588x_1x_3 - 0.00272291x_1x_4 + 0.0017004x_2x_3 + 0.0038428x_2x_4 \\ & - 0.000198969x_3x_4 + 1.86025x_1x_2x_3 \cdot 10^{-5} - 1.88719x_1x_2x_4 \cdot 10^{-6} \\ & + 2.50923x_1x_3x_4 \cdot 10^{-5} - 5.62199x_2x_3x_4 \cdot 10^{-5} + w_2 \left(5 \cos \left(\frac{2\pi}{100}x_5 \right) - 2 \right). \end{aligned} \quad (\text{B.4})$$

Appendix C. Supplementary data

More at https://colab.research.google.com/github/SMTorg/smt/blob/master/tutorial/NotebookRunTestCases_Paper_SMT_v2.ipynb.

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.advensoft.2023.103571>.

References

- [1] Mader CA, Martins JRRA, Alonso JJ, van der Weide E. ADjoint: An approach for the rapid development of discrete adjoint solvers. *AIAA J* 2008;46:863–73.
- [2] Kennedy M, O'Hagan A. Bayesian calibration of computer models. *J R Stat Soc Ser B Stat Methodol* 2001;63:425–64.
- [3] Hwang JT, Martins JRRA. A fast-prediction surrogate model for large datasets. *Aerosp Sci Technol* 2018;75:74–87.
- [4] Martins JRRA, Ning A. *Engineering design optimization*. Cambridge University Press; 2021.
- [5] Bouhlel MA, Hwang JT, Bartoli N, Lafage R, Morlier J, Martins JRA. A Python surrogate modeling framework with derivatives. *Adv Eng Softw* 2019;135:102662.
- [6] Bouhlel MA, Martins J. Gradient-enhanced kriging for high-dimensional problems. *Eng Comput* 2019;35:157–73.
- [7] Pedregosa F, Varoquaux G, Gramfort A, Thirion VMB, Grisel O, et al. Scikit-learn: Machine learning in Python. *J Mach Learn Res* 2011;12:2825–30.

¹⁶ <https://github.com/jbussemaker/SBArchOpt>

¹⁷ <https://github.com/jbussemaker/SBArchOpt>

- [8] Lataniotis C, Marelli S, Sudret B. Uqlab 2.0 and uqcloud: open-source vs. cloud-based uncertainty quantification. In: SIAM conference on uncertainty quantification. 2022.
- [9] Faraci A, Beaupaire P, Gayton N. Review on Python toolboxes for Kriging surrogate modelling. In: ESREL. 2022.
- [10] Krügener M, Zapata Usandivaras J, Bauerheim M, Urbano A. Coaxial-injector surrogate modeling based on Reynolds-averaged Navier–Stokes simulations using deep learning. *J Propuls Power* 2022;38:783–98.
- [11] Ming D, Williamson D, Guillas S. Deep Gaussian process emulation using stochastic imputation. *Technometrics* 2022;1–12.
- [12] Eliáš J, Vořechovský M, Sadílek V. Periodic version of the minimax distance criterion for Monte Carlo integration. *Adv Eng Softw* 2020;149:102900.
- [13] Drouet V, Balesdent M, Brevault L, Dubreuil S, Morio J. Multi-fidelity algorithm for the sensitivity analysis of multidisciplinary problems. *J Mech Des* 2023;145:1–22.
- [14] Karban P, Pánek D, Orosz T, Petrášová I, Doležel I. FEM based robust design optimization with Agros and Ārtap. *Comput Math Appl* 2021;81:618–33.
- [15] Kudela J, Matousek R. Recent advances and applications of surrogate models for finite element method computations: a review. *Soft Comput* 2022;26:13709–33.
- [16] Chen Y, Dababneh F, Zhang B, Kassae S, Smith BT, Liu K, et al. Surrogate modeling for capacity planning of charging station equipped with photovoltaic panel and hydropneumatic energy storage. *J Energy Res Technol* 2020;142:050907.
- [17] Jasa J, Bortolotti P, Zalkind D, Barter G. Effectively using multifidelity optimization for wind turbine design. *Wind Energy Sci* 2022;7:991–1006.
- [18] Wang W, Tao G, Ke D, Luo J, Cui J. Transpiration cooling of high pressure turbine vane with optimized porosity distribution. *Appl Therm Eng* 2023;223:119831.
- [19] Savage T, Almeida-Trasvina HF, del Río-Chanona EA, Smith R, Zhang D. An adaptive data-driven modelling and optimization framework for complex chemical process design. *Comput Aided Chem Eng* 2020;48:73–8.
- [20] Chan A, Pires AF, Polacek T. Trying to elicit and assign goals to the right actors. In: Conceptual modeling: 41st international conference. 2022.
- [21] Hutter F, Osborne MA. A kernel for hierarchical parameter spaces. 2013, arXiv.
- [22] Bussemaker JH, Ciampa PD, Nagel B. System architecture design space exploration: An approach to modeling and optimization. In: AIAA aviation 2020 forum. 2020.
- [23] Fouda MEA, Adler EJ, Bussemaker J, Martins JRRA, Kurtulus DF, Boggero L, et al. Automated hybrid propulsion model construction for conceptual aircraft design and optimization. In: 33rd congress of the international council of the aeronautical sciences. 2022.
- [24] Bussemaker JH, Bartoli N, Lefebvre T, Ciampa PD, Nagel B. Effectiveness of surrogate-based optimization algorithms for system architecture optimization. In: AIAA aviation 2021 forum. 2021.
- [25] Balandat M, Karrer B, Jiang D, Daulton S, Letham B, Wilson A, et al. BoTorch: A framework for efficient Monte-Carlo Bayesian optimization. *Adv Neural Inf Process Syst* 2020;33:21524–38.
- [26] Adams B, Bohnhoff W, Dalbey K, Ebeida M, Eddy J, Eldred M, et al. Dakota, a multilevel parallel object-oriented framework for design optimization, parameter estimation, uncertainty quantification, and sensitivity analysis: Version 6.13 user's manual. Technical report, Albuquerque, NM (United States: Sandia National Lab.(SNL-NM)); 2020.
- [27] Roustant O, Ginsbourger D, Deville Y. DiceKriging, DiceOptim: Two R packages for the analysis of computer experiments by Kriging-based metamodeling and optimization. *J Stat Softw* 2012;51:1–55.
- [28] Zhang Y, Tao S, Chen W, Apley D. A latent variable approach to Gaussian process modeling with qualitative and quantitative factors. *Technometrics* 2020;62:291–302.
- [29] Chang TH, Wild SM. ParMOO: A Python library for parallel multiobjective simulation optimization. *J Open Source Softw* 2023;8:4468.
- [30] Garrido-Merchán EC, Hernández-Lobato D. Dealing with categorical and integer-valued variables in Bayesian optimization with Gaussian processes. *Neurocomputing* 2020;380:20–35.
- [31] Halstrup M. Black-box optimization of mixed discrete-continuous optimization problems (Ph.D. thesis), TU Dortmund; 2016.
- [32] Roustant O, Padonou E, Deville Y, Clément A, Perrin G, Giorla J, et al. Group kernels for gaussian process metamodels with categorical inputs. *SIAM J Uncertain Quant* 2020;8:775–806.
- [33] Zhou Q, Qian PZG, Zhou S. A simple approach to emulation for computer models with qualitative and quantitative factors. *Technometrics* 2011;53:266–73.
- [34] Saves P, Diouane Y, Bartoli N, Lefebvre T, Morlier J. A mixed-categorical correlation kernel for Gaussian process. *Neurocomputing* 2023;550:126472.
- [35] Pelamatti J, Brevault L, Balesdent M, Talbi E-G, Guerin Y. Efficient global optimization of constrained mixed variable problems. *J Global Optim* 2019;73:583–613.
- [36] Horn D, Stork J, Ier N-JS, Zaefferer M. Surrogates for hierarchical search spaces: The Wedge-Kernel and an automated analysis. In: Proceedings of the genetic and evolutionary computation conference. 2019.
- [37] Hung Y, Joseph VR, Melkote SN. Design and analysis of computer experiments with branching and nested factors. *Technometrics* 2009;51:354–65.
- [38] Audet C, Hallé-Hannan E, Le Digabel S. A general mathematical framework for constrained mixed-variable blackbox optimization problems with meta and categorical variables. *Oper Res Forum* 2023;4:1–37.
- [39] Saves P, Nguyen Van E, Bartoli N, Diouane Y, Lefebvre T, David C, Defoort S, Morlier J. Bayesian optimization for mixed variables using an adaptive dimension reduction process: applications to aircraft design. In: AIAA scitech 2022. 2022.
- [40] Conde Arenzana R, López-Lopera A, Mouton S, Bartoli N, Lefebvre T. Multi-fidelity Gaussian process model for CFD and wind tunnel data fusion. In: ECCOMAS aerobest. 2021.
- [41] Rufato RC, Diouane Y, Henry J, Ahlfeld R, Morlier J. A mixed-categorical data-driven approach for prediction and optimization of hybrid discontinuous composites performance. In: AIAA aviation 2022 forum. 2022.
- [42] Gorissen D, Crombecq K, Couckuyt I, Dhaene T, Demeester P. A surrogate modeling and adaptive sampling toolbox for computer based design. *J Mach Learn Res* 2010;11:2051–5.
- [43] Williams CK, Rasmussen CE. Gaussian processes for machine learning. MA: MIT press Cambridge; 2006.
- [44] Bouhlel MA, Bartoli N, Regis R, Otsmane A, Morlier J. Efficient Global Optimization for high-dimensional constrained problems by using the Kriging models combined with the Partial Least Squares method. *Eng Optim* 2018;50:2038–53.
- [45] Bouhlel MA, He S, Martins J. Scalable gradient-enhanced artificial neural networks for airfoil shape design in the subsonic and transonic regimes. *Struct Multidiscip Optim* 2020;61:1363–76.
- [46] Kwan LS, Pitrou A, Seibert S. Numba: A LLVM-based python JIT compiler. In: Proceedings of the second workshop on the LLVM compiler infrastructure in HPC. 2015.
- [47] Zaefferer M, Horn D. A first analysis of kernels for Kriging-based optimization in hierarchical search spaces. 2018, arXiv.
- [48] Jin R, Chen W, Sudjianto A. An efficient algorithm for constructing optimal design of computer experiments. *J Statist Plann Inference* 2005;2:545–54.
- [49] Garnett R, Osborne M, Hennig P. Active learning of linear embeddings for Gaussian processes. In: Uncertainty in artificial intelligence - Proceedings of the 30th conference. 2013.
- [50] Jones D. A taxonomy of global optimization methods based on response surfaces. *J Global Optim* 2001;21:345–83.
- [51] Lafage R. egobox, a Rust toolbox for efficient global optimization. *J Open Source Softw* 2022;7:4737.
- [52] Jones DR, Schonlau M, Welch WJ. Efficient global optimization of expensive black-box functions. *J Global Optim* 1998;13:455–92.
- [53] Deng X, Lin CD, Liu K, Rowe RK. Additive Gaussian process for computer models with qualitative and quantitative factors. *Technometrics* 2017;59:283–92.
- [54] Cuesta-Ramirez J, Le Riche R, Roustant O, Perrin G, Durantin C, Gliere A. A comparison of mixed-variables Bayesian optimization approaches. *Adv Model Simul Eng Sci* 2021;9:1–29.
- [55] Rebonato R, Jaecel P. The most general methodology to create a valid correlation matrix for risk management and option pricing purposes. *J Risk* 2001;2:17–27.
- [56] Rapisarda F, Brigo D, Mercurio F. Parameterizing correlations: a geometric interpretation. *IMA J Manag Math* 2007;18:55–73.
- [57] Bouhlel MA, Bartoli N, Regis R, Otsmane A, Morlier J. An improved approach for estimating the hyperparameters of the Kriging model for high-dimensional problems through the Partial Least Squares method. *Math Probl Eng* 2016;2016:6723410.
- [58] Cheng GH, Younis A, Hajikolaie KH, Wang GG. Trust region based mode pursuing sampling method for global optimization of high dimensional design problems. *J Mech Des* 2015;137:021407.
- [59] Karlsson R, Bliet L, Verwer S, de Weerd M. Continuous surrogate-based optimization algorithms are well-suited for expensive discrete problems. In: Artificial intelligence and machine learning. 2021.
- [60] Pelamatti J, Brevault L, Balesdent M, Talbi E-G, Guerin Y. Bayesian optimization of variable-size design space problems. *Opt Eng* 2021;22:387–447.
- [61] Hebbal A, Brevault L, Balesdent M, Talbi E-G, Melab N. Bayesian optimization using deep Gaussian processes with applications to aerospace system design. *Opt Eng* 2021;22:321–61.
- [62] Wildberger N. A rational approach to trigonometry. *Math Horiz* 2007;15:16–20.
- [63] Cho H, Kim Y, Lee E, Choi D, Lee Y, Rhee W. Basic enhancement strategies when using bayesian optimization for hyperparameter tuning of deep neural networks. *IEEE Access* 2020;8:52588–608.
- [64] Zuniga MM, Sinoquet D. Global optimization for mixed categorical-continuous variables based on Gaussian process models with a randomized categorical space exploration step. *INFOR Inf Syst Oper Res* 2020;58:310–41.
- [65] Lindauer M, Eggensperger K, Feurer M, AB, Deng D, Benjamins C, et al. SMAC3: A versatile Bayesian optimization package for hyperparameter optimization. *J Mach Learn Res* 2022;23:1–9.
- [66] Picheny V, Berkeley J, Moss H, Stojic H, Granta U, Ober S, et al. Trieste: Efficiently exploring the depths of black-box functions with TensorFlow. 2023, arXiv.
- [67] Cowen-Rivers AI, Ly W, Wang Z, Tutunov R, Jianye H, Wang J, et al. HEB0: Heteroscedastic evolutionary Bayesian optimisation. 2020, arXiv.
- [68] Jiang H, Shen Y, Li Y, Zhang W, Zhang C, Cui B. OpenBox: A Python toolkit for generalized black-box optimization. 2023, arXiv.

- [69] Kandasamy K, Vysyaraju KR, Neiswanger W, Paria B, Collins C, Schneider J, et al. Tuning hyperparameters without grad students: Scalable and robust bayesian optimisation with dragonfly. *J Mach Learn Res* 2020;21:3098–124.
- [70] Roy S, Crossley WA, Stanford BK, Moore KT, Gray JS. A mixed integer efficient global optimization algorithm with multiple infill strategy - Applied to a wing topology optimization problem. In: *AIAA scitech 2019 forum*. 2019.
- [71] Müller J, Shoemaker CA, Piché R. SO-MI: A surrogate model algorithm for computationally expensive nonlinear mixed-integer black-box global optimization problems. *Comput Oper Res* 2013;40:1383–400.
- [72] Tran T, Sinoquet D, Da Veiga S, Mongeau M. Derivative-free mixed binary necklace optimization for cyclic-symmetry optimal design problems. *Opt Eng* 2021.
- [73] Meliani M, Bartoli N, Lefebvre T, Bouhlef MA, Martins JRRA, Morlier J. Multi-fidelity efficient global optimization: Methodology and application to airfoil shape design. In: *AIAA aviation 2019 forum*. 2019.
- [74] Lee H. Gaussian processes. Springer Berlin Heidelberg; 2011, p. 575–7.
- [75] López-Lopera AF, Idier D, Rohmer J, Bachoc F. Multioutput Gaussian processes with functional data: A study on coastal flood hazard assessment. *Reliab Eng Syst Saf* 2022;218:108139.
- [76] Berthelin G, Dubreuil S, Salaün M, Bartoli N, Gogu C. Disciplinary proper orthogonal decomposition and interpolation for the resolution of parameterized multidisciplinary analysis. *Internat J Numer Methods Engrg* 2022;123:3594–626.
- [77] Cardoso I, Dubreuil S, Bartoli N, Gogu C, Salaün M, Lafage R. Disciplinary surrogates for gradient-based optimization of multidisciplinary systems. In: *ECCOMAS Aerobest*. 2023.
- [78] Platt J, Penny S, Smith T, Chen T, Abarbanel H. A systematic exploration of reservoir computing for forecasting complex spatiotemporal dynamics. *Neural Netw* 2022;153:530–52.
- [79] Charayron R, Lefebvre T, Bartoli N, Morlier J. Multi-fidelity Bayesian optimization strategy applied to overall drone design. In: *AIAA scitech 2023 forum*. 2023.
- [80] Charayron R, Lefebvre T, Bartoli N, Morlier J. Towards a multi-fidelity and multi-objective Bayesian optimization efficient algorithm. *Aerosp Sci Technol* 2023;142:108673.
- [81] Wold H. Soft modelling by latent variables: The non-linear iterative partial least squares (NIPALS) approach. *J Appl Probab* 1975;12:117–42.
- [82] Priem R, Diouane Y, Bartoli N, Dubreuil S, Saves P. High-dimensional efficient global optimization using both random and supervised embeddings. In: *AIAA aviation 2023 forum*. 2023.
- [83] Betz W, Papaioannou I, Straub D. Numerical methods for the discretization of random fields by means of the Karhunen–Loève expansion. *Comput Methods Appl Mech Engrg* 2014;271:109–29.
- [84] Menz M, Dubreuil S, Morio J, Gogu C, Bartoli N, Chiron M. Variance based sensitivity analysis for Monte Carlo and importance sampling reliability assessment with Gaussian processes. *Struct Saf* 2021;93:102116.
- [85] Ginsbourger D, Le Riche R, Carraro L. Kriging is well-suited to parallelize optimization. Springer Berlin Heidelberg; 2010, p. 131–62.
- [86] Roux E, Tillier Y, Kraria S, Bouchard P-O. An efficient parallel global optimization strategy based on Kriging properties suitable for material parameters identification. *Arch Mech Eng* 2020;67.