

# Adaptive sampling methodologies to guide the design of reactive materials towards user defined region of interest

Raphael Sala<sup>a</sup>, Yasser Sami<sup>a</sup>, Matthieu Jonckheere<sup>a</sup>, Alain Esteve<sup>a</sup>, Carole Rossi<sup>a,\*</sup>

<sup>a</sup>LAAS-CNRS, University of Toulouse, 31077 Toulouse, France

---

## Abstract

The discovery and optimization of materials still remains a significant challenge when dealing with a very large feature space, limited data, and, if the experiments and/or calculations are expensive to perform. This paper presents intelligent sampling methods designed to guide experiments or computations towards user-defined specific regions, termed "regions of interest," within vast and complex feature spaces. The focus of this work is to compare several adaptive sampling methodologies to identify 50 optimized Al/CuO thermite materials that meet user specifications, while minimizing the number of samples to reduce experimental costs. We considered Bayesian optimization and active learning techniques, both driven by specific learning schemes, to guide the sampling task. Particularly, we introduced two variations of the original ParEGO algorithm and evaluated their effectiveness in sampling optimized materials within the whole feature space against active learning methods. This work showed that, using a limited initial dataset of 100 points, the active learning approach is more effective to navigate in a vast design space as it leverages uncertainties and predictions from a surrogate model, combined with an acquisition function that prioritizes decision-making on unexplored data.

---

---

\*Corresponding author

Email address: [rossi@laas.fr](mailto:rossi@laas.fr) (Carole Rossi)

## 1. INTRODUCTION

Over the past 20 years, metal-based reactive materials have provided hope to power future generations of miniature autonomous systems and nanosatellites, and to provide high-energy actuations (thermal and mechanical) within a reduced volume ( $\text{cm}^{-3}$ ). They are energy-dense, benign, harmless to the environment yet feature extremely high combustion temperatures, thus providing actuations in all environments, *i.e.* even in deep-sea or in harsh-environment. They can therefore be employed for enabling safe, instantaneous shutdown of high-voltage electric currents, in anti-tamper devices, for maneuvers in-space or for welding repairs underground in geothermal or in disaster-stricken areas which is not possible with any other traditional CHNO energetic materials. One of the most interesting class of metal-based reactive material are thermites, also named metastable interstitial composites, composed of metallic particles (mostly Al) mixed with metal oxide particles [1–8]. They stay inert and stable until subjected to a sufficiently strong thermal stimulus, after which they undergo fast burning with release of high amount of chemical energy (up to  $16 \text{ kJ.cm}^{-3}$ ). Their burn rate and pressurization rate can be tuned in the ranges of  $10^{-1} - 10^4 \text{ m.s}^{-1}$  and  $1 - 400 \text{ MPa.}\mu\text{s}^{-1}$ , respectively, by adjusting the metal and oxide particle size, the metal purity, the powder density and the stoichiometry [2, 9–14]. Al is the fuel of choice due to the combination of its high energy density, reactivity, nontoxic nature, availability and low cost, but other fuels like B [15], Si [16], Mg [17], Zr [18, 19], Ti [20] are being explored in combination with Al for their high combustion energy or propensity to alloy exothermically with Al [21]. Copper oxide (CuO) molybdenum oxide ( $\text{MoO}_3$ ), tungsten oxide ( $\text{WO}_3$ ), iodine oxide ( $\text{I}_2\text{O}_5$ ), bismuth oxide ( $\text{Bi}_2\text{O}_3$ ) and iron oxide ( $\text{Fe}_2\text{O}_3$ ) [4, 22–24] are the many commonly used oxidizers. The size of the metal fuel and oxidizer particles can also be varied typically from 0.1 to  $10 \mu\text{m}$ .

The effective deployment of such promising energy-dense materials faces a significant hurdle arising from both the enormous material design space and lack of reliable design guidelines to experimental groups and engineers. Not only the nature of reactants (metal and oxidizer) but also the size of particles, the purity of the metallic fuel, the powder density, stoichiometric conditions do influence the bulk properties and thus the energetic performance.

Presently, to design a thermite for a particular application, one typically starts with an educated guess of the metal and metallic oxidizer composition, based on preliminary studies indicating that this materials choice could perform well. If the initial material composition shows promising results, materials are then synthesized varying the process parameters, or powder size and compaction, to improve performances further. But other groups may have different ideas and hence explore other parts of the design space. This approach is both imprecise and inefficient. A standard combustion test is a single-shot experiment that demands substantial material quantities (usually tens of milligrams) and require robust and costly safety measures.

In that context, the development of intelligent sequential experimental design capable of swiftly identifying optimal thermites *i.e.* which meet user-specifications, within limited experimental budgets is imperative. Numerous Bayesian optimization techniques [25, 26] has emerged as promising approaches for extracting trends and patterns from large design spaces and handling multiple objectives. However, traditional Bayesian optimization aims at maximizing or minimizing a given objective function  $f_{obj}$  with as few function evaluations as possible. Designing an optimal thermite for a given application requires exploring and identifying points within a user-defined region of the design space (referred to as the region of interest) in as few steps as possible. This is distinct from multi-objective optimization problems, as in this case, we focus on efficiently exploring a specified interval of the  $f_{obj}$ , denoted as  $K$ . Active learning technique [27–32] is an alternative intelligent sampling strategy to guide the exploration towards specified interval of the  $f_{obj}$ . It focuses on improving a predictive model, typically a Gaussian Process Regression (GPR), by selectively choosing the most informative data points for evaluation.

Both Bayesian optimization and active learning methodologies have seen an exponential growth in popularity in the past decades. Similarities and differences between the vast families of adaptive sampling, active learning, and Bayesian optimization were reported in [33]. In this study, the recently developed adaptive machine learning workflow [34] is compared with a modified ParEGO scheme to demonstrate the efficacy of Bayesian methods in rationally guiding physics-based computations in search of optimal thermites with desired functional properties.

## 2. METHODOLOGY

### 2.1. Active learning approach : GPR with a Custom Acquisition Function

Firstly a simple GPR is trained to predict both the value and the uncertainty of pressure, temperature at any point within the design space. Subsequently, an acquisition function tailored to guide the sampling towards a region  $K$  avoids unnecessary evaluations in other non interesting areas of the design space. Let  $n$  be the number of inputs and  $m$  the number of outputs (in this paper,  $n \in [1, 3]$  and  $m = 2$ ). We consider values returned by the  $GPR : \mathbb{R}^n \mapsto \mathbb{R}^m$ , to evaluate quality of sampled points based on their variance and an "interest" score. Interest means that the point falls within  $K$  region of the objective function. An interest function is therefore defined as  $I : x \mapsto \mathbb{P}(GPR(x) \in K)$ . This methodology is referred to as **irbs** for *Interest Region Bayesian Sampling*.

It has to be noted that,  $I$  directs sampling towards interesting regions, which may result in exploitation without sufficient exploration: sampled points may be too close to each other, providing limited information about the behavior of  $f_{obj}$  in  $K$ . With the goal to sample new points far enough from already chosen points, we added an additional terms designed to encourage exploration. The choice of a new point  $x$  is done according to its variance,  $\mathbb{V}(GP(x))$ , its interest,  $I(x)$  and this additional term,  $C(x)$ , defined as follows :

$$C_{s,d,A}(u) = 1 - \frac{1}{1 + \exp(A - s\Gamma_d(u))}$$

$$\Gamma_d(u) = \sum_{x \in X_k} \exp(-d\|u - x\|^2)$$

Here,  $x \in X_k$  represents the points that have already been explored, and the parameters  $d$ ,  $s$ , and  $A$  are defined in [34]. Based on numerical experiments, these parameters are set to  $A = 5$ ,  $s = 1$ , and  $d = 0.04$ . The function  $C$  is designed to penalize points that are close to those already sampled, and this approach is referred to as **irbs\_C**.

The algorithm for **irbs\_C** is given below. The algorithm for **irbs** is identical, except that lines 9 is omitted, and the last term in line 10 ( $C(x)$ ) used to compute  $Acq(x)$  is removed.

---

#### Algorithm 1 irbs\_C: Interest Region Bayesian Sampling + Coverage

---

**Require:**  $N_{init}, N_{iter}, \theta$

- 1: Initialize  $X_{init}$  using Latin Hypercube Sampling
- 2: Initialize  $D_{N_{init}} = (X_{init}, y_{init})$
- 3: **for**  $n = N_{init}$  to  $N_{init} + N_{iter} - 1$  **do**
- 4:   Train GP model  $\mathcal{G} \mathcal{P}$  on  $D_n$
- 5:   Initialize an empty list  $Acq\_values = []$
- 6:   **for**  $x \in D_n[0]$  **do** ▷ Get feature values of each sample
- 7:     Compute  $V(x) = \mathbb{V}(GP(x))$
- 8:     Compute  $I(x) = \mathbb{P}(GP(x) \in K)$
- 9:     Compute  $C(x) = 1 - \frac{1}{1 + \exp(A - s\Gamma_d(x))}$
- 10:    Compute  $Acq(x) = V(x) + I(x) + C(x)$
- 11:    Append  $-Acq(x)$  to  $Acq\_values$  ▷ "-" because SHGO minimizes the function
- 12:   **end for**
- 13:   Optimize  $Acq\_opti = \text{shgo}(D_n[0], Acq\_values)$
- 14:   Compute  $x_{n+1} = \arg \max_x Acq\_opti(x)$
- 15:   Evaluate  $y_{n+1} = f_{obj}(x_{n+1})$
- 16:   Update dataset  $D_{n+1} = D_n \cup (x_{n+1}, y_{n+1})$
- 17: **end for**
- 18: **return**  $D_{N_{init} + N_{iter}}$

---

$N_{iters}$  is the number of iteration.  $D_n$  is the dataset containing  $n$  points evaluated by  $f_{obj}$ .

## 2.2. Bayesian approach : ParEGO with a Custom Objective Function

The Efficient Global Optimization (EGO) algorithm was designed for optimizing expensive black-box functions. EGO also uses a GPR to model the objective function and selects new points to evaluate by maximizing the expected improvement (EI). The expected improvement balances exploration and exploitation by considering both the predicted value and the uncertainty of the prediction [35]. ParEGO is an extension of EGO to multi-objective optimization. Instead of optimizing a single objective, ParEGO converts multiple objectives into a single scalar objective using a parameterized scalarizing weight vector. This scalarization step allows the GPR to handle multiple objectives by creating a single aggregated objective function. At each iteration, a weight vector  $\lambda$  is drawn uniformly at random from the set

$$\Lambda = \left\{ \lambda = (\lambda_1, \lambda_2, \dots, \lambda_m) \mid \sum_{j=1}^m \lambda_j = 1 \wedge \forall j, \lambda_j = \frac{l}{s}, l \in \{0, \dots, s\} \right\}$$

where  $m$  is the number of objectives (outputs). The parameter  $s$  is a hyper-parameter that controls the size of  $\Lambda$ . A smaller, compared to the number of algorithm's iteration,  $s$  increases the probability of obtaining extreme weight vectors, while a larger  $s$  allows for more balanced weight vectors. We use  $s = 1000$ . The scalar objective function  $f_\lambda(x)$  is then computed using the augmented Tchebycheff function:

$$f_\lambda(x) = \max_{i \in \llbracket 1, m \rrbracket} \{\lambda_i f_i(x)\} + \rho \sum_{i=1}^m \lambda_i f_i(x)$$

where  $f_i(x)$  are the individual objective functions, and  $\rho$  is a small positive value, the paper [36] suggest  $\rho = 0.05$ . The term with  $\rho$  is added because the nonlinearity of the function allows points in nonconvex regions of the Pareto front to become minimizers, leading to the discovery of nonsupported solutions. Meanwhile, the linear component of the function ensures that solutions weakly dominated by Pareto optimal solutions are rewarded less, favoring true Pareto optimal outcomes[36].

The next point to sample is therefore chosen by maximizing the expected improvement with respect to the scalar objective function  $f_\lambda$  using a genetic algorithm (GA). This involves initializing a temporary population of solution vectors, some of which are mutants of previously evaluated points while others are generated randomly. The genetic algorithm iteratively evaluates the expected improvement of these solutions using the Gaussian Process model, on  $f_\lambda$ , and applies selection, recombination, and mutation operations to form a new population. This process continues until a predefined number of evaluations is reached, ultimately returning the solution that maximizes the expected improvement after a large number of population's generation [37].

### 2.2.1. Modification of the Scalar Objective Function

ParEGO uses a parameterized scalarizing weight vector to explore the search space by considering different objectives to varying extents during each iteration. This approach typically maximizes the objectives using the function  $\max_{i \in \{1, \dots, m\}} \{\lambda_i f_i(x)\}$ . However, since we are interested in exploring the region  $K$ , we need to transform the values given by  $f_i$ . We define functions,  $\forall i \in \llbracket 1, m \rrbracket, g_i : \mathbb{R}^m \rightarrow \mathbb{R}$ , that maps  $x$  to a scalar value indicating the distance with the region  $K = K_1 \times \dots \times K_m$ , as,

$$g_i(x) = 1 - \text{dist}(f_i(x), K_i) \cdot \text{slope}$$

*slope* is a parameter set at 10 for this study in order to accentuate the importance to fall into the region  $K$ . The expression of  $g$  resembles a tent, which is reflected in the algorithm's name. The hyper-parameters for this algorithm include  $s$  and *slope*. Additionally, there are parameters inherited from the genetic algorithm [37], with the only one referenced later being *num\_gen*, which determines the number of generations to iterate and converge to an optimal population of points with outputs falling with  $K$ . The method is named **sGA\_tent**, where 's' stands for *scalarization*, reflecting that our optimization problem doesn't require a Pareto front. Below is the algorithm for this method.

---

**Algorithm 2** sGA\_tent

---

**Require:**  $N_{init}, N_{iter}, \theta$ 

```
1: Initialize  $X_{init}$  using Latin Hypercube Sampling
2: Initialize  $D_{N_{init}} = (X_{init}, y_{init})$ 
3: for  $n = N_{init}$  to  $N_{init} + N_{iter} - 1$  do
4:   Choose a random  $\lambda \in \Lambda$ 
5:   Initialize an empty list  $y_\lambda = []$ 
6:   for  $y \in D_n[1]$  do ▷ Get target values of each sample
7:     Append  $\max_{i \in [1, m]} \{\lambda_i g_i(y)\} + \rho \sum_{i=1}^m \lambda_i g_i(y)$  to  $y_\lambda$ 
8:   end for
9:   Train GP model  $\mathcal{GP}$  on  $(D_n[0], y_\lambda)$ 
10:  Select a population of points,  $X_{pop}$  from  $D_n[0]$  with some mutants
11:  for  $i=1$  to  $\text{num\_gen}$  do
12:    Compute score =  $EI \circ \mathcal{GP}(X_{pop})$ 
13:    Reproductive selection on  $X_{pop}$  : Replacement, crossover, mutation according to score
14:  end for
15:  Compute  $x_{n+1} = \arg \max_{x \in X_{pop}} EI \circ \mathcal{GP}(x)$ 
16:  Evaluate  $y_{n+1} = f_{obj}(x_{n+1})$ 
17:  Update dataset  $D_{n+1} = D_n \cup (x_{n+1}, y_{n+1})$ 
18: end for
19: return  $D_{N_{init}+N_{iter}}$ 
```

---

### 2.2.2. Replacing SHGO by a Genetic Algorithm

We also explored the replacement of the Simplicial Homology Global Optimization (SHGO) algorithm by a genetic algorithm. This enables to focus solely on the interest function,  $I$ . The issues of point proximity are mitigated because mutations and crossovers in the genetic algorithm facilitate continuous and natural exploration, rather than just exploitation. So we replace  $f_\lambda$  by the irbs scheme. This approach allows us to directly maximize the probability of being in the region of interest without unnecessary complexities. The algorithm is given below :

---

**Algorithm 3** GA\_GP

---

**Require:**  $N_{init}, N_{iter}, \theta$ 

```
1: Initialize  $X_{init}$  using Latin Hypercube Sampling
2: Initialize  $D_{N_{init}} = (X_{init}, y_{init})$ 
3: for  $n = N_{init}$  to  $N_{init} + N_{iter} - 1$  do
4:   Train GP model  $\mathcal{GP}$  on  $D_n$ 
5:   Initialize an empty list  $Itr\_values = []$ 
6:   for  $x \in D_n[0]$  do ▷ Get feature values of each sample
7:     Compute  $I(x) = \mathbb{P}(GP(x) \in K)$ 
8:     Append  $I(x)$  to  $Itr\_values$ 
9:   end for
10:  Select a population of points,  $X_{pop}$  from  $D_n[0]$  with some mutants
11:  for  $i=1$  to  $\text{num\_gen}$  do
12:    Compute  $\text{improvement\_itr} = EI \circ I(X_{pop})$ 
13:    Reproductive selection on  $X_{pop}$  : Replacement, crossover, mutation acc.  $\text{improvement\_itr}$ 
14:  end for
15:  Compute  $x_{n+1} = \arg \max_{x \in X_{pop}} EI \circ I(x)$ 
16:  Evaluate  $y_{n+1} = f_{obj}(x_{n+1})$ 
17:  Update dataset  $D_{n+1} = D_n \cup (x_{n+1}, y_{n+1})$ 
18: end for
19: return  $D_{N_{init}+N_{iter}}$ 
```

---

### 2.3. Computational details and Metrics

To optimize simulation time, we parallelized the algorithms for evaluating  $f_{obj}$ , enabling us to obtain multiple new input points at each iteration rather than just one. For this particular study, we consider a batch size of 8 points.

To compare the performance of the four previously presented methods (irbs, irbs\_C, sGA\_tent, and GA\_GP), we sampled points into the design space (feature space) until we obtained 50 points of interest (i.e., points within the interest region  $K$ ). We used three metrics to evaluate the performances of the methods: **Efficiency**, **Coverage**, and **Voronoi volume**, with calculations performed at the end of each experiment i.e. after the 50 new interesting points are sampled.

- **Efficiency** (in %) is defined as the ratio of points of interest (sampled within  $K$ ) to the total number of sampled points. A higher **Efficiency** indicates a more effective method.
- **Coverage** is calculated as ratio of the the total volume covered by hyper-spheres with a predefined radius (set to 0.025), centered at each sampled point of interest, divided by the total number of points of interest (sampled within  $K$ ). We expressed it as a percentage of the total design space volume by dividing it by the volume of a hyper-sphere corresponding to the dimensionality of the design space. A higher coverage indicates that the sampled points of interest are more dispersed.

Calculating the volume of hyper-spheres in  $\mathbb{R}^n$ , i.e. where hyper-sphere clusters exist, is not straightforward. For clusters, consisting of one or two hyper-spheres which do not intersect the design space boundary limits, the volume is computed analytically. Hyper-spheres that intersect the design space boundaries i.e. considering normalized data, that corresponds to points sampled outside  $[0, 1]^n$  are ignored. For clusters with more than two intersecting hyper-spheres, we use the octree technique. This involves creating a parallelised around the clusters and subdividing it into smaller equal parts. We sum the volumes of the subdivisions that are entirely within the clusters and ignore those outside. For subdivisions intersecting the clusters, we recursively subdivide them further. This process is repeated to obtain a precise volume estimate with narrow boundaries.

In addition, for each algorithm, we compute the :

- **Voronoi volume** which is the volume of the region surrounding each sampled point. It offers more detailed insights than *coverage* by revealing the distribution and extent of the areas controlled by each point. This allows for a deeper understanding of the spatial arrangement and influence of the sampled points.

To calculate the Voronoi volume, we first construct the Voronoi diagram by partitioning the design space according to the distance from each sampled point. For points whose Voronoi cells do not intersect the design space boundaries, the volume is computed analytically based on the geometry of the convex polytope that forms the Voronoi cell. For cells intersecting the boundaries, we exclude the portion of the cell that extends beyond the normalized space  $[0, 1]^n$  from the volume calculation.

## 3. RESULTS

The proposed intelligent sampling strategies are analyzed and discussed through the discovery of Al/CuO thermite reactive material for two applications: thermal plugging [38] and high-actuation such as pyro-fuzing [39, 40]. For welding applications, the motivation is to design Al/CuO thermites that generate combustion temperatures and pressures in the range 2000 - 4500 K and 1 - 2 MPa, respectively. These specified Temperature-Pressure values define the region of interest 1 of the property space (referred to as R1) for the exploration. The second application requires very high pressure and temperature, i.e., in the range of 40 - 56 MPa and 5800 - 7000 K, respectively. These specified Temperature-Pressure values define the region of interest 2 of the property space (referred to as R2).

### 3.1. Feature and Property spaces of Al/CuO

A reactive thermite material composed of an Al/CuO mixture exhibits five key characteristics: radii of CuO and Al particles,  $r_{CuO}, r_{Al}$ , the Al over CuO molar ratio,  $\phi$ , Al purity,  $Y_{Al}$ , and compaction,  $\alpha_p$ . The range of input features

values are given in Table 1 corresponding to technological constraints. The main properties are the combustion pressure and temperature, which are expected to fall within  $[2 \times 10^5; 1.5 \times 10^8]$  Pa and  $[600; 7500]$  K, respectively. The thermodynamic model used to predict these combustion properties is detailed in [41–44].

We will conduct two different sampling experiments: the first involves two unknown features ( $r_{Al}$  and  $\phi$ ) considering that the size of oxidizer is not important as it releases its oxygen rapidly upon heating. In a second experiment, we have added  $r_{CuO}$  as unknown feature to validate the previous assumption. In this second experiment, the unknown features are :  $r_{Al}$ ,  $\phi$  and  $r_{CuO}$ .

Table 1: Description of the feature and property space considered in this work with their range of variation

Feature	Description	Range
$r_{Al}$	Aluminum particle radius	0.1 - 10 $\mu\text{m}$
$r_{CuO}$	Copper oxide particle radius	0.1 - 10 $\mu\text{m}$
$\phi$	Aluminum richness	1 - 4
$\alpha_p$	Powder density	0.2 - 0.5
$Y_{Al}$	Al purity	0.5 - 0.9
Property	Description	Range
$T_{\max}$	Combustion temperature	600 - 7500 K
$P_f$	Combustion pressure	0.2 - 150 MPa

Each experiment starts with the same 100 initial points to train the GPR and stops after finding **50 new points of interest**. This stopping criterion simulates a real-world scenario where data acquisition must succeed within operational limits. Furthermore, it facilitates the interpretation of the **Coverage** and **Voronoi volume** metrics, as the number of points of interest is constant for all the methods.

## 3.2. Sampling in 2D Design Space

### 3.2.1. Identifying Al/CuO Thermite Materials with Properties Within the R1 Region

In this section, we are searching for Al/CuO materials, composed of Al and CuO particles, that generate combustion temperatures and pressures (the targets) in the range of 2000 - 4500 K and 1 - 2 MPa, respectively. Two features are variable,  $r_{Al}$  and  $\phi$ , as we consider that  $r_{Al} = r_{CuO}$ . The quality of Al powder is fixed by the supplier (Al purity is of 0.8) and the powder density ( $\alpha_p$ ) is set at the maximum possible experimentally, 0.5.

We ran the four algorithms and plotted the spatial distribution of sampled points for each, as shown in Figure 1. To find 50 points of interest ( $n_{itr}=54^1$ ), irbs explored 247 points and irbs\_c explored 417, whereas GA\_GP and sGA\_tent explored only 103 and 151, respectively. Therefore, the modified EGO-based algorithms are significantly more efficient. Looking now at the distribution of points in the 2D  $r_{Al}$  -  $\phi$  design space, the GA\_GP algorithm concentrate sampling in the same region : high  $\phi$  (3-4) and low  $r_{Al}$  (below 2  $\mu\text{m}$ ). Points sampled by sGA\_tent are much less concentrated. We also observe in Figure 1 that one area of interest is rapidly identified (yellow points) but the sampling continues very locally. In contrast, the irbs\_c algorithm demonstrates a stronger exploratory capability, as the initial yellow points of interest are in different regions of the space compared to the later ones (red). Next, we present Violin and Kernel Density Estimate (KDE) plots (Figures 2 and 3) detailing the distribution of the target values (combustion pressure and temperature) of the Al/CuO thermites contained in each database (100 training points + points sampled by each strategy to get 50 points of interest).

Violin plots confirm the tendencies described above. The points in the dataset are significantly less dispersed when sampled with sGA\_tent and GA\_GP, as the median point (white dot) is closer to the region of interest delineated by red vertical lines. Moreover, the plots clearly show that irbs and irbs\_c allow for less localized sampling, as evidenced by a slight bump between 20 - 40 MPa (Figure 2a). Interestingly, considering the combustion temperature,

<sup>1</sup> since the initial 100 points used to train the GPR already included 4 points of interest

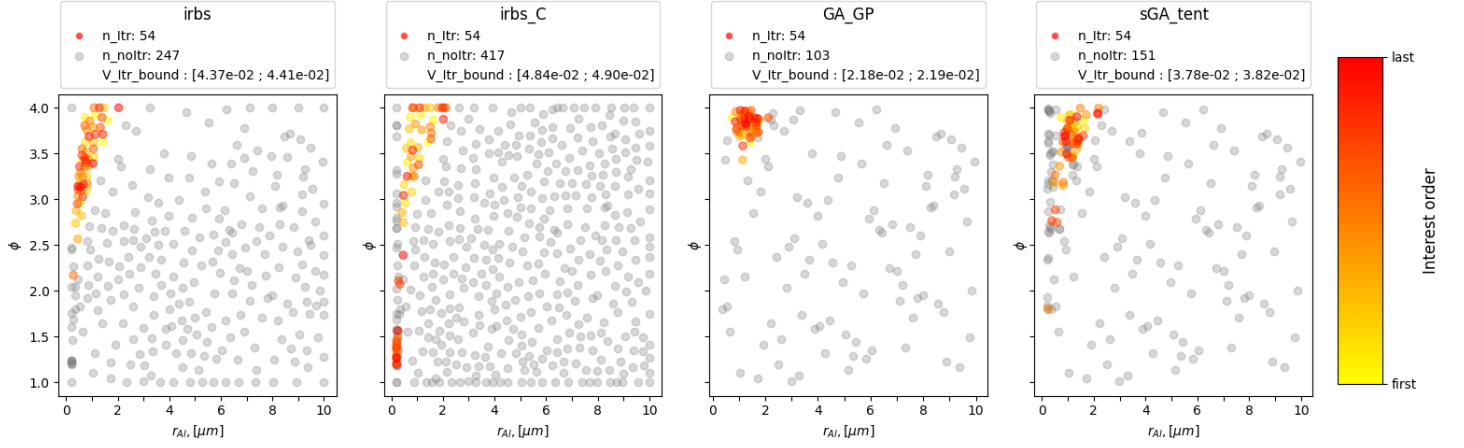


Figure 1: Spatial distribution of sampled points. The points of interest are represented as yellow-to-red balls, indicating their sampling order, while gray balls denote non-interesting points. The sampling stopped after finding 50 points in the region R1.

GA\_tent outperforms the three other algorithms, showing a good spread of points within the region of interest, which is beneficial for users. These observations are confirmed in KDE plots (Figure 3). irbs\_C and sGA\_tent sampled 50 points of interest that resulted in a wider range of combustion pressures and temperatures.

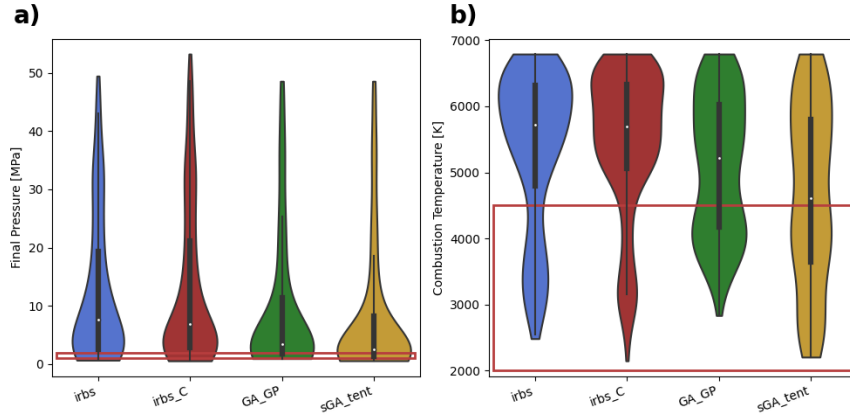


Figure 2: Violin plots of the targets: a) Final pressure, b) Combustion temperature obtained with each algorithm. The red lines mark the limits of the region of interest.



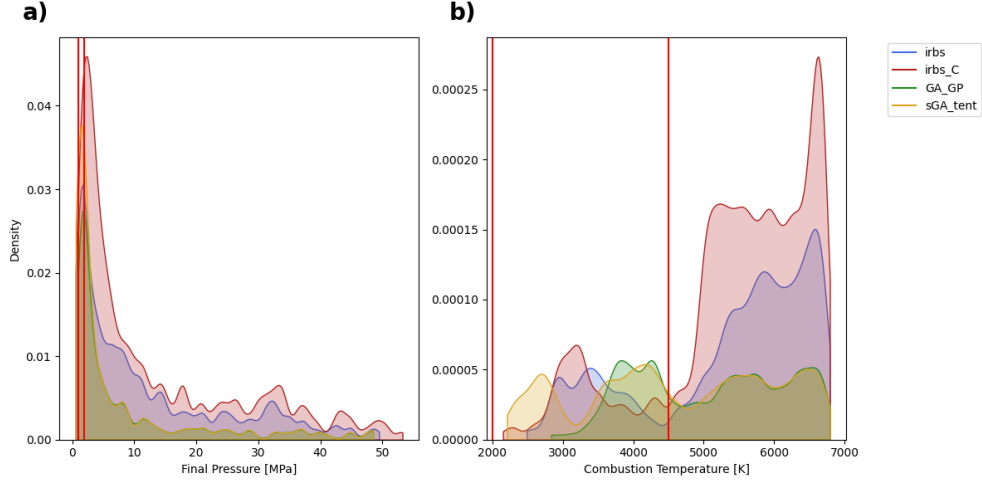


Figure 3: KDE plots of the targets: a) Final pressure, b) Combustion temperature obtained with each algorithm. The red lines mark the limits of the region of interest.

### 3.2.2. Identifying Al/CuO Thermite Materials with Properties Within the R2 Region

In this section, we are now searching for Al/CuO materials that generate combustion pressures and temperatures in the range of 40 - 56 MPa and 5800 - 7000 K, respectively. Still, only two features are variable,  $r_{Al}$  and  $\phi$ . Figure 4 shows the spatial distribution of sampled points for each of the four algorithms, with the points of interest highlighted in color. Herein  $n_{itr}=56$  since the initial 100 points used to train the GPR include 6 points of interest. As in the exploration of the 2D design space to find thermites within the region of interest R1, the EGO-based algorithms proved to be more efficient: the 50 new points were found in 95 iterations for GA\_GP and 145 iterations for sGA\_tent, compared to 178 and 263 for irbs and irbs\_C, respectively. GA\_GP also confirms its tendency to find points very locally, while the other three methods allow for a good dispersion of the points of interest. This is evident in the following Violin and KDE plots (Figure 5 and 6).

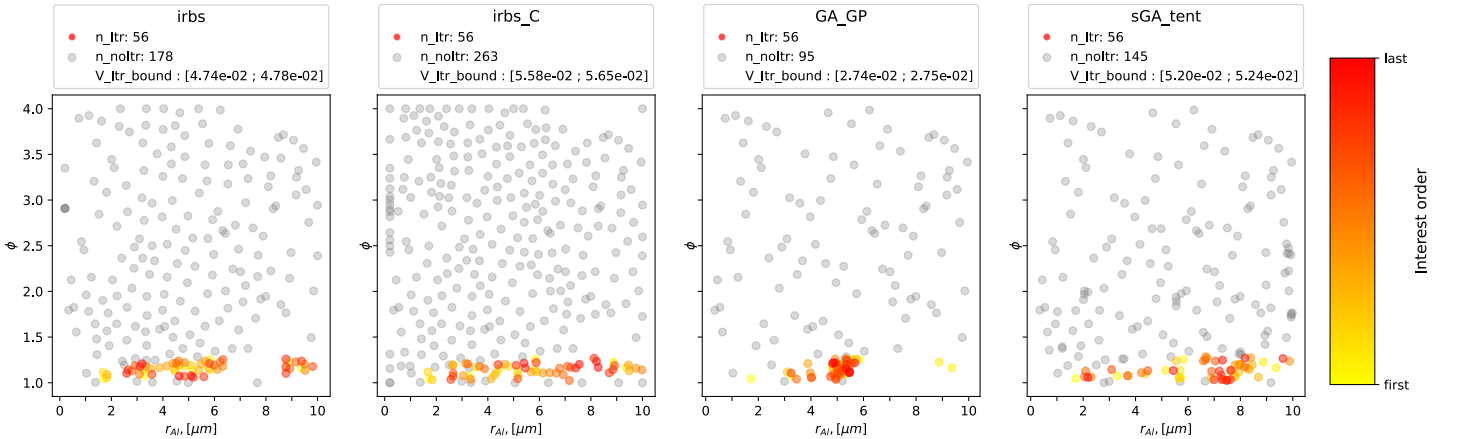


Figure 4: Spatial distribution of sampled points. The points of interest are represented as yellow-to-red balls, indicating their sampling order, while gray balls denote non-interesting points. The sampling stopped after finding 50 points in the region of interest.

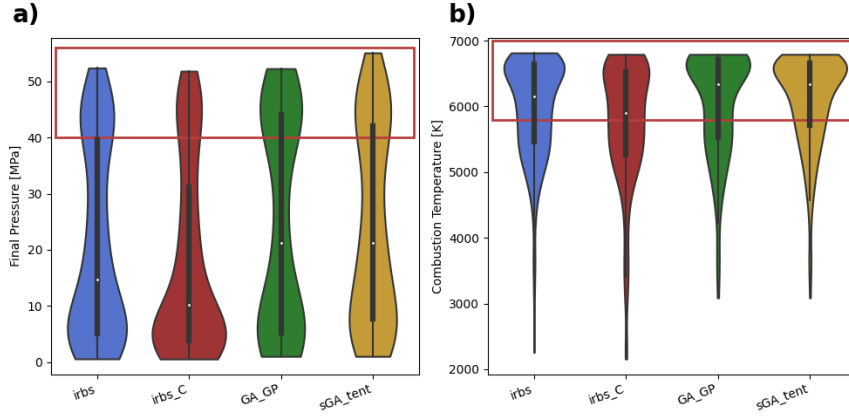


Figure 5: Violin plots of the targets: a) Final pressure, b) Combustion temperature obtained with each algorithm. The red lines mark the limits of the region of interest.

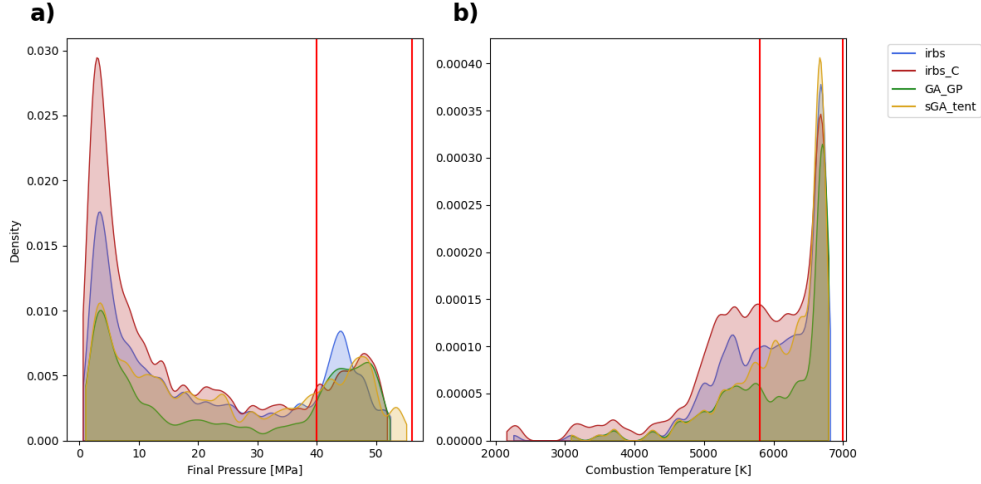


Figure 6: KDE plots of the targets: a) Final pressure, b) Combustion temperature obtained with each active sampling scenario. The red lines mark the limits of the region of interest.

### 3.2.3. Comparison of the methods

The Table 2 reports the two metrics, **Coverage** and **Efficiency**, calculated after the search of 50 points of interest in the regions of interest R1 and R2 using each of the four algorithms. Clearly, with an **Efficiency** > 87%, the most effective method is the GA\_GP algorithm. It quickly identifies points of interest as the genetic algorithm (GA) consistently converges around previously identified interest points, which can be an advantage for designing materials with limited iterations. The less effective is irbs\_C due to the dispersive term  $C(x)$  in the acquisition function. If the user seeks 50 thermite configurations that are well-dispersed throughout the feature space to make a selection based on technological constraints, irbs\_C yields the best results, achieving the higher coverage. It is noteworthy that sGA\_tent offers a good compromise between efficiency and spatial distribution. sGA\_tent is low impacted by the exploratory limitations of the GA because it does not optimize the same function each time, thereby reducing the likelihood of sampling closely spaced points.

Table 2: Comparison of the algorithms in terms of **Coverage** and **Efficiency**, as defined in Section 2.3 and considering the two regions of interest (R1 and R2).

Region of Interest	Algorithm	Coverage (in %)	Efficiency (in %)
R1	irbs	$41.4 \pm 0.2$	24.88
	irbs_C	<b><math>45.9 \pm 0.3</math></b>	13.48
	GA_GP	$20.6 \pm 0.1$	<b>87.72</b>
	sGA_tent	$35.8 \pm 0.2$	47.62
R2	irbs	$43.3 \pm 0.2$	37.31
	irbs_C	<b><math>51.1 \pm 0.3</math></b>	22.83
	GA_GP	$25.0 \pm 0.1$	<b>98.04</b>
	sGA_tent	$47.5 \pm 0.2$	49.50

As last metrics, Figures 7 and 8 give the **Voronoi volume** distribution that represents the spatial influence of each sampled interesting point. The bar plot is sorted in increasing order, highlighting the extent of zones controlled by the points. The first subplot considers only the feature space, capturing the distribution of interest points, while the second subplot includes the target space to reveal fluctuations in the objective function. **Voronoi volume** metric confirms that irbs\_C effectively samples points that are well-dispersed in the design space, as indicated by the high median and first quartile range (QR) values. This results in fewer small regions with closely spaced points. The spatial distribution of the points of interest sampled using irbs and sGA\_tent appears to be comparable. However, the 1QR and lower whisker of irbs\_C are consistently larger than those of the other methods, while the third QR and upper whisker are smaller. This indicates that irbs\_C provides a more uniform coverage of the space compared to sGA\_tent. Finally, the Voronoi diagram in Figure 8 reveals that irbs\_C captures greater variability in  $f_{obj}$  due to the larger volume of sampled points, which is valuable in practice as it provides more comprehensive information for users.

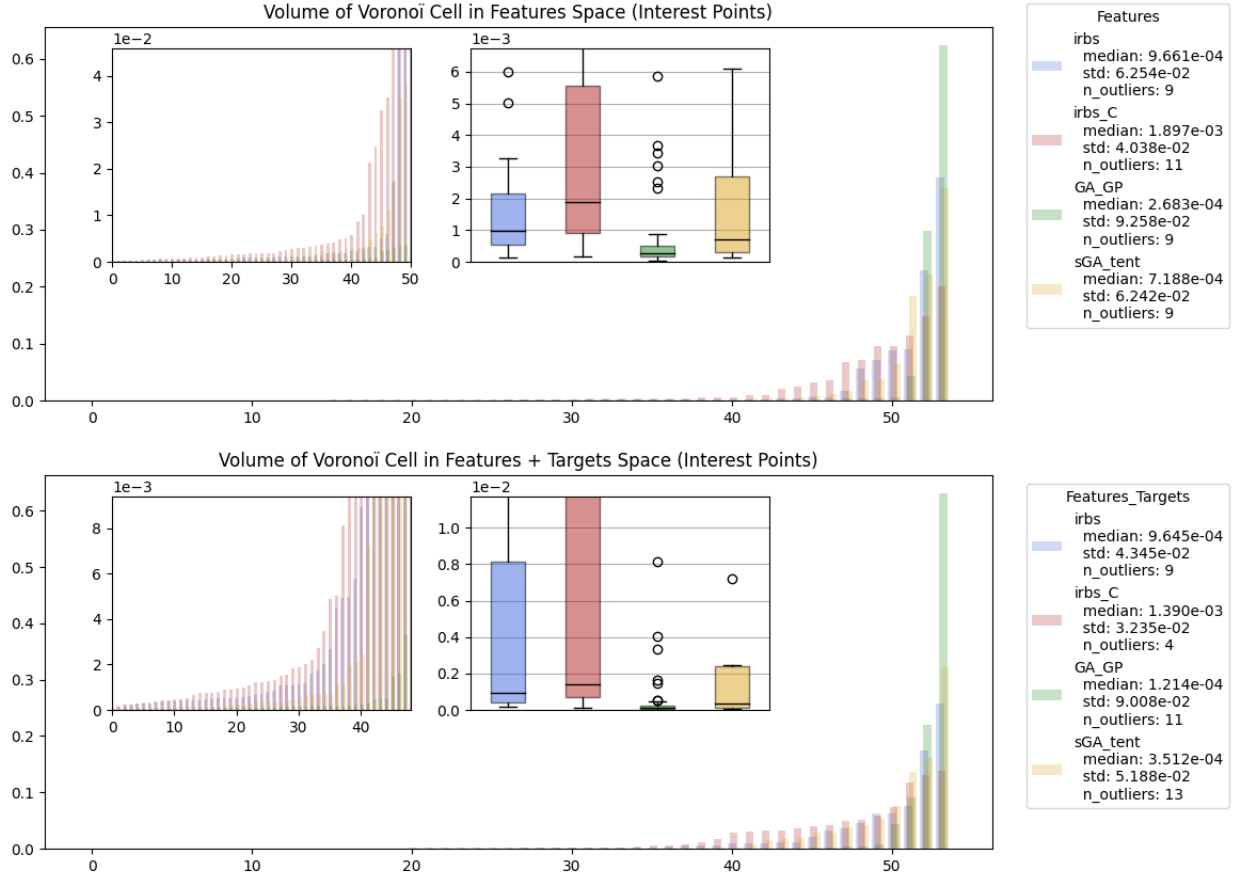


Figure 7: Voronoi volume distribution within the unit hypercube illustrates the spatial influence of each sampled point in region R1. The bar plot, sorted in ascending order, highlights the extent of the regions controlled by each point. The first subplot focuses solely on the feature space, showcasing the distribution of interest points. The second subplot is currently empty due to computation errors encountered during analysis.

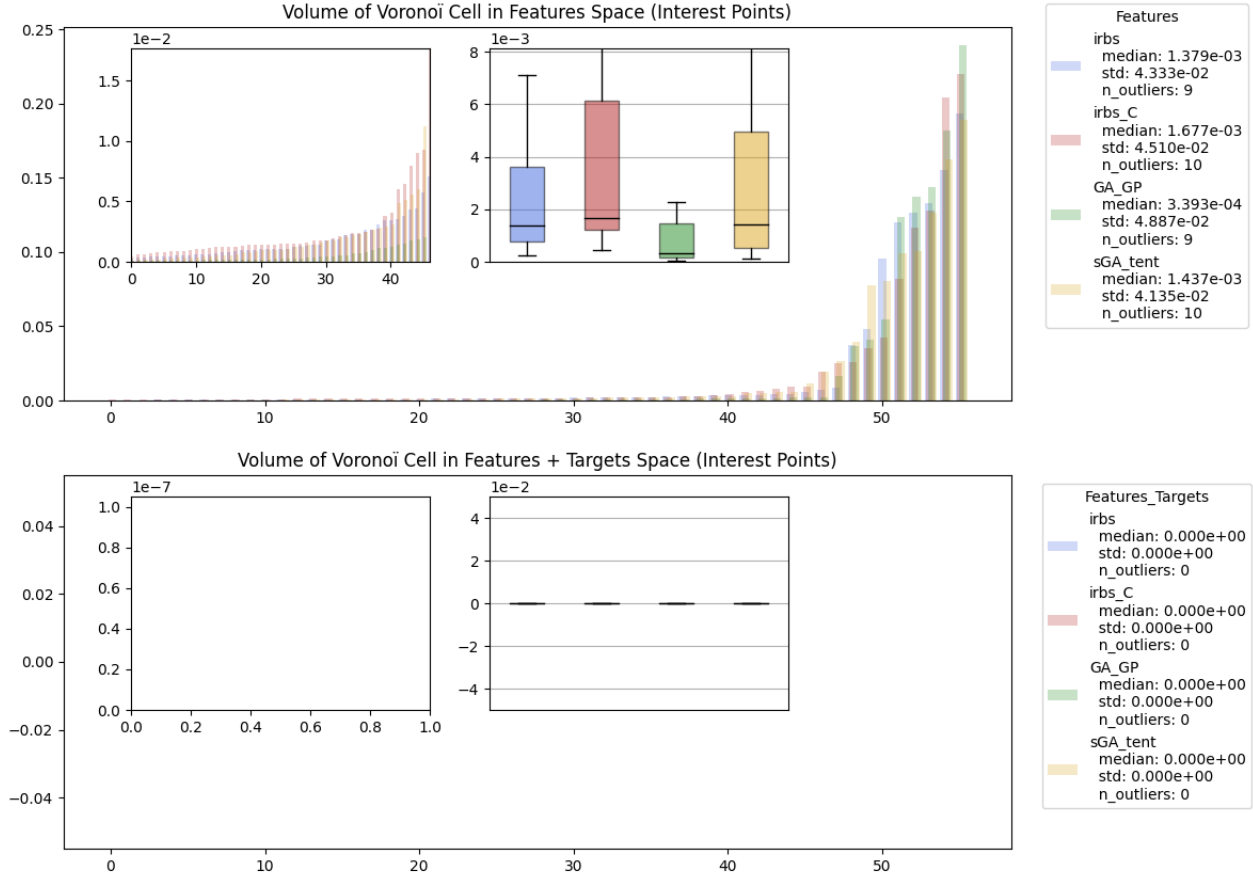


Figure 8: Voronoi volume distribution within the unit hypercube illustrates the spatial influence of each sampled point in region R2. The bar plot, sorted in ascending order, highlights the extent of the regions controlled by each point. The first subplot focuses solely on the feature space, showcasing the distribution of interest points. The second subplot is currently empty due to computation errors encountered during analysis.

### 3.3. Sampling in a 3D Design Space

#### 3.3.1. Identifying Al/CuO Thermite Materials with Properties Within the R1 Region

Now we will consider searching Al/CuO materials that generate combustion temperatures and pressures in the range of 2000 - 4500 K and 1 - 2 MPa, respectively (R1) but considering three variable features :  $r_{Al}$ ,  $\phi$  and  $r_{CuO}$ . The quality of Al powder is fixed by the supplier (Al purity is of 0.8) and the powder density ( $\alpha_p$ ) is set at the maximum possible experimentally, 0.5. We ran the four algorithms and plotted the spatial distribution of sampled points for each, as shown in Figure 9.

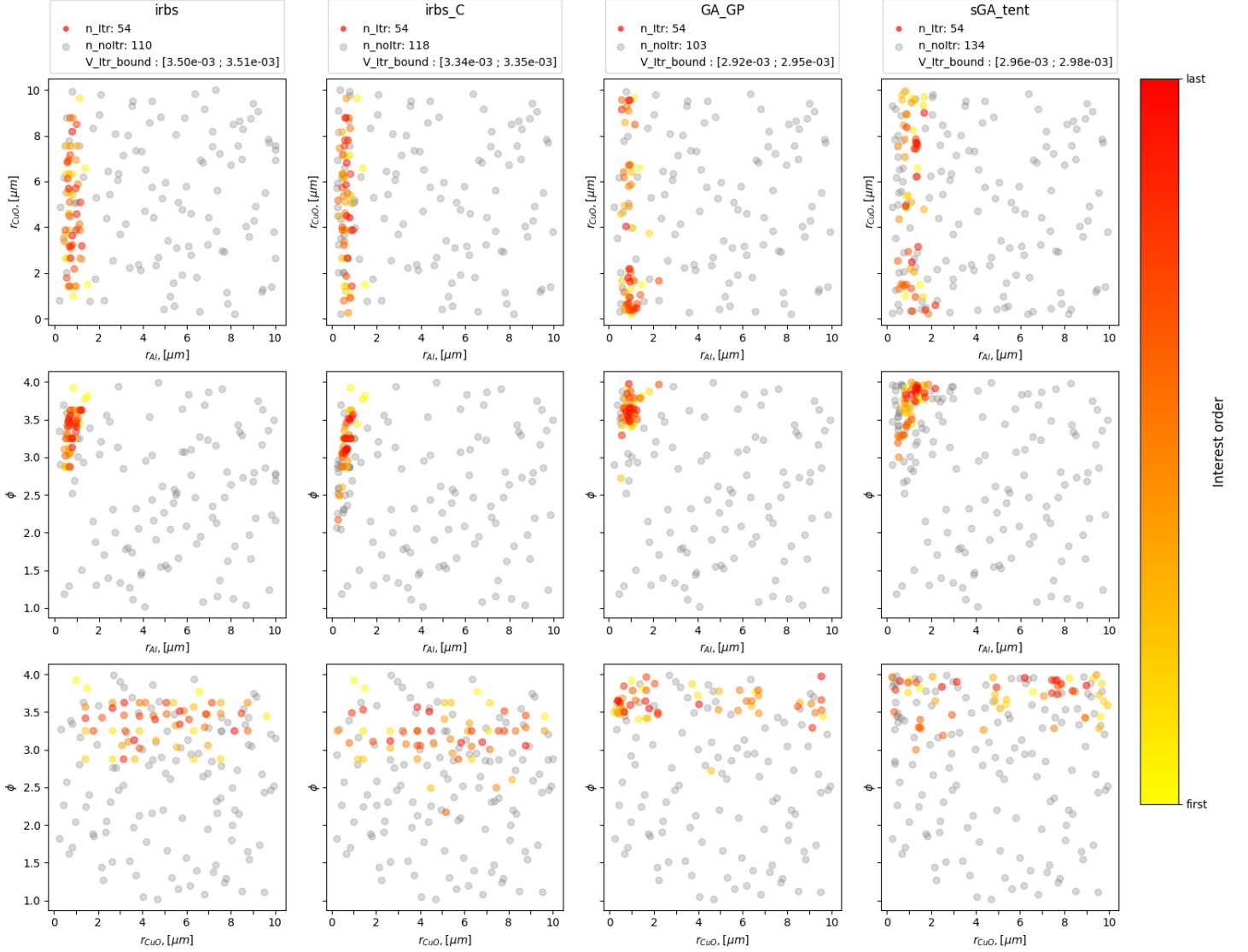


Figure 9: Spatial distribution of sampled points. The points of interest are represented as yellow-to-red balls, indicating their sampling order, while gray balls denote non-interesting points. The sampling stopped after finding 50 points in the region R1.

To find 50 points of interest ( $n_{itr}=54$ ), *irbs* and *irbs\_c* explored 110 and 118 points respectively. *GA\_GP* and *sGA\_tent* sampled only 103 and 134, respectively. Therefore, the modified EGO-based algorithms lost efficiency compared to active learning technique with the addition of one feature. Analyzing the distribution of points in the 3D  $r_{Al}$ - $r_{CuO}$ - $\phi$  design space, it is evident that all methods sample points within similar regions of the feature space. Specifically,  $r_{CuO}$  covers the full range of values,  $r_{Al}$  stays below 2  $\mu m$ , and  $\phi$  ranges from 3 to 4. These results confirm our hypothesis that the feature  $r_{CuO}$  can be disregarded, as CuO decomposes rapidly regardless of its size. We also observe in the diagrams of Figure 9 that the distribution of yellow and red points is uniformly spread throughout the feature space. This indicates that, for all methods, the sampling effectively explores the entire feature space and does not become restricted to a local region.

Next, we built Violin and Kernel Density Estimate (KDE) plots (Figures 9 and 11) detailing the distribution of the

target values (combustion pressure and temperature) of the Al/CuO thermites contained in each database (100 training points + points sampled by each strategy to get 50 points of interest). All the graphs are similar, regardless of the method chosen for sampling the new points. This confirms the previous analysis. In the case with 3 variable features, the algorithms demonstrate equivalent behavior and performance.

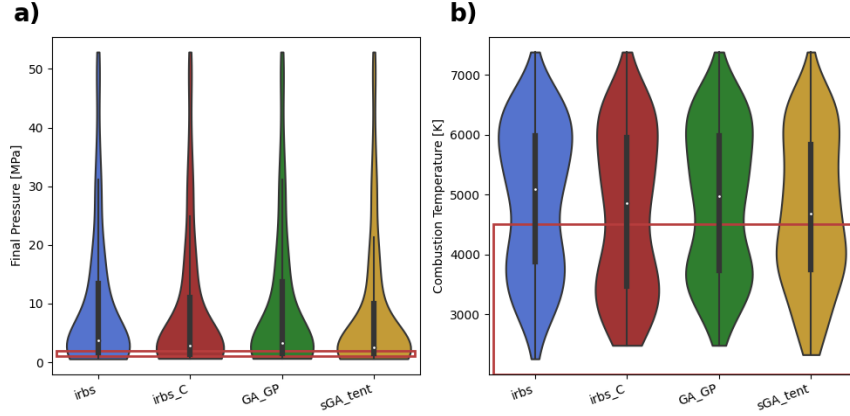


Figure 10: Violin plots of the targets: a) Final pressure, b) Combustion temperature obtained with each algorithm. The red lines mark the limits of the region of interest.

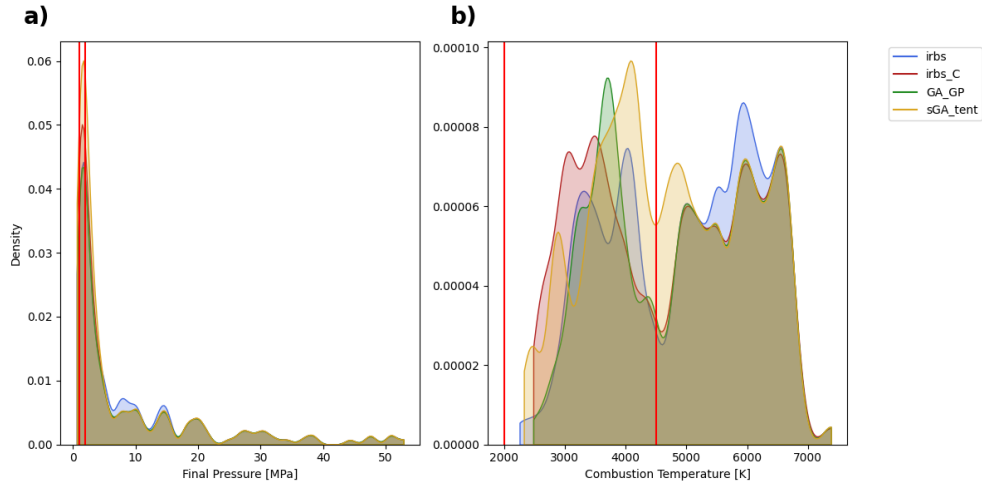


Figure 11: KDE plots of the targets: a) Final pressure, b) Combustion temperature obtained with each algorithm. The red lines mark the limits of the region of interest.

### 3.3.2. Identifying Al/CuO Thermite Materials with Properties Within the R2 Region

This section provides a brief summary of the results obtained in the search for Al/CuO materials that produce combustion temperatures and pressures within the ranges of 5800 - 7000 K and 40 - 56 MPa, respectively (R2), considering three unknown variable features:  $r_{Al}$ ,  $\phi$ , and  $r_{CuO}$ . The spatial distribution of the sampled points for each feature is illustrated in Figure 12.

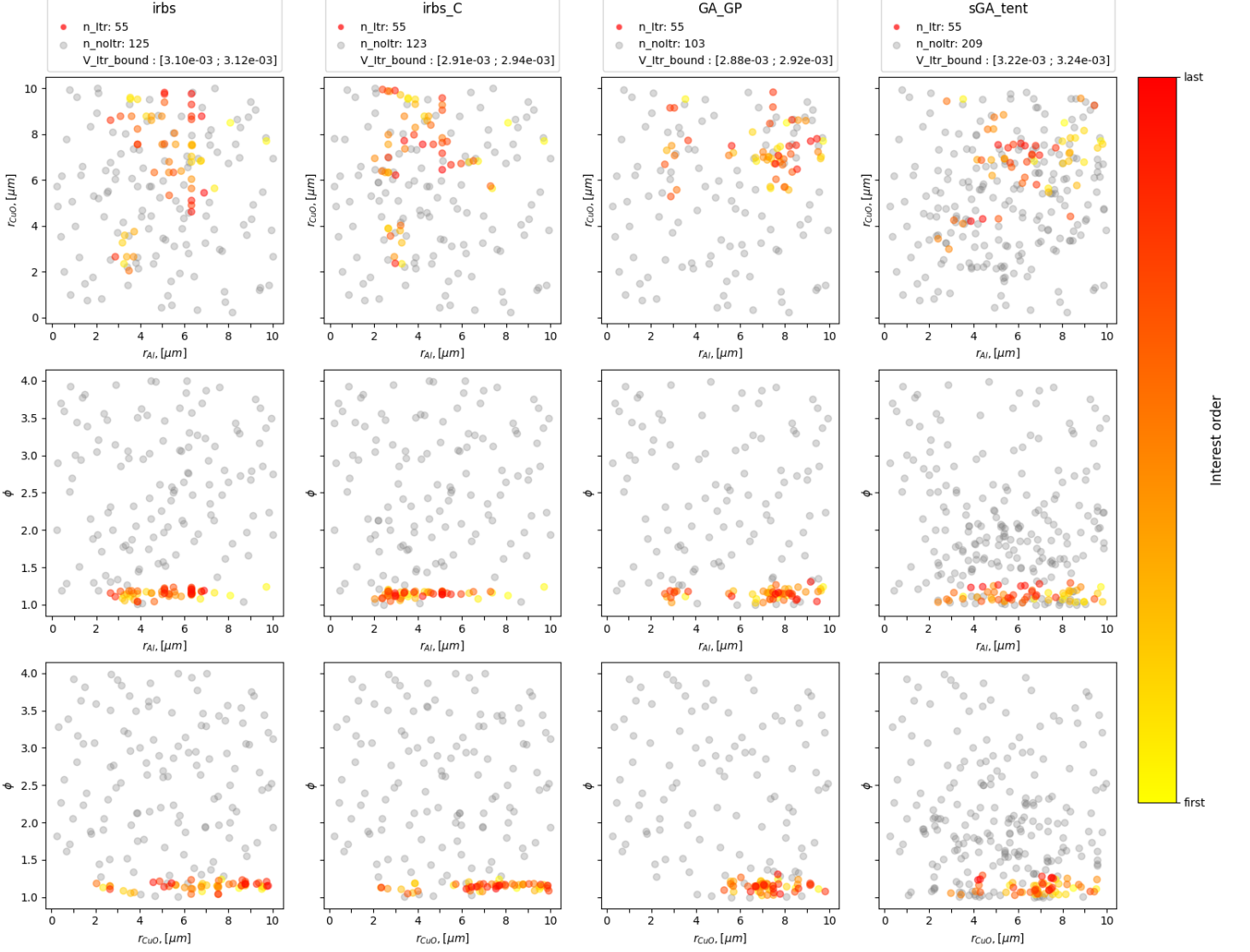


Figure 12: Spatial distribution of sampled points. The points of interest are represented as yellow-to-red balls, indicating their sampling order, while gray balls denote non-interesting points. The sampling stopped after finding 50 points in the region R2.

To find 50 points of interest ( $n_{itr}=55$ ), *irbs* and *irbs\_c* explored 125 and 123 points respectively. *GA\_GP* and *sGA\_tent* explored 103 and 209, respectively. The *sGA\_tent* algorithm lost its efficiency compared to active learning technique with the addition of one feature. Analyzing the distribution of points in the 3D  $r_{Al}$ - $r_{CuO}$ - $\phi$  feature space, it is evident that all methods sample points within similar regions of the feature space. Specifically,  $r_{CuO}$  and  $r_{Al}$  span the full range of values and  $\phi$  is close to 1.

We also observe in the diagrams of Figure 12 that the distribution of yellow and red points is uniformly spread throughout the feature space. This indicates that, for all methods, the sampling effectively explores the entire feature space and does not become restricted to a local region.

Next, we built Violin and Kernel Density Estimate (KDE) plots (Figures 12 and 14) detailing the distribution of the target values (combustion pressure and temperature) of the Al/CuO thermites contained in each database (100 training



points + points sampled by each strategy to get 50 points of interest). GA\_GP and sGA\_tent perform better than irbs and irbs\_C in identifying points that are well-distributed within the target R2 zone. However, the differences between the EGO-based algorithm and the active learning approach are not significant.

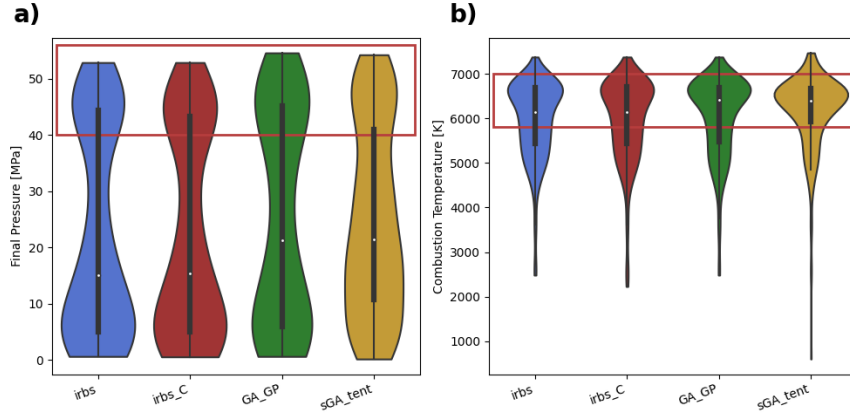


Figure 13: Violin plots of the targets: a) Final pressure, b) Combustion temperature obtained with each algorithm. The red lines mark the limits of the region of interest.

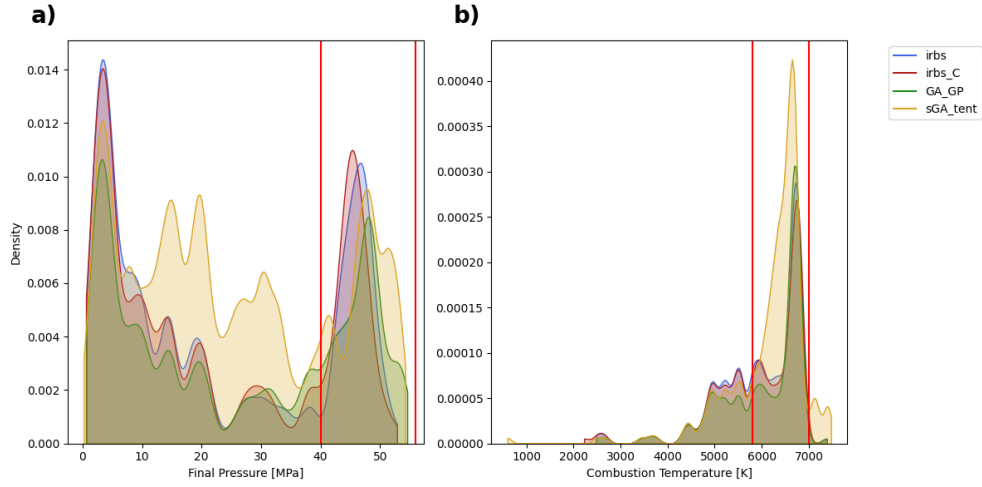


Figure 14: KDE plots of the targets: a) Final pressure, b) Combustion temperature obtained with each algorithm. The red lines mark the limits of the region of interest.

### 3.3.3. Comparison of the methods

Table 3 gives the two metrics, **Coverage** and **Efficiency**, calculated after the search of 50 points of interest in the regions of interest R1 using each of the four algorithms. Clearly, with an Efficiency close to 90% GA\_GP is the most effective to quickly identifies points of interest uniformly spread in the feature space (**coverage** reaches 80% i.e. point of interest are sampled within 80% of the entire design space). While lower that GP\_GA, irbs and irbs\_C feature good efficiencies between 62-78%. All methods do not show significant differences in spatial distribution. Interestingly irbs features a very high coverage for R1 region.

Table 3: Comparison of the algorithms in terms of **Coverage** and **Efficiency**, as defined in Section 2.3 and considering the two regions of interest (R1 and R2).

Experiment	Algorithm	Coverage (in %)	Efficiency (in %)
R1	irbs	<b><math>99.2 \pm 0.1</math></b>	78.12
	irbs_C	$94.6 \pm 0.1$	69.44
	GA_GP	$83.0 \pm 0.4$	<b>87.72</b>
	sGA_tent	$84.0 \pm 0.3$	58.82
R2	irbs	$86.4 \pm 0.3$	62.5
	irbs_C	$81.3 \pm 0.4$	64.10
	GA_GP	$80.6 \pm 0.6$	<b>86.20</b>
	sGA_tent	<b><math>89.7 \pm 0.3</math></b>	30.48

In this experiment, GA\_GP outperforms the other algorithms in quickly identifying points of interest. The expanded feature space has resulted in several significant regions of interest, making it easier to identify these points for both irbs and irbs\_C. However, sGA\_tent struggles to capture these points efficiently, likely due to the loss of information during scalarization with  $f_\lambda$ .

Examining the **Voronoi metric** shown in Figure 16, it is more evident that irbs performs better, as indicated by its higher median, longer lower whisker, shorter upper whisker, and fewer outliers. These outliers represent large, unexplored regions, typically found at the boundaries of clusters. Fewer and smaller outliers generally indicate more effective cluster coverage of the space. Additionally, in the second subplot, irbs\_C demonstrates higher initial volumes, ordered increasingly, capturing fluctuations in  $f_{obj}$ .

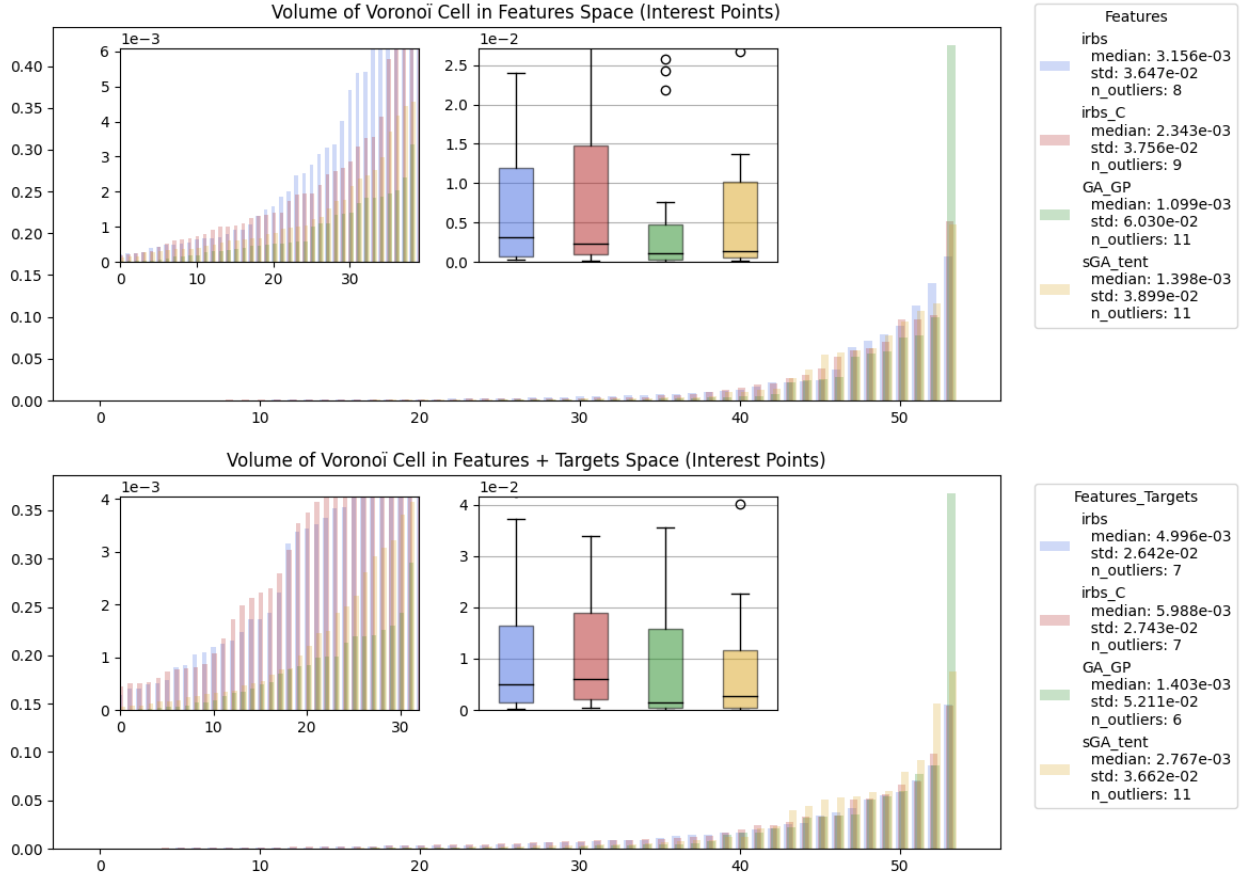


Figure 15: Voronoi volume distribution within the unit hypercube illustrates the spatial influence of each sampled point in region R1. The bar plot, sorted in ascending order, highlights the extent of the regions controlled by each point. The first subplot focuses solely on the feature space, showcasing the distribution of interest points. The second subplot is currently empty due to computation errors encountered during analysis.

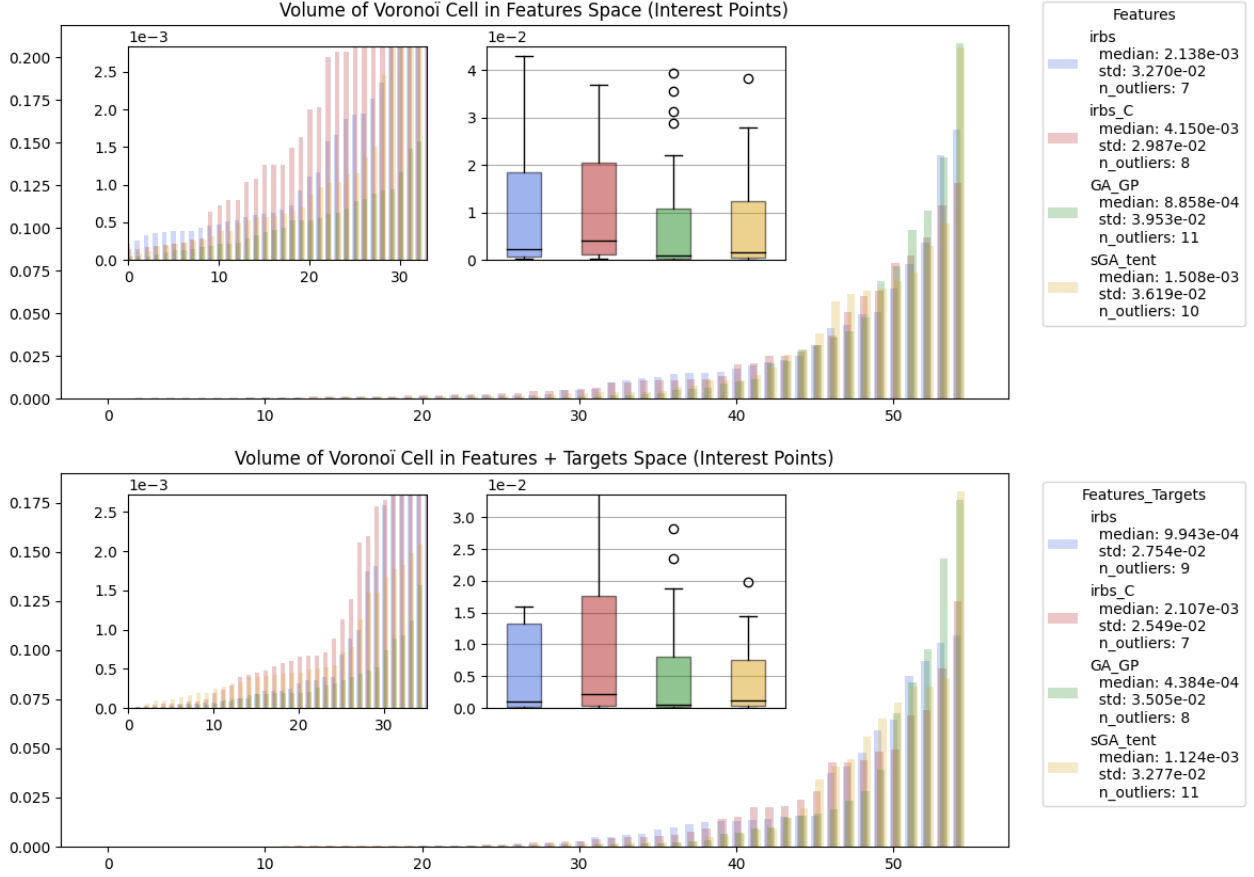


Figure 16: Voronoi volume distribution within the unit hypercube illustrates the spatial influence of each sampled point in region R2. The bar plot, sorted in ascending order, highlights the extent of the regions controlled by each point. The first subplot focuses solely on the feature space, showcasing the distribution of interest points. The second subplot is currently empty due to computation errors encountered during analysis.

## 4. CONCLUSION AND PERSPECTIVE

In this research we introduced two variations of the initial ParEGO algorithm (sGA\_tent and GA\_GP) and compared their efficacy in sampling materials of interest within a 2D ( $r_{Al}$ ,  $\phi$ ) and 3D design space ( $r_{Al}$ ,  $r_{CuO}$ ,  $\phi$ ) with active learning methods presented in a recent paper [34]. Following conclusions can be drawn:

- GA\_GP, which optimizes the Gaussian Process directly using a genetic algorithm based on the interest function, demonstrated primarily strong exploitative behavior.
- While sGA\_tent showed a good ability to identify points of interest that are relatively well-dispersed throughout the feature space, the results are not compelling for high-dimensional feature space. Indeed, as the feature-space dimension increases, sGA\_tent becomes less effective (as observed when comparing results obtained for 2D and 3D feature space). This decline in performance is likely due to the algorithm compressing all outputs into a single value during the scalarization step, which restricts the information available in each dimension. This

"over-simplification" of the original function may lead to a significant loss of valuable information, thereby limiting the algorithm's ability to effectively exploit the data.

Future work should focus on further refining these algorithms. The current hyperparameters were chosen based on intuition and with limited benchmarking, which may not fully capture their optimal performance. A more rigorous approach is required to identify the most suitable algorithm for this type of problem. For instance, the  $\rho$  term in the scalarized objective function of sGA\_tent might be redundant, as it is designed for convergence toward a Pareto front, which is irrelevant to our problem. Instead, efforts should be directed toward tuning the genetic algorithm to more effectively utilize information from the Gaussian Process.

- The custom active learning algorithms (irbs and irbs\_C) which use an acquisition function that linearly combines tailored factors ( $I$  and  $C$ ) within a GPR algorithm, perform very well in selecting Al/CuO thermite compositions tailored for the two specific applications.

It should be noted that irbs\_C, which incorporates the coverage term  $C$ , slightly increases the number of interest points by carefully selecting new samples. For future studies, it would be beneficial to adapt the coverage function to better align with the size of the feature space, such as by expanding it to 5 variables. Currently, the coverage function provides local information, generally improving the coverage score by preventing spheres around interest points from intersecting. However, it will be interesting to develop a function that truly acts as a "coverage function" to enhance the Voronoi volume score, which is a global measure of how uniformly points are distributed. While this new metric would be relevant and interpretable, it could also introduce errors. Additionally, it will be computationally expensive and may be prone to frequent errors due to numerical approximations, leading to difficulties in computation. Future research should focus on optimizing the implementation of this metric or developing a new method to achieve a similar score with greater efficiency and accuracy.

## Acknowledgments

This work was primarily funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grand agreement No. 832889 - PyroSafe). We wish to acknowledge Claudia Ramirez for her work in setting up the github platform and pipeline architecture.

## Data availability

The data that support the findings of this study are available within the article. All the datasets, source codes, and algorithms in Python will be freely available by contacting [rossi@laas.fr](mailto:rossi@laas.fr).

## Authors contribution

C.R. conceived the presented research and supervised the project with the support from A.E. and M.J. C.R. conceived and planned the experiments. R.S. developed the algorithm, solved the mathematical issues and performed the computations. R.S. carried out the numerical experiment and plotted the graphs. Y.S. verified the algorithms and exchanged ideas with R.S. M.J. assisted R.S. and Y.S. on the theoretical formalism and active learning part, especially M.J. proposed the acquisition function. C.R. A.B. lead the manuscript writing. All authors provided critical feedback and analysis.

## References

- [1] I. Abdallah, J. Zapata, G. Lahiner, B. Warot-Fonrose, J. Cure, Y. Chabal, A. Esteve, and C. Rossi, "Structure and Chemical Characterization at the Atomic Level of Reactions in Al/CuO Multilayers," *ACS Applied Energy Materials*, vol. 1, pp. 1762–1770, Apr. 2018. Publisher: American Chemical Society.
- [2] Y. Aly, M. Schoenitz, and E. L. Dreizin, "Aluminum-metal reactive composites," *Combustion Science and Technology*, vol. 183, no. 10, pp. 1107–1132, 2011.

- [3] V. E. Sanders, B. W. Asay, T. J. Foley, B. C. Tappan, A. N. Pacheco, and S. F. Son, "Reaction propagation of four nanoscale energetic composites (al/moo3, al/wo3, al/cuo, and b12o3)," *Journal of Propulsion and Power*, vol. 23, no. 4, pp. 707–714, 2007.
- [4] B. S. Bockmon, M. L. Pantoya, S. F. Son, B. W. Asay, and J. T. Mang, "Combustion velocities and propagation mechanisms of metastable interstitial composites," *Journal of Applied Physics*, vol. 98, p. 064903, 09 2005.
- [5] K. M. De Souza and M. J. de Lemos, "Advanced one-dimensional modeling of thermite reaction for thermal plug and abandonment of oil wells," *International Journal of Heat and Mass Transfer*, vol. 205, p. 123913, 2023.
- [6] V. Singh, T. Wu, E. Hagen, L. Salvagnac, C. Tenaillieu, A. Estève, M. R. Zachariah, and C. Rossi, "How positioning of a hard ceramic tib2 layer in al/cuo multilayers can regulate the overall energy release behavior," *Fuel*, vol. 349, p. 128599, 2023.
- [7] T. Wu, V. Singh, B. Julien, C. Tenaillieu, A. Estève, and C. Rossi, "Pioneering insights into the superior performance of titanium as a fuel in energetic materials," *Chemical Engineering Journal*, vol. 453, p. 139922, 2023.
- [8] T. Wu, B. Julien, H. Wang, S. Pelloquin, A. Esteve, M. R. Zachariah, and C. Rossi, "Engineered porosity-induced burn rate enhancement in dense al/cuo nanothermites," *ACS Applied Energy Materials*, vol. 5, no. 3, pp. 3189–3198, 2022.
- [9] P. Gandhi, M. Schoenitz, and E. Dreizin, "Evaluation and design of metal-based gas-generating energetic materials," *Combustion and Flame*, vol. 249, p. 112615, 2023.
- [10] M. R. Zachariah, "Nanoenergetics: Hype, reality and future," *Propellants, Explosives, Pyrotechnics*, vol. 38, p. 7, 2013.
- [11] C. Rossi, "Metallized reactive materials – a road to clean and sustainable pyrotechnics," *Propellants, Explosives, Pyrotechnics*, vol. 48, no. 5, p. e202380531, 2023.
- [12] X. Zhou, M. Torabi, J. Lu, R. Shen, and K. Zhang, "Nanostructured energetic composites: Synthesis, ignition/combustion modeling, and applications," *ACS Applied Materials & Interfaces*, vol. 6, no. 5, pp. 3058–3074, 2014.
- [13] T. Wu, G. Lahiner, C. Tenaillieu, B. Reig, T. Hungria, A. Esteve, and C. Rossi, "Unexpected enhanced reactivity of aluminized nanothermites by accelerated aging," *Chemical Engineering Journal*, vol. 418, p. 129432, 2021.
- [14] C. Rossi, "Engineering of Al/CuO Reactive Multilayer Thin Films for Tunable Initiation and Actuation," *Propellants, Explosives, Pyrotechnics*, vol. 44, no. 1, pp. 94–108, 2019.
- [15] M. Mursalat, M. Schoenitz, D. E. L., A. Neveu, and F. Francqui, "Spherical boron powders prepared by mechanical milling in immiscible liquids," *Powder Technology*, vol. 388, pp. 41–50, 2021.
- [16] A. Abraham, N. W. Piekil, C. J. Morris, and E. L. Dreizin, "Combustion of energetic porous silicon composites containing different oxidizers," *Propellants, Explosives, Pyrotechnics*, vol. 41, no. 1, pp. 179–188, 2016.
- [17] X. Liu, M. Schoenitz, and E. L. Dreizin, "Preparation, ignition, and combustion of magnesium-calcium iodate reactive nano-composite powders," *Chemical Engineering Journal*, vol. 359, pp. 955–962, 2019.
- [18] G. C. Egan, E. J. Mily, J.-P. Maria, and M. R. Zachariah, "Probing the reaction dynamics of thermite nanolaminates," *The Journal of Physical Chemistry C*, vol. 119, no. 35, pp. 20401–20408, 2015.
- [19] V. Singh, T. Wu, C. Tenaillieu, T. Hungria, A. Estève, and C. Rossi, "Investigating the reaction mechanism of zirconium as a fuel in reactive multilayer films via multimodal analysis," *Chemical Engineering Journal*, vol. 495, p. 153357, 2024.
- [20] W. Zhao, H. Wang, D. J. Kline, X. Wang, T. Wu, J. Xu, H. Ren, and M. R. Zachariah, "Influence of titanium addition on performance of boron-based thermites," *Chemical Engineering Journal*, vol. 438, p. 134837, 2022.
- [21] A. L. Polk, K. Lee, A. Chintersingh, M. R. Flickinger, E. R. Valdes, W. L. Gardner, and T. P. Weihs, "Effect of composition and process control agents on the microstructure and ignition properties of ball-milled al-zr powders," *Powder Technology*, vol. 427, p. 118729, 2023.
- [22] S. F. Son, B. W. Asay, T. J. Foley, R. A. Yetter, M. H. Wu, and G. A. Risha, "Combustion of nanoscale al/moo3 thermite in microchannels," *Journal of Propulsion and Power*, vol. 23, no. 4, pp. 715–721, 2007.
- [23] H. Wang, G. Jian, W. Zhou, J. B. DeLisio, V. T. Lee, and M. R. Zachariah, "Metal iodate-based energetic composites and their combustion and biocidal performance," *ACS Applied Materials & Interfaces*, vol. 7, no. 31, pp. 17363–17370, 2015.
- [24] L. Glavier, G. Taton, J. M. Ducéré, V. Baijot, P. Pinon, T. Calais, A. Estève, M. Djafari-Rouhani, and C. Rossi, "Nanoenergetics as pressure generator for nontoxic impact primers: Comparison of al/bi2o3, al/cuo, al/moo3 nanothermites and al/ptfe," *Combustion and Flame*, vol. 162, no. 5, pp. 1813–1820, 2015.
- [25] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. de Freitas, "Taking the human out of the loop: A review of bayesian optimization," *Proceedings of the IEEE*, vol. 104, no. 1, pp. 148–175, 2016.
- [26] J. Knowles and E. J. Hughes, "Multiobjective optimization on a budget of 250 evaluations," in *Evolutionary Multi-Criterion Optimization* (C. A. Coello Coello, A. Hernández Aguirre, and E. Zitzler, eds.), (Berlin, Heidelberg), pp. 176–190, Springer Berlin Heidelberg, 2005.
- [27] A. O. Oliynyk, E. Antono, T. D. Sparks, L. Ghadbeigi, M. W. Gaultois, B. Meredig, and A. Mar, "High-throughput machine-learning-driven synthesis of full-heusler compounds," *Chemistry of Materials*, vol. 28, no. 20, pp. 7324–7331, 2016.
- [28] T. Ueno, T. D. Rhone, Z. Hou, T. Mizoguchi, and K. Tsuda, "Combo: An efficient bayesian optimization library for materials science," *Materials Discovery*, vol. 4, pp. 18–21, 2016.
- [29] R. Ouyang, E. Ahmetcik, C. Carbogno, M. Scheffler, and L. M. Ghiringhelli, "Simultaneous learning of several materials properties from incomplete databases with multi-task siso," *Journal of Physics: Materials*, vol. 2, p. 024002, 2019.
- [30] G. Pilania, J. E. Gubernatis, and T. Lookman, "Structure classification and melting temperature prediction in octet ab solids via machine learning," *Physical Review B*, vol. 91, p. 214302, Jun 2015.
- [31] P. C. Jennings, S. Lysgaard, J. S. Hummelshøj, T. Vegge, and T. Bligaard, "Genetic algorithms for computational materials discovery accelerated by machine learning," *npj Computational Materials*, vol. 5, no. 1, p. 46, 2019.
- [32] H. Matter, "Selecting optimally diverse compounds from structure databases : A validation study of two-dimensional and three-dimensional molecular descriptors," *Journal of Medicinal Chemistry*, vol. 40, no. 8, pp. 1219–1229, 1997.
- [33] F. Di Fiore, M. Nardelli, and L. Mainini, "Active learning and bayesian optimization: A unified perspective to learn with a goal," *Archives of Computational Methods in Engineering*, Apr 2024.
- [34] L. Carreira, L. Pillemont, Y. Sami, N. Richard, A. Esteve, M. Jonckheere, and C. Rossi, "Targeted nano-energetic material exploration through bayesian algorithm implementation," *Energy Materials Frontiers*, 2024.
- [35] D. R. Jones, M. Schonlau, and W. J. Welch, "Efficient global optimization of expensive black-box functions," *Journal of Global optimization*,

- vol. 13, no. 4, pp. 455–492, 1998.
- [36] J. Knowles, “Parego: A hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems,” Tech. Rep. TR-COMPSYSBIO-2004-01, University of Manchester, September 2004.
  - [37] J. Knowles and E. J. Hughes, “Multiobjective optimization on a budget of 250 evaluations,” in *International Conference on Evolutionary Multi-Criterion Optimization*, pp. 176–190, Springer, 2005.
  - [38] K. M. de Souza and M. J. S. de Lemos, “Detailed numerical modeling and simulation of  $\text{Fe}_2\text{O}_3\text{Al}$  thermite reaction,” *Propellants, Explosives, Pyrotechnics*, vol. 46, no. 5, pp. 806–824, 2021.
  - [39] A. Nicollet, L. Salvagnac, V. Baijot, A. Estève, and C. Rossi, “Fast circuit breaker based on integration of  $\text{Al/CuO}$  nanothermites,” *Sensors and Actuators A: Physical*, vol. 273, pp. 249–255, 2018.
  - [40] N. Banerjee, Y. Xie, M. Rahman, H. Kim, and C. H. Mastrangelo, “From chips to dust: The mems shatter secure chip,” in *2014 IEEE 27th International Conference on Micro Electro Mechanical Systems (MEMS)*, pp. 1123–1126, 2014.
  - [41] E. Tichtchenko, V. Folliet, O. Simonin, B. Bédard, L. Glavier, A. Esteve, and C. Rossi, “Combustion model for thermite materials integrating explicit and coupled treatment of condensed and gas phase kinetics,” *Proceedings of the Combustion Institute*, vol. 39, no. 3, pp. 3637–3645, 2023.
  - [42] V. Baijot, M. Djafari-Rouhani, C. Rossi, and A. Esteve, “A multi-phase micro-kinetic model for simulating aluminum based thermite reactions,” *Combustion and Flame*, vol. 180, pp. 10–19, 2017.
  - [43] V. Baijot, L. Glavier, J.-M. Ducéré, M. DjafariRouhani, C. Rossi, and A. Estève, “Modeling the pressure generation in aluminum-based thermites,” *Propellants, Explosives, Pyrotechnics*, vol. 40, no. 3, pp. 402–412, 2015.
  - [44] E. Tichtchenko, B. Bedat, O. Simonin, L. Glavier, D. Gauchard, A. Esteve, and C. Rossi, “Comprehending the influence of the particle size and stoichiometry on  $\text{Al/CuO}$  thermite combustion in close bomb: A theoretical study,” *Propellants, Explosives, Pyrotechnics*, vol. 48, no. 7, p. e202200334, 2023.