



**HAL**  
open science

## Looking for a hyper polyhedron within the multidimensional space of Design Space from the results of Designs of Experiments

Diane Manzon, Badih Ghattas, Magalie Claeys-Bruno, Sophie Declomesnil, Christophe Carité, Michelle Sergent

### ► To cite this version:

Diane Manzon, Badih Ghattas, Magalie Claeys-Bruno, Sophie Declomesnil, Christophe Carité, et al.. Looking for a hyper polyhedron within the multidimensional space of Design Space from the results of Designs of Experiments. *Chemometrics and Intelligent Laboratory Systems*, 2023, 232, pp.104712. 10.1016/j.chemolab.2022.104712 . hal-04021786

**HAL Id: hal-04021786**

**<https://amu.hal.science/hal-04021786>**

Submitted on 23 Aug 2023

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

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

# Looking for a hyper polyhedron within the multidimensional space of Design Space from the results of Designs of Experiments

Diane Manzon <sup>a,b</sup>, Badih Ghattas <sup>c,\*</sup>, Magalie Claeys-Bruno <sup>a</sup>, Sophie Declomesnil <sup>b</sup>,  
Christophe Carité <sup>b</sup>, Michelle Sergent <sup>a</sup>

<sup>a</sup> Institut Méditerranéen de Biodiversité et d'Ecologie marine et continentale, Aix-Marseille Université, UMR CNRS IRD Avignon Université, Site de l'Etoile, Marseille, France

<sup>b</sup> 4D pharma plc, 5th Floor, 9 Bond Court, LS1 2JZ, Leeds, UK

<sup>c</sup> AMSE, Aix Marseille Université, CNRS, Marseille, France

## Abstract

In pharmaceutical studies, the Quality by Design (QbD) approach is increasingly being implemented to improve product development. Product quality is tested at each step of the manufacturing process, allowing a better process understanding and a better risk management, thus avoiding manufacturing defects. A key element of QbD is the construction of a Design Space (DS), i.e., a region in which the specifications on the output parameters should be met. Among the various possible construction methods, Designs of Experiments (DoE), and more precisely Response Surface Methodology, represent a perfectly adapted tool. The DS obtained may have any geometrical shape; consequently, the acceptable variation range of an input may depend on the value of other inputs. However, the experimenters would like to directly know the variation range of each input so that their variation domains are independent. In this context, we developed a method to determine the “Proven Acceptable **Independent** Range” (PAIR). It consists of looking for all the hyper polyhedra included in the multidimensional DS and selecting a hyper polyhedron according to various strategies. We will illustrate the performance of our method on different DoE cases.

**Keywords:** Quality by Design (QbD); Design of Experiments (DoE); Response Surface Methodology (RSM); Design Space (DS); Proven Acceptable Independent Range (PAIR)

## 1. Introduction

Since the early 2000s, the Quality by Design (QbD) approach in the pharmaceutical industry has been encouraged by the Food and Drug Administration [1]. The guideline ICH Q8 explains that the quality of a product must be tested at each step of the manufacturing process [2]. This concept consists of different elements (Fig. 1).

The first element is the definition of a Quality Target Product Profile that sets the objectives of the development [3]. The second element is the list establishment of properties regarding the product quality, called Critical Quality Attributes (CQA) [4]. The third element is a crucial step. On the one hand, the risk assessment identifies critical inputs variables of a process or a formula that could impact the CQA. On the other hand, a relationship between inputs and CQA is then determined [5]. The last element of QbD is the control strategy that leads to continuous quality improvement [6].

The implementation of QbD in the pharmaceutical industry improves product development by increasing process understanding [7] and

providing better risk management in order to deliver a product of good quality [8].

The risk assessment step is a crucial step in identifying critical input parameters of the process that may impact on CQA. Different approaches can be used [9,10] and among these, Design of Experiments (DoE), and more precisely modelling studies (Response Surface Methodology, RSM [11]) are an essential tool [12]. It is used to predict the variation of the CQA studied over the whole variation range of the input parameters. When there are many inputs and CQA, it is necessary to find a “compromise zone” in which all CQA will comply with the specifications and for this purpose, multi-criteria optimisation methods, such as the desirability function [13], are used. With this method, an optimum point i.e. the point that has the highest overall desirability value can be calculated. This optimum point gives the conditions of the factors that best meet the specifications, i.e. the experimental conditions for which the CQA tend towards their target values. The position of this optimum point will depend on the specifications and the weight attributed to each CQA. In addition, Food and Drug Administration recommends the

\* Corresponding author.

E-mail address: [badih.ghattas@gmail.com](mailto:badih.ghattas@gmail.com) (B. Ghattas).

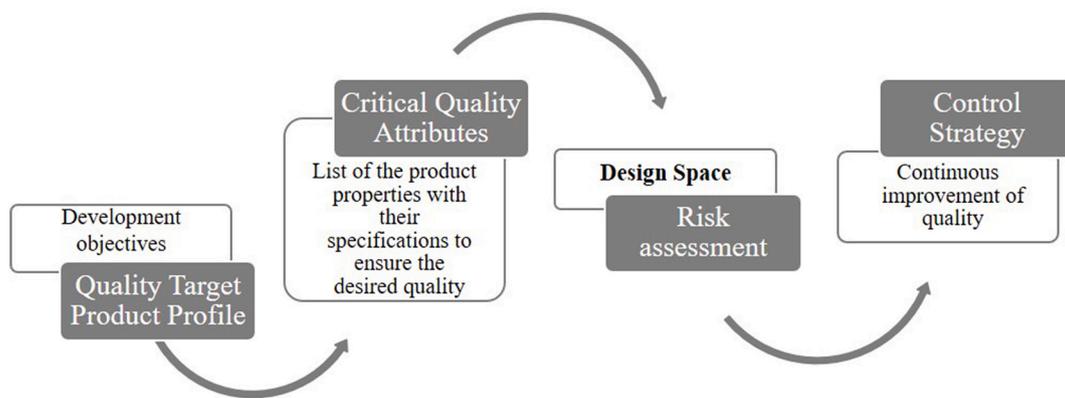


Fig. 1. Elements of QbD approach.

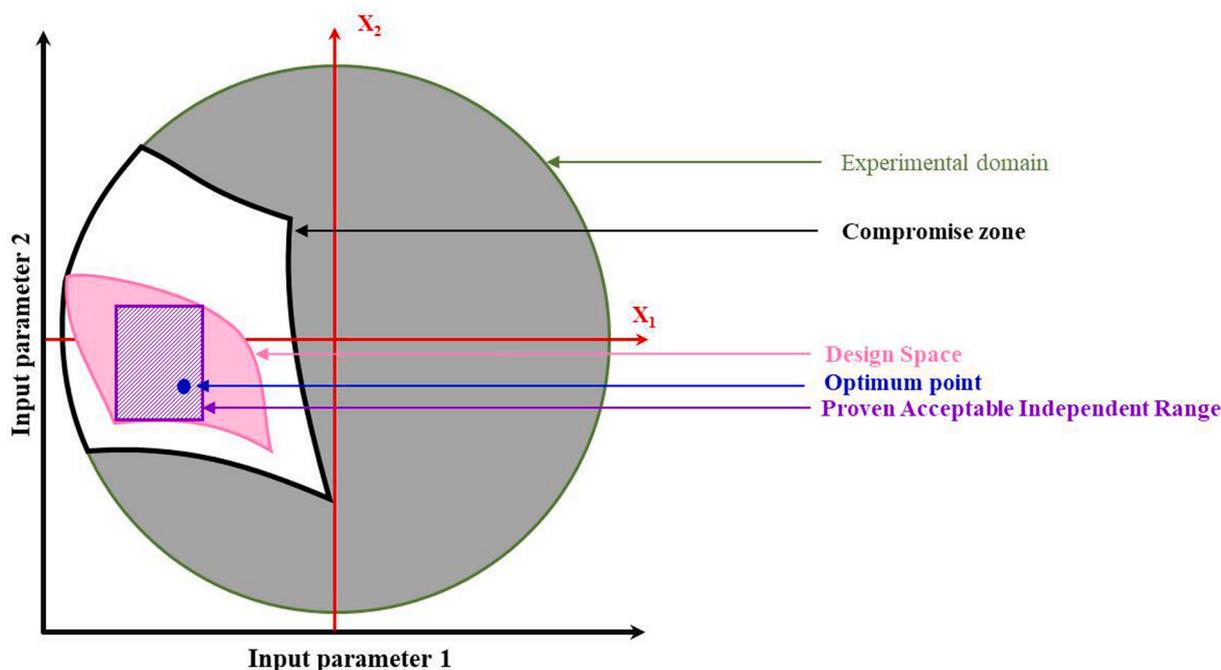


Fig. 2. Illustration of QbD concepts in experimental domain - 2D representation.

establishment of a Design Space (DS), defined in the ICHQ8 guideline as “the multidimensional combination and interaction of input variables and process parameters that have been demonstrated to provide assurance of quality” [2]. In other words, it means that DS is a region in which all input parameters can be varied without altering the quality of the product. Different methods can be used to define this DS and we have opted for probabilistic approaches. In a previous work, we proposed a comparison of four DS construction methods from the DoE results in the QbD context to highlight their advantages and their limits [14]. The difficulty in constructing the DS is that any geometric shape will be obtained, which means when we look for the variation range of a parameter, we have to set the values for the other parameters. This variation range is called Proven Acceptable Range (PAR) [2]. According to the guideline ICH Q8, PAR is defined as “a characterised range of a process parameter for which operation within this range, while keeping other parameters constant, will result in producing a material meeting relevant quality criteria” [2]. For example, if a first parameter (pH) is maintained at 5, the second parameter (temperature) can be varied between 15 and 25 °C. However, if another pH value is selected, the temperature variation range will be different.

To overcome this problem, we seek to determine the variation range

of each input, so that their respective variation domains are independent, inside the DS. The problem then turns out to be the number of dimensions, i.e. the number of input parameters. For example, for a number of dimensions equal to 2, we look for the largest rectangle inside the DS, which may seem quite easy, but when working with many inputs and many CQA, the challenge is to find the largest hyper polyhedron in the multidimensional space of the DS, whose faces will be parallel to the input axis, which is much more complicated. We called this concept “Proven Acceptable Independent Range”. To summarise, Fig. 2 shows the different concepts we have just detailed for a two-dimensional (2D) case.

In this article, we present an algorithm which looks for all possible hyper polyhedra within the DS, and determines the PAIR using various strategies and considering different selection criteria. We apply this method on different DoE cases to compare these different strategies and to show that the proposed approach might be useful and it is flexible enough to satisfy different needs of the user in pharmaceutical and chemical studies.

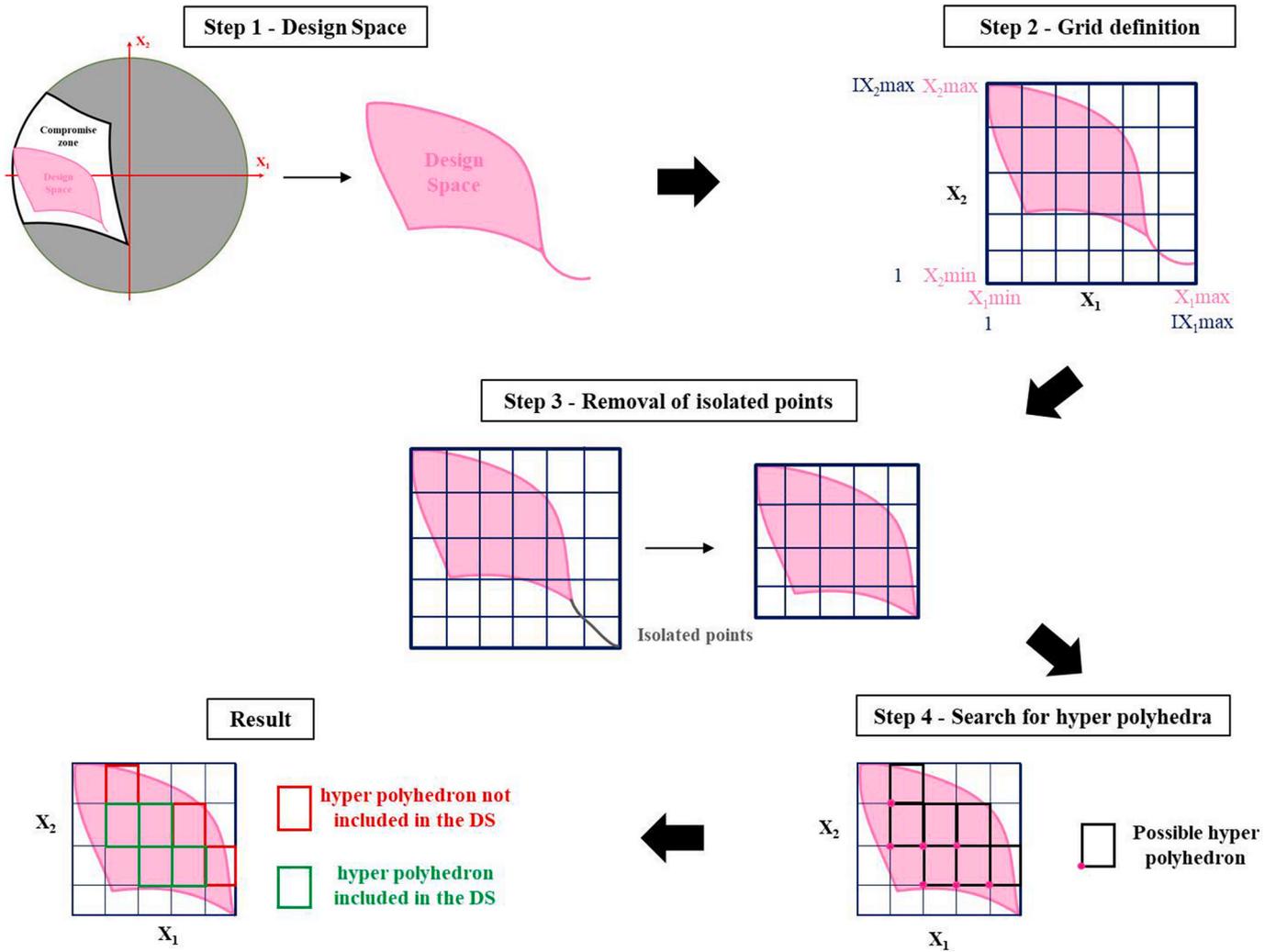


Fig. 3. Search method of hyper polyhedra included in the DS.

## 2. Methods

Different approaches, based on robust setpoint and using Monte-Carlo simulations, can be used to determine the PAR of inputs [15]. These approaches consider in general a limited number of inputs and are time-consuming mainly when the variation step is small.

Our objective is to propose a novel approach to determine the PAIR that can be used even in high dimension. Our methodology is based on two main stages:

- Retrieving the set of all possible hyper polyhedra included in the DS
- Selecting a hyper polyhedron based on a user specified criterion

We now describe these stages in detail.

### 2.1. Retrieving the set of all eligible hyper polyhedra

Our objective at this stage is to look for all possible hyper polyhedra that are inscribed inside the DS.

#### 2.1.1. Principle

This is done in four steps illustrated in Fig. 3.

The first step consists of constructing a DS using a suitable method described in a previous work [14]. The second step creates a regular grid covering this DS which will facilitate the search for hyper polyhedra. The discretization step of the grid is set by the user. Depending on the

shape of the DS, and after many 2D tests, we found that “isolated” points cannot lead to interesting inscribed hyper plans and should be eliminated. To exclude these points, we proceeded in two steps. First, we detect the contour of the DS using the notion of close neighbours and then, we identify and remove the isolated points from the DS. These intermediate steps constitute the third step of the method. The fourth step consists in searching, from each point of the DS, among all the hyper polyhedra in the grid, those which are inscribed in the DS.

Fig. 3 presents a simple 2D case. The major complexity of this approach lies in the exploration of the DS when the dimension  $k$ , corresponding to the number of factors in the study, is high.

#### 2.1.2. Description of the different steps of the algorithm

After fixing the inputs, our algorithm consists of three parts (construction of a grid point, removal of isolated points and search for the hyper polyhedra included in the DS) which will be detailed below.

**2.1.2.1. Inputs of the algorithm.** The algorithm takes as input two elements: a table of dimension  $n \times k$  where  $n$  corresponds to the number of points in the DS,  $k$  represents the number of factors (dimension) and a variation step, defined during the construction of the DS.

**2.1.2.2. Grid choice.** The regular grid is determined using the empirical range ( $max - min$ ) in each dimension together with the fixed step. In order to have the same scale in all dimensions, we recoded the DS to

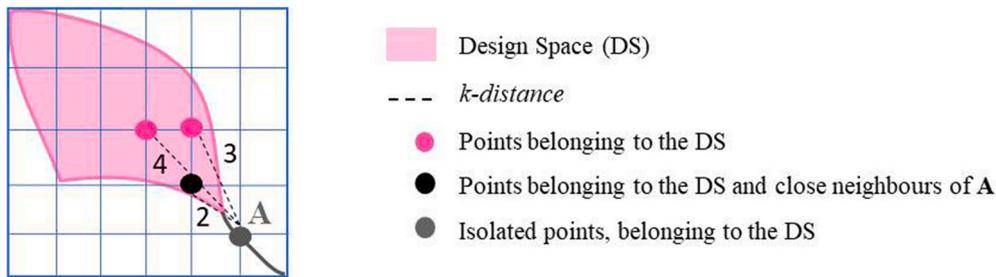


Fig. 4. Detection of isolated points – 2D example.

### Inputs

**DS** = Design Space (table  $n \times k$ )

**k** = number of factors (dimensions)

**step** = discretization step

### Algorithm

1. Calculation of the minimum and the maximum ( $X_{\min}$  et  $X_{\max}$ ) of each dimension of the DS
  2. Normalization of the DS to have a variation step equal to 1 : definition of the **I-DS** (I for Integer)
  3. Calculation of the **Icontour**: points of the **I-DS** whose number of  $k$ -close neighbours\* is less than  $3^k - 1$
  4. Calculation of isolated points : points of the Icontour whose number of 1-close neighbours\*\* is less than  $k$ 
    - a. **IcontourSPI** (**Icontour** without isolated points) = Icontour
    - b. As long as IcontourSPI contains isolated points
      - i. Calculation of the minkowski distance for all points of IcontourSPI between them
      - ii. Keep only points with a number of 1-close neighbours\*\* greater than or equal to  $k$
  5. Calculation of **I-DS-SPI** (**I-DS** without isolated points) = I-DS – Icontour + IcontourSPI
  6. **Looking for all possible hyper polyhedra** in the I-DS-SPI : a hyper polyhedron is characterized by its two diagonal points IPI and IPF
    - a. For each **IPI** point of the I-DS-SPI:
      - i. Looking for **all points on the I-DS-SPI** which have  $k$  coordinates greater than the coordinates of the IPI point under consideration = **candidate IPI points**
      - ii. For each candidate IPF point:
        - **Generation of all points of the polyhedron**
          - I-DS-SPI membership test
            - Yes: Keep the IPI points and the IPF point considered = a possible hyper polyhedron
            - No: Move on to the next candidate IPI point
7. **Conservation of all hyper polyhedra included** in the **I-DS-SPI**

\* a  $k$ -close neighbour of a point is a point which is less than  $k$  away from it, in the sense of minkowski distance.

\*\* a 1-close neighbour of a point is a point distant from it by 1, in the sense of minkowski distance.

Fig. 5. Algorithm – Exploring all the hyper polyhedra inscribed within the DS.

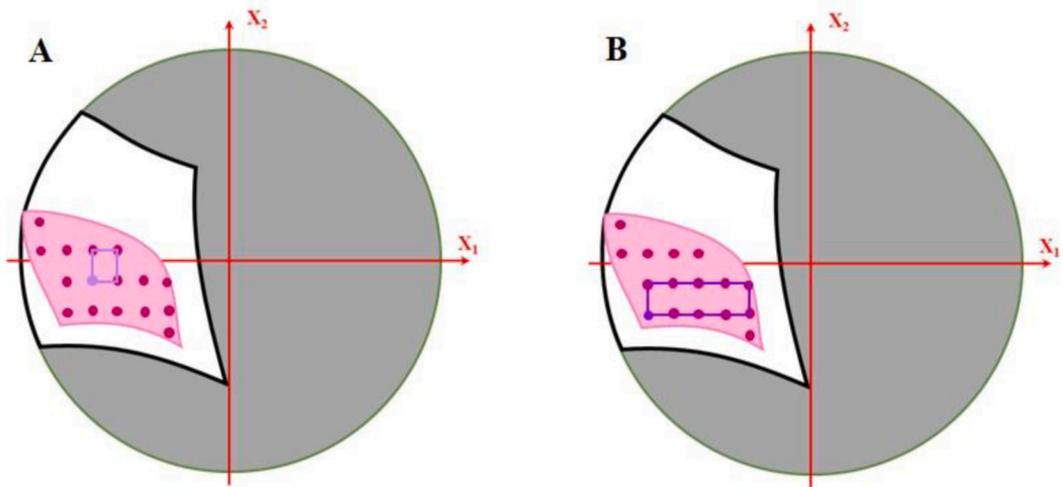


Fig. 6. Strategy 1 - A: a possible rectangle, B: the largest area rectangle.

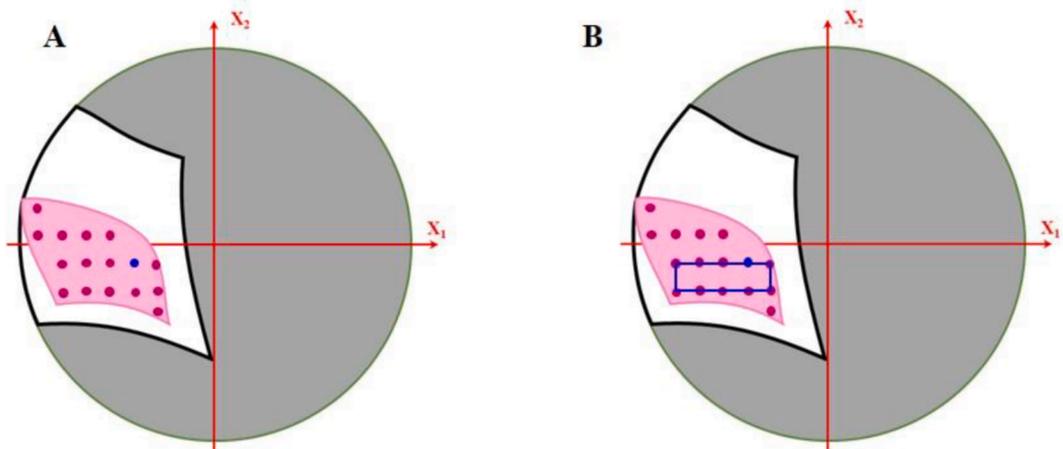


Fig. 7. Strategy 1 - A: Optimum in blue, B: the rectangle with the largest area including the optimum. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

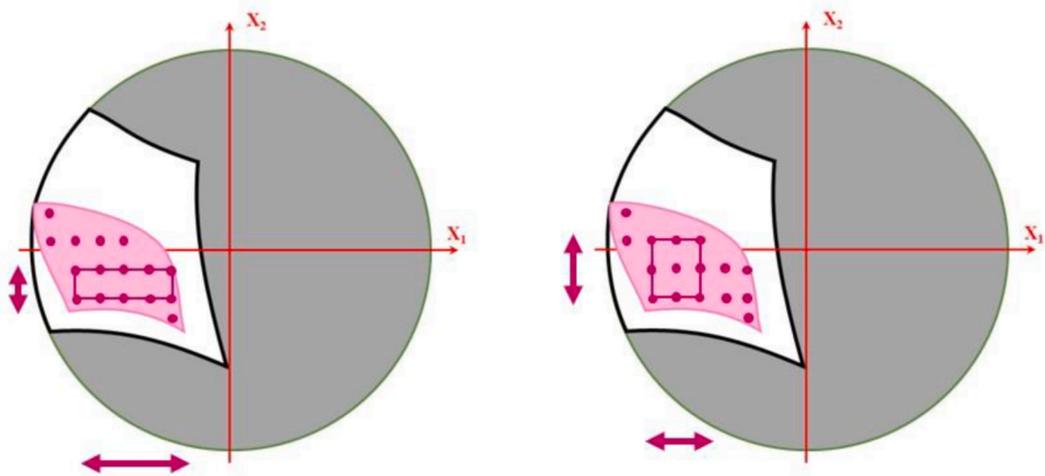


Fig. 8. Strategy 2.

**Table 1**  
Studied parameters – 2D case study.

Inputs	Outputs and its specifications
$X_1$ = Liquid (g) [2500; 3000]	$Y_1$ : 20% < Granulometry < 45%
$X_2$ = Mixing time (s) [180; 360]	$Y_2$ : Friability < 0.5
	$Y_3$ : Hardness > 11
	$Y_4$ : Transmission ratio > 80%
	$Y_5$ : Cohesion index > 600

have a discretization step equal to 1 so as to simplify the calculations. This second part is illustrated in step 2 of Fig. 3.

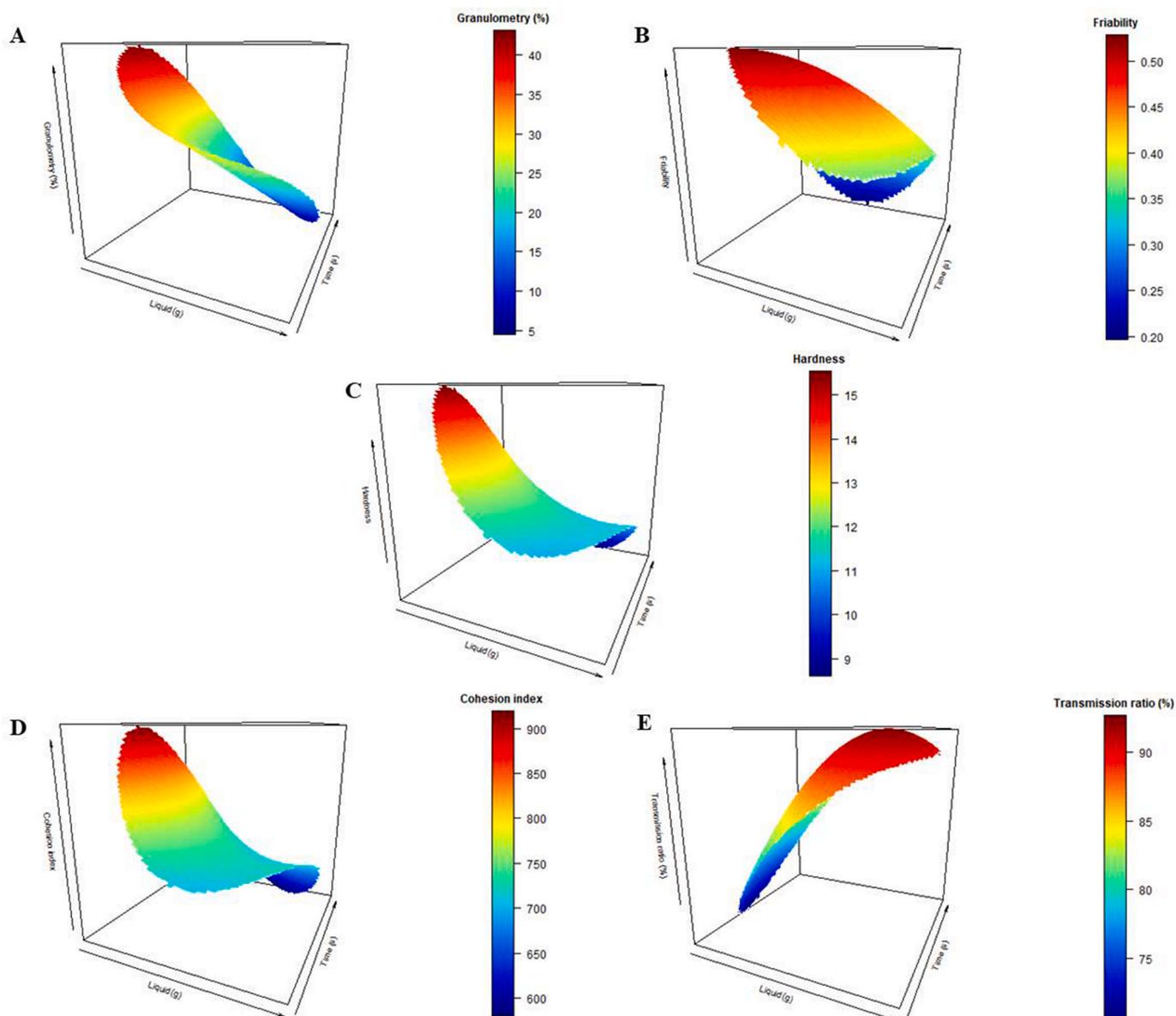
**2.1.2.3. Removal of isolated points.** The detection of isolated points in a DS goes through an intermediate step which consists of calculating the contour of this DS. To do this, we use the notion of close neighbours of a point. We consider that a point belongs to the contour if and only if its number of close neighbours is less than  $3^k - 1$ . Thus, a point A in dimension  $k$  belonging to the DS is isolated, if its  $k$ -distance, e.g. the number of points on the DS which are at a distance from A less than or

equal to  $k$ , is strictly less than  $k + 1$  points. The detection of an isolated point is shown in Fig. 4.

The DS is represented by the pink zone. We consider the point A (in grey), and we calculate the  $k$ -distance from this point to the other points. The value of the  $k$ -distance is noted next the dotted lines. We remark that one dot, in black, has a  $k$ -distance of 2, therefore point A is an isolated point and will be removed from the DS. This process is repeated until all isolated points are completely removed from the DS.

**2.1.2.4. Search for the possible hyper polyhedra included in the DS.** Exploring a multidimensional grid for a fixed dimension  $k$ , may be naturally done using  $k$  nested loops. These nested loops are created dynamically. Therefore, our algorithm and the corresponding code can be used for any dimension without modification. This allows us to reduce the complexity of a  $k$ -dimensional space exploration.

For each point (called IPI) from the DS, all points whose  $k$  coordinates are higher than those of IPI are tested. Each of these points (called IPF) together with IPI, define a different polyhedron. If the actual IPF is outside the DS, then it is ignored, as well as any point whose at



**Fig. 9.** Projection plans.  
A:  $Y_1$  - Granulometry, B:  $Y_2$  - Friability, C:  $Y_3$  - Hardness, D:  $Y_4$  - Transmission ratio, E:  $Y_5$  - Cohesion index.

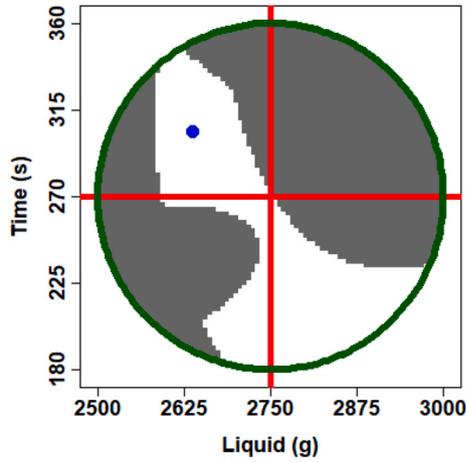


Fig. 10. Compromise zone.

**Table 2**  
Coordinates of maximum desirability and response predictions at this point.

Parameters	Optimum
Liquid (g)	2637.5
Time (s)	303.75
Granulometry (%)	26.79
Friability	0.41
Hardness	12.45
Transmission ratio (%)	84.30
Cohesion index	789.27

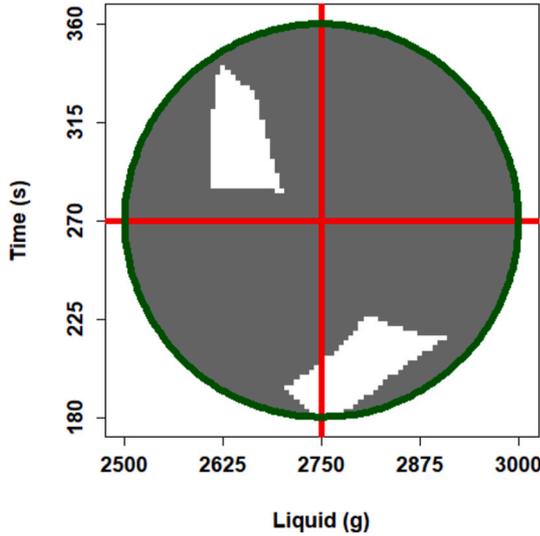
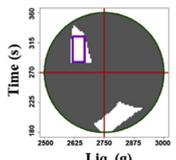


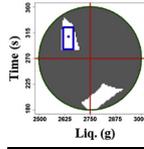
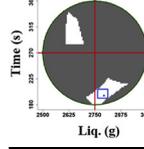
Fig. 11. Design Space constructed by the reliability method ( $\alpha = 5\%$ ).

**Table 3**  
PAIR\_2D with the largest area (Strategy 1).

	Liquid (g)	[2618.75; 2668.75]
	Mixing time (s)	[285.75; 324]
	Nb points	533
	Nb rectangles	17,235
	Time	$T$
	Area	1912.5

**Table 4**

PAIR\_2D containing the optimum or another point and with the largest area (Strategy 1).

		
		Liquid (g)
Mixing time (s)	[285.75; 324]	[191.25; 207]
Nb points	533	533
Nb rectangles	3287	1013
Time	$T$	$T$
Area	1912.5	787.5

least one coordinate is higher than that of IPF. Else, all the points inside the polyhedron (IPI-IPF) are tested to ensure that the corresponding polyhedron is entirely within the DS. Fig. 5 describes the algorithm.

Alternatively, we can use in step 6 the WSP algorithm [16]. This algorithm is generally used to construct Space-Filling Designs whose points will be uniformly distributed in space [16]. In our study, we use WSP algorithm to select  $N$  points from an initial set of  $n$  candidate points (DS), so that the selected points are at least one  $d_{\min}$  away from each other. Thus, we can reduce the number of points contained in the DS, which will reduce the calculation time.

## 2.2. Strategies for selecting a hyper polyhedron

Once all the hyper polyhedra inscribed within the DS are determined, we need to select one. Different strategies may be used depending mainly on the desired criterion.

### 2.2.1. Strategy 1: volume criterion

The volume of all the hyper polyhedra included in the DS is calculated and the one with the largest volume is retained (Fig. 6).

Fig. 6A shows a first possible rectangle (which is a square in this example) and Fig. 6B shows the rectangle with the largest volume (in purple) in this DS.

In some cases, we may choose the largest hyper polyhedron that contains a specific optimum point. However, it is possible that the volume of this hyper polyhedron is smaller. Fig. 7 shows the above example.

Considering the blue point as optimum, we find the largest rectangle that contains this optimum.

By applying this strategy, it may happen that several hyper polyhedra have the same largest volume, especially in two dimensions. This is less probable in higher dimensions. In this case, we may choose among the optimal hyper polyhedra one randomly, or consider another strategy detailed in the next section.

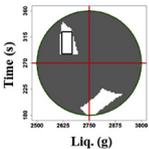
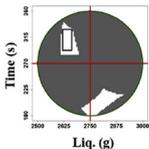
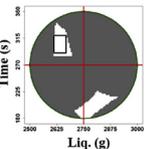
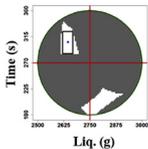
### 2.2.2. Strategy 2: volume and range variation criteria

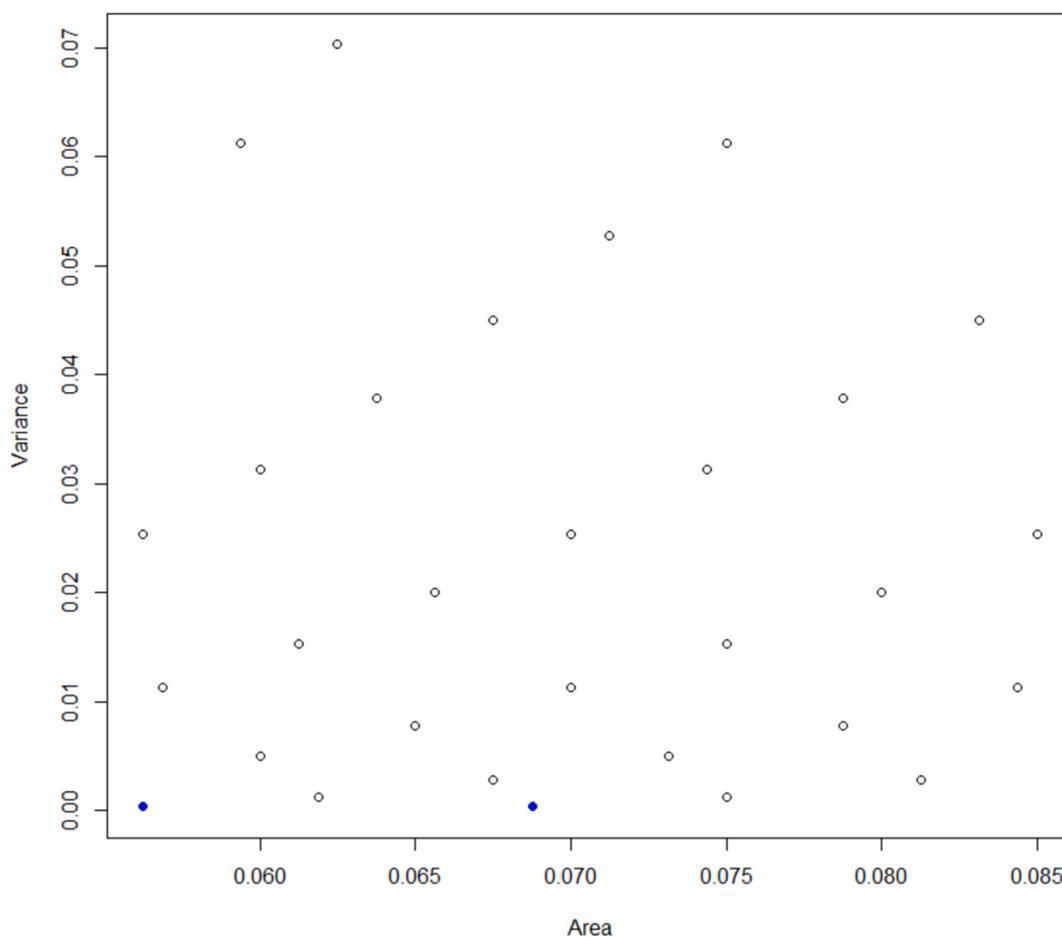
In some cases, volume optimisation may not be satisfactory. Indeed, several hyper polyhedra may have the same largest volume (Fig. 8) or the factors do not have the same importance from a control process point of view. It is up to the experimenter to choose the most suitable one according to its knowledge of the process. To answer this problem, one idea is to account for the length of the edges of the hyper polyhedron which determines the variation range of the different factors. We can then use the weighted variance of the lengths of the edges:

$$\sigma_L^2 = \sum_{i=1}^k (w_i L_i - \bar{L})^2 \text{ with } \bar{L} = \frac{\sum_{i=1}^k w_i L_i}{w_i} \quad (\text{Eq. 1})$$

Where  $w_i$  is the weight of dimension  $i$ ,  $L_i$  is the edge length and  $\bar{L}$  is the

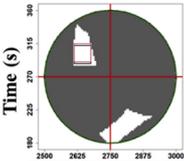
**Table 5**  
PAIR\_2D obtained using the WSP algorithm (Strategy 1).

				
Liquid (g)	[2618.75; 2668.75]	[2618.75; 2662.5]	[2612.5; 2668.75]	[2618.75; 2668.75]
Mixing time (s)	[285.75; 324]	[292.5; 328.5]	[290.25; 319.5]	[285.75; 324]
Nb points	264 (50%)	134 (25%)	36 (less than 10%)	265 (50%) + optimum
Nb rectangles	8738	4148	1240	8650
Time	$t/2$	$t/4$	$t/8$	$t/2$
Area	1912.5	1575	1645.3	1912.5



**Fig. 12.** Edges variance versus area for the top 1% of the largest rectangles.

**Table 6**  
PAIR\_2D obtained with strategy 2.

		Liquid (g)	[2612.5; 2675]	[2612.5; 2675]
		Mixing time (s)	[288; 312.75]	[290.25; 315]
		Nb points	533	
		Nb rectangles	162	
		Time	$T$	
		Area	1546.88	

weighted average of edge lengths. Weights are fixed by the user.

If the experimenter cannot precisely control a factor and prefers to have a larger variation range on one of the factors, we will assign a greater weight to that factor and maximise the weighted variance of the edge lengths (rectangle on the left-hand graph). On the other hand, if the experimenter prefers to have a homogeneous variation range on these factors because it has good control over them, then we will minimize the variance of the edge lengths by assigning identical weights (square on the right-hand graph). Fig. 8 illustrates this strategy in a simple two-dimensional representation.

One can minimize this weighted variance over all the admissible rectangles. In the case of uniform weights, squares will have zero variance even if they have small edges. To avoid such bias, we explore among all the admissible rectangles, only those having large volume, for

**Table 7**  
Studied parameters and DS - 3D case study.

Case	3D <sup>18</sup>		
Inputs	Amount of kaolinite in % - $X_1 = [5; 15]$ Sintering temperature in °C - $X_2 = [900; 1000]$ Sintering time in h - $X_3 = [2; 4]$		
Outputs and its specifications	$Y_1$ : Mechanical strength >22.5 MPa $Y_2$ : Porosity >33.7% $Y_3$ : Permeability >800 L h <sup>-1</sup> m <sup>-2</sup> .bar <sup>-1</sup>		
Response surfaces			
Optimum point	Amount of kaolinite in %	5	
	Sintering temperature in °C	1000	
	Sintering time in h	4	
	Mechanical strength in MPa	40.59	
	Porosity in %	40.86	
	Permeability in L.h <sup>-1</sup> .m <sup>-2</sup> .bar <sup>-1</sup>	1034.88	
Design space			

instance the 1% largest rectangles. Indeed, this strategy would consider not only the volume but also the length of factor variation range.

In the next section, we will apply this PAIR construction method on different DoE cases.

### 3. Experiments

We apply our construction method described in the previous section on different DoE cases. First, we use a 2D case study for easier graphical representation and then, we use different DoE cases (3D and 4D) from the literature. To select a polyhedron, we apply and compare the

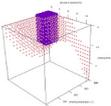
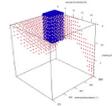
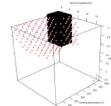
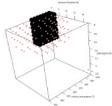
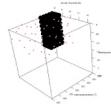
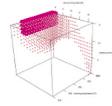
different strategies described in the previous section.

For all cases, we work on a HP personal computer with Intel® Core™ i5-4300 M CPU processor and 4Go of RAM. All experiments were realized with the RStudio interface [17]. In the presentation of results, the units of areas and volumes are ignored.

#### 3.1. Case study in two-dimensional space

We use here a case study introduced in our previous work [14] with 2 inputs and 5 outputs (Table 1). In this study, a second order polynomial model was postulated and a Doehlert design with 12 experiments was

**Table 8**  
Determination of the PAIR with the two strategies – 3D case study.

	Strategy 1			Strategy 2		
	All points	Optimum $Kao = 5\%$ $Temp = 1000\text{ }^\circ\text{C}$ $Time = 4\text{ h}$	WSP algorithm			
						
$Kao$ (%)	[5; 7]	[5; 7]	[5; 8]	[5; 6.5]	[5; 7]	[5; 6]
$Temp$ ( $^\circ\text{C}$ )	[980; 1000]	[980; 1000]	[985; 1000]	[970; 1000]	[980; 1000]	[940; 1000]
$Time$ (h)	[3.3; 4]	[3.3; 4]	[3.4; 4]	[3.4; 4]	[3.3; 4]	[3.7; 4]
Nb points	746	746	317 (50%)	158 (25%)	66 (10%)	746
Nb HP	12,788	351	6181	3856	1449	124
$Time$	$t$	$t$	$t/2$	$t/4$	$t/8$	$t$
Volume	0.112	0.112	0.108	0.108	0.112	0.072

Kao: Amount of kaolinite, Temp: sintering temperature, Time: sintering time, HP: hyper polyhedra.

performed.

From the results of the DoE, the models were used to predict responses throughout the experimental domain. The response surfaces are represented in projections plans (Fig. 9).

In order to satisfy all response specifications, we use desirability function to determine a compromise zone (Fig. 10).

The white zone represents the compromise region in which all the outputs comply with the specifications.

We also determined the optimum point (Table 2) that is represented in blue on Fig. 10.

Then, we constructed a DS by an approach considering the prediction quality, called reliability method with a fixed  $\alpha$  risk of 5% (Fig. 11). This method consists of constructing a confidence interval for each response and for each point in the compromise zone. It considers the accuracy of the predicted values taking into account the variance function of the predicted response in each point [14]. If these confidence interval comply with the specifications for all responses, the point will be kept in the DS.

It has 533 points and is divided into two distinct zones. The fixed discretization step is 0.025 and it has 2 isolated points. The calculation time depends on the computer used. In this case study, we found 17,235 rectangles included in the DS in 8.4 min. Let us note  $t$ , this reference time for the first phase of our methodology.

We then applied the different strategies, detailed in section 2.2.

### 3.1.1. Strategy 1: volume criterion

We look for the rectangle with the largest area included in the DS. The results are presented in Table 3.

We can notice that, for this case study, the optimum point, corresponding to a liquid of 2645 g and a mixing time of 306 s, belongs to PAIR. Among the 17,235 rectangles included in the DS, we are looking for the largest area rectangle that contains the optimum.

Table 4 presents the results for two different points, the optimum maximising the desirability and another point, corresponding to a liquid of 2693.75 g and a mixing time of 195.75 s chosen for various reasons, such as feasibility.

We then use the WSP algorithm to create 3 subsets of points corresponding to 50%, 25% and about 10% of the total number of points present in the DS after removing the isolated points. We also tested a subset of points (50% of the total number of DS points) including the optimum. Table 5 presents the results.

If we consider 50% of the total number of points, we notice that the calculation time is divided by 2. We obtain the same PAIR than that with all DS points. Then, the fewer points in the DS, the shorter the calculation time, but the area of the rectangle is also smaller.

Including the optimum point, we obtain the same PAIR than that with all DS points.

In this first case, we notice that, using the WSP algorithm, the calculation time is shorter than by considering all the points of the DS. Using 50% of the DS points, we find the same PAIR.

### 3.1.2. Strategy 2: volume and variation range criteria

For this case study, strategy 1 showed that the rectangle with the largest area had a greater variation range for the mixing time (input  $X_2$ ) than the liquid (input  $X_1$ ). We can imagine that the experimenter would like to have homogeneous variation range for the factors. For that, we selected 1% of the largest rectangles and considered those having the minimal edges variance by assigning an identical weight to each factor. Fig. 12 shows for the top 1% area rectangles, their edges variance versus their area.

Several rectangles achieve a minimum variance of 0.0003125 (blue points in Fig. 12) and among these, the largest area corresponds to 0.06875 and is achieved by two rectangles shown in Table 6.

The areas are smaller than that with strategy 1 but we obtain more homogeneous variation range on the input parameters, as expected. The choice of the best variation range for input parameters is left to the experimenter.

### 3.2. DoE cases from the literature

We have chosen to work on two case studies, one three-dimensional (3D) and the other four-dimensional (4D).

For each case study, we first present the studied parameters. Then, we construct a DS with the reliability method ( $\alpha = 5\%$ ) and apply the different strategies to determine the PAIR.

#### 3.2.1. 3D case study

The 3D case study concerns the optimisation of phosphate/kaolinite microfiltration membrane for treatment of industrial wastewater [18]. For this case, a second order polynomial model was postulated and a Box-Behnken design with 15 experiments was performed. The studied parameters and the DS constructed by the reliability method with a fixed  $\alpha$  risk of 5% are presented in Table 7.

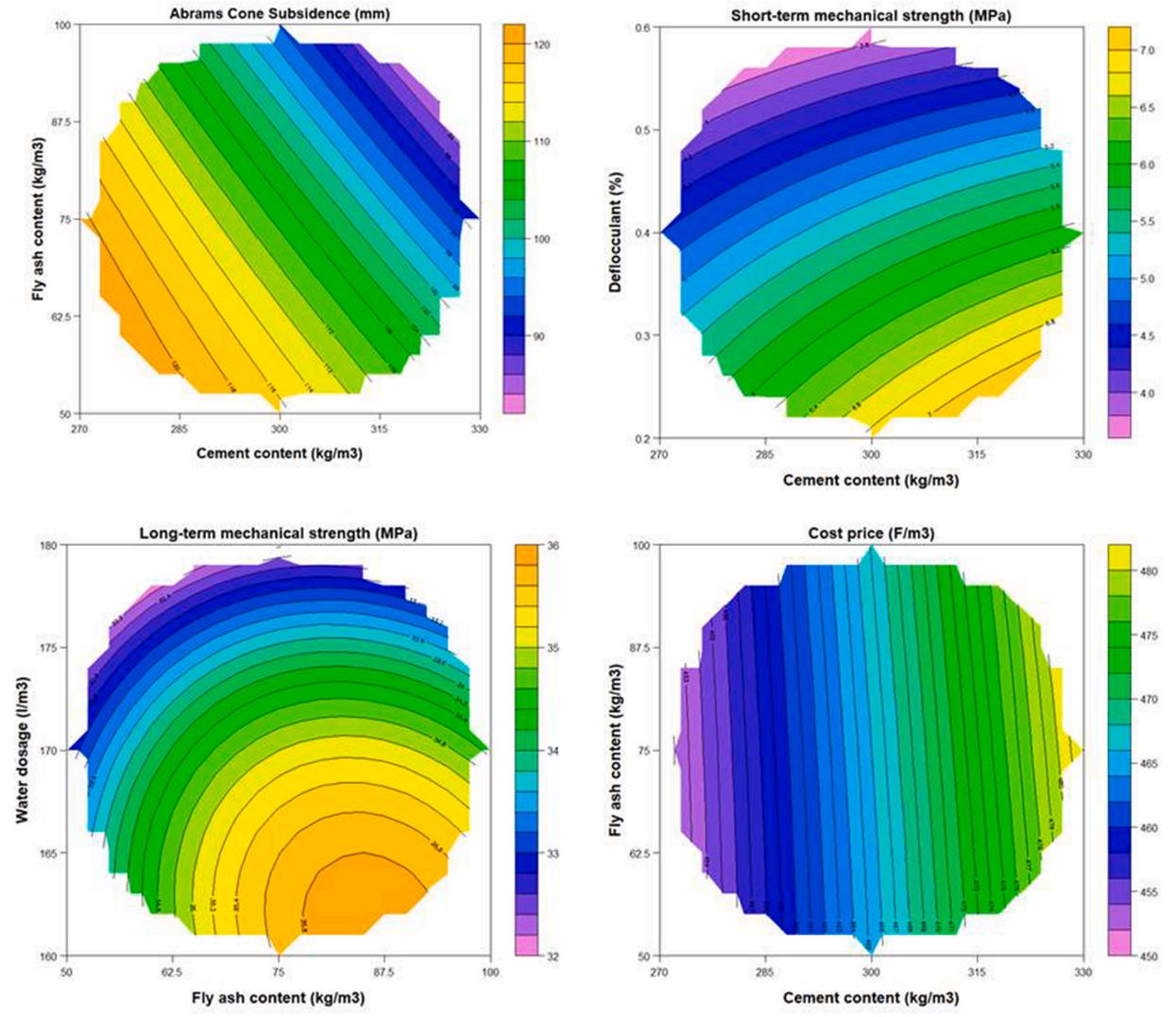
This DS has 746 points. The fixed discretization step is 0.1 and it has 112 isolated points. Table 8 presents the results of the PAIR determination with the different strategies. For this case, we found 12,788 hyper polyhedra included in the DS in 14.24 min, noted  $t$  as reference time.

In strategy 1, we notice that when we work with all points, the PAIR found include the optimum point. When we include the optimum point in the search for the hyper polyhedron with the largest volume, we find

**Table 9**

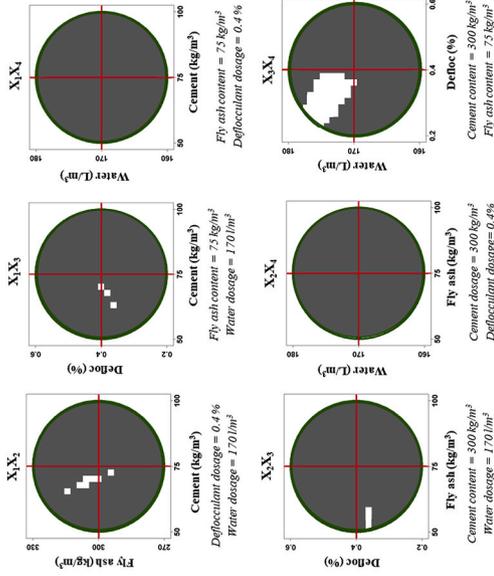
Studied parameters and DS - 4D case study.

Case	4D <sup>19</sup>
Inputs	Cement content in kg/m <sup>3</sup> - X <sub>1</sub> = [270; 330] Fly ash content in kg/m <sup>3</sup> - X <sub>2</sub> = [50; 100] Deflocculant dosage in % - X <sub>3</sub> = [0.2; 0.6] Water dosage in l/m <sup>3</sup> - X <sub>4</sub> = [160; 180]
Outputs and its specifications	Y <sub>1</sub> : 100 mm < Abrams Cone Subsidence < 150 mm Y <sub>2</sub> : Short-term mechanical strength > 5 MPa Y <sub>3</sub> : Long-term mechanical strength > 28 MPa Y <sub>4</sub> : Cost price < 465 F/m <sup>3</sup>
Response surfaces	



288  
80  
0.24  
174  
107.52  
6.05  
32.45  
453.20

Cement content ( $\text{kg}/\text{m}^3$ )  
Fly ash content ( $\text{kg}/\text{m}^3$ )  
Deflocculant dosage (%)  
Water dosage ( $\text{L}/\text{m}^3$ )  
Abrams Cone Subsidence (mm)  
Short-term mechanical strength (MPa)  
Long-term mechanical strength (MPa)  
Cost price ( $\text{F}/\text{m}^3$ )



the same hyper polyhedron than the one found with all DS points.

When we consider a subset of DS points (using WSP algorithm), we notice that the calculation times are shorter than those with all DS points. With 50% and 25% of the DS points, we find almost the same hyper polyhedron as when considering all DS points (the volume is slightly smaller). Then, when we considered 10% of the DS points, the hyper polyhedron found is the same than the one found with all DS points.

In strategy 2, we can imagine that the sintering temperature is a critical factor for which the experimenter would like to have a larger variation range. From the 12,788 hyper polyhedra, we selected 1% of the largest hyper polyhedra. We maximised the weighted variance of edge lengths by assigning a weight of 3 to this critical factor and a weight of 1 for the other factors. We obtain a hyper polyhedron with a smaller volume, but as desired, the acceptable variation range of the sintering temperature is greater than that obtained with strategy 1.

### 3.2.2. 4D case study

The 4D case study deals with the formulation of hydraulic concrete [19]. In this study, a second order polynomial model was postulated and a Doehlert design with 23 experiments was performed. Table 9 presents the studied parameters and some sectional plans of DS constructed by the reliability method with a fixed  $\alpha$  risk of 5%.

This DS has 2426 points. The fixed discretization step is 0.1 and it has 486 isolated points. Table 10 presents the results of the PAIR determination with the different strategies. We found a total of 9765 hyper polyhedra included in the DS in 38.5 min, noted  $t$  as reference time.

In strategy 1, we notice that the optimum point does not belong to the found PAIR. Nevertheless, the coordinates of the optimum point do not allow finding a hyper polyhedron. When we consider a subset of DS points (using the WSP algorithm), the calculation times are shorter than that with all DS points. With 50% of the DS points, we found the same PAIR than that with all DS points. With about 10% of the DS points, we found a hyper polyhedron with a smaller volume than that with all DS points.

In strategy 2, we can imagine that the experimenter would like to have more homogeneous variation range on the factors. For this, we have selected 1% of the larger hyper polyhedra. Then, we minimized the variance of the edge lengths weighted with an identical weight for each of the factors. We can see that the volume of this hyper polyhedron is smaller than that of strategy 1 but, as expected, the variation ranges on the factors are more homogeneous.

## 4. Conclusion

The final objective of the Quality by Design approach is to define a DS. Regardless of the construction method chosen, it has any geometric shape. Therefore, the acceptable variation range of an input parameter may depend on the value of another input parameter. To solve this issue, we developed a two-step algorithm that allows the determination of acceptable variation range on each input so that their respective variation domains are independent. The methodology consists in searching for all hyper polyhedra inscribed in any  $k$ -dimensional space and selecting one according to different strategies.

We applied this method to different 2D, 3D and 4D case studies. The calculation times are very acceptable. In addition, the search for the hyper polyhedra inscribed in the DS can be parallelized to save even more calculation time. The selection of a hyper polyhedron depends on the experimental context. We proposed two strategies and compared them with each other.

To conclude, the notion of strategy remains somewhat subjective and with elements that must be determined by the experimenter.

### Author statement

Diane Manzon: Conceptualization, Methodology, Software, Writing -

**Table 10**

Determination of the PAIR with the two strategies – 4D case study

CD: cement dosage, FC: fly ash content, DD: deflocculant dosage, WD: water dosage, HP: hyper polyhedra.

	Strategy 1		Strategy 2		
	All points	Optimum <i>CD</i> = 288 kg/m <sup>3</sup> <i>FC</i> = 80 kg/m <sup>3</sup> <i>DD</i> = 0.24% <i>WD</i> = 174 l/m <sup>3</sup>	WSP algorithm		
<i>CD</i> (kg/m <sup>3</sup> )	[291; 303]	-	[291; 303]	[297; 303]	[291; 297]
<i>FC</i> (kg/m <sup>3</sup> )	[65; 80]	-	[65; 80]	[60; 80]	[65; 75]
<i>DD</i> (%)	[0.28; 0.32]	-	[0.28; 0.32]	[0.28; 0.32]	[0.28; 0.32]
<i>WD</i> (l/m <sup>3</sup> )	[174; 175]	-	[174; 175]	[174; 175]	[173; 175]
Nb points	2426	-	967 (50%)	237 (about 10%)	2426
Nb HP	9765	-	4881	1111	98
Time	<i>t</i>	-	<i>t</i> /2.5	<i>t</i> /8	<i>t</i>
Volume	7.2	-	7.2	4.8	4.8

Original Draft, Badih Ghattas: Conceptualization, Methodology, Software, Writing - Review & Editing, Magalie Claeys-bruno: Writing - Review & Editing, Supervision, Sophie Declomesnil: Funding Acquisition, Christophe Carité: Funding Acquisition, Michelle Sergent: Writing - Review & Editing, Supervision.

#### 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.

#### References

- [1] J. Woodcock, The concept of pharmaceutical quality, *Am. Pharmaceut. Rev.* 7 (2004) 10–15.
- [2] ICH Q8 (R2) Pharmaceutical Development, European Medicines Agency, 2018. Published September 17, <https://www.ema.europa.eu/en/ich-q8-r2-pharmaceutical-development>. (Accessed 4 September 2022).
- [3] S. Sarasija, S.K.A. Roy, Quality by design: an overview, *Indian Drugs* 52 (2015) 5–11, <https://doi.org/10.53879/id.52.02.p0005>.
- [4] K. Pramod, M.A. Tahir, N.A. Charoo, S.H. Ansari, J. Ali, Pharmaceutical product development: a quality by design approach, *Int J Pharm Investig* 6 (3) (2016) 129–138, <https://doi.org/10.4103/2230-973X.187350>.
- [5] S. Kalyan, A. Parle, Quality by design: changing outlook of pharmaceutical development, *Int. J. Pharma Sci. Res.* 10 (9) (2019) 4100–4108, [https://doi.org/10.13040/IJPSR.0975-8232.10\(9\).4100-08](https://doi.org/10.13040/IJPSR.0975-8232.10(9).4100-08).
- [6] N.S. Ganesh, C. Shah, V. Chandu, Quality by Design approach: regulatory requirement for development of quality pharmaceuticals, *Pharma Sci Monit- Int J Pharm Sci.* 9 (2) (2018) 312–340.
- [7] S. Beg, M. Rahman, S. Swain, Quality by design applications in pharmaceutical product development, *Pharma Focus Asia* 38 (2020) 1–5.
- [8] H.B. Grangeia, C. Silva, S.P. Simões, M.S. Reis, Quality by design in pharmaceutical manufacturing: a systematic review of current status, challenges and future perspectives, *Eur. J. Pharm. Biopharm.* 147 (2020) 19–37, <https://doi.org/10.1016/j.ejpb.2019.12.007>.
- [9] J. Lepore, J. Spavins, PQLI Design Space, *J Pharm Innov* 3 (2008) 79–87, <https://doi.org/10.1007/s12247-008-9034-2>.
- [10] F. Boukouvala, F. Muzzio, M. Ierapetritou, Design Space of pharmaceutical processes using data-driven-based methods, *J Pharm Innov* 5 (3) (2010) 119–137, <https://doi.org/10.1007/s12247-010-9086-y>.
- [11] L.A. Sarabia, M.C. Ortiz, Response surface methodology, in: *Vol Comprehensive Chemometrics*, Elsevier., 2009, pp. 345–390.
- [12] I.M. Fukuda, C.F.F. Pinto, C. dos S. Moreira, A.M. Saviano, F.R. Lourenço, Design of experiments (DoE) applied to pharmaceutical and analytical quality by design (QbD), *Braz J Pharm Sci* 54 (2018), <https://doi.org/10.1590/s2175-97902018000001006>.
- [13] G. Derringer, R. Suich, Simultaneous optimization of several response variables, *J. Qual. Technol.* 12 (4) (1980) 214–219, <https://doi.org/10.1080/00224065.1980.11980968>.
- [14] D. Manzon, M. Claeys-Bruno, S. Declomesnil, C. Carité, M. Sergent, Quality by design: comparison of design space construction methods in the case of design of experiments, *Chemometr. Intell. Lab. Syst.* 200 (2020), 104002, <https://doi.org/10.1016/j.chemolab.2020.104002>.
- [15] MODDE® pro 12, sartorius stedim data analytics AB, Umeå, Sweden, <https://umetrics.com/>.
- [16] J. Santiago, M. Claeys-Bruno, M. Sergent, Construction of space-filling designs using WSP algorithm for high dimensional spaces, *Chemometr. Intell. Lab. Syst.* 113 (2012), <https://doi.org/10.1016/j.chemolab.2011.06.003>.
- [17] RStudio Team, RStudio, Integrated Development for R. RStudio, Inc., Boston, MA, 2016. <http://www.rstudio.com/>.
- [18] A. Belgada, F.Z. Charik, B. Achiou, et al., Optimization of phosphate/kaolinite microfiltration membrane using Box–Behnken design for treatment of industrial wastewater, *J. Environ. Chem. Eng.* 9 (1) (2021), 104972, <https://doi.org/10.1016/j.jece.2020.104972>.
- [19] F. Louvet, L. Delplanque, Les Plans d'expériences : Une Approche Pragmatique et Illustrée = Design of Experiments : The French Touch, *Expérimentique.*, 2005.