



HAL
open science

Space-filling designs for mixtures

C. Gomes, M. Claeys-Bruno, M. Sergent

► **To cite this version:**

C. Gomes, M. Claeys-Bruno, M. Sergent. Space-filling designs for mixtures. *Chemometrics and Intelligent Laboratory Systems*, 2018, 174, pp.111-127. 10.1016/j.chemolab.2018.01.013 . hal-02079830

HAL Id: hal-02079830

<https://amu.hal.science/hal-02079830>

Submitted on 11 Apr 2019

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.

Space-filling designs for mixtures

C. Gomes^a, M. Claeys-Bruno^{b,*}, M. Sergent^b

^a L'OREAL Research & Innovation, 93601, Aulnay-sous-bois, France

^b Aix Marseille Univ, Univ Avignon, CNRS, IRD, IMBE, Marseille, France

ARTICLE INFO

Keywords:

Formulation

Space-filling designs

Mixture designs

ABSTRACT

Uniform experimental designs (Space Filling Designs) are now widely used with independent variables, particularly in numerical simulation. A lot of construction methods have been developed to uniformly cover the domain of variables, but they remain rarely applied for the study of mixtures. In this article, we propose methods to build space-filling designs for mixtures that can be used to modelize complex phenomena in formulation. Various algorithms like Kennard and Stone, WSP, Strauss and Dmax algorithms, to construct these designs are detailed and compared with respect to uniformity criteria.

1. Introduction

In many industries products are obtained by mixing two or more components or ingredients, and the properties of the products depend on the proportion of each ingredient in the mixture. Issues related to mixtures can thus be defined as the study of the influence of the relative amounts of several components (called *proportions*) on the manifestations of a physico-chemical phenomenon (called *responses*).

In the designs of experiments for mixtures, the factors studied are therefore the proportions of each ingredient. These proportions have two important properties: their values are dimensionless numbers - a value of 0.1 has the same signification whatever the nature of the component - and the sum of all the proportions is equal to 1, which indicates that the components of the mixture cannot be independently controlled.

We therefore have the following constraints:

$$x_i \geq 0 \text{ and } \sum_{i=1}^q x_i = 1 \text{ for } i = 1, 2, \dots, q$$

where q is the number of components and x_i is the proportion of component i contained in the mixture.

As a result of these constraints, the experimental space is a regular polyhedron, or simplex, with q vertices in a space with $q-1$ dimensions.

If the ingredients must comply with upper or lower limits, the space of interest will be a sub-space of the initial simplex.

When the aim is to model a response and to study its behavior, classical models of mixtures (Scheffé model, etc.) and classical designs such as the Scheffé designs [1] [2] or D-optimal designs are commonly used to estimate the coefficients of the model. However, whatever the degree of the model, classical designs position the points at the periphery of the experimental space and are potentially ill-adapted to deal with irregularities inside the space. To limit these risks, we propose to explore the whole experimental space by uniformly distributing points within it.

Numerous methods have been developed to distribute points uniformly throughout a space when all the parameters are independently controlled (Space-Filling Designs [3–5]). Among these methods, we have selected a small number based on different principles: some rely on points selection (Kennard and Stone [6], and Wootton, Sergent, Phan-Tan-Luu (WSP) algorithms [7]), others use the principle of repulsion between points (maximal entropy designs [8] or Strauss designs [9]), while a new algorithm relies on step-by-step construction. All of these methods can uniformly distribute points in a space for independent variables, but some modifications must be made if we want to apply them to studies of mixtures and to take the specificities of these studies into account. In this article, we will start by presenting the intrinsic criteria through which the uniformity of a distribution of points can be assessed, then we will provide details on the construction algorithms used and the modifications made. Finally, we will compare the relative advantages and disadvantages of the different methods used to construct uniform designs for mixtures.

* Corresponding author.

E-mail address: m.claeys-bruno@univ-amu.fr (M. Claeys-Bruno).

2. Intrinsic uniformity criteria

To assess the quality of the structure of a set of points making up a uniform experimental design, quantitative criteria must be used. Among the available criteria, we retained criteria involving calculation of the Euclidean distances. These criteria appeared relevant to us to assess the quality of the distribution of points in a space with D dimensions.

- The **Mindist** criterion [10–12] defines the smallest distance between any two points in a distribution:

$$Mindist = \min_{x_i \in X} \min_{x_j, j \neq i \in X} dist(x_i, x_j) \quad (1)$$

where $X = \{x_1, x_2, \dots, x_N\} \subset [0, 1]^D$ is a set of N points in D dimensions.

A high **Mindist** value is synonymous with a good distribution of the points throughout the variables space. In contrast, distributions of points

with a low **Mindist** value have at least two very closely-spaced points that could be qualified as a cluster. When clusters are present, the distribution of points in the space is no longer ideal in terms of uniformity.

- The **MeanMin** criterion defines the mean of the minimal distances, and provides information on the minimal distances between points:

$$MeanMin = \frac{1}{N} \sum_{i=1}^N \min_{j \neq i} dist(x_i, x_j) \quad (2)$$

A set of N points could be termed uniform if the **MeanMin** value is large and close to the **Mindist** value.

- The **coverage** criterion [13] measures how much of the space is covered by the points and characterizes the homogeneity of the distribution of the points:

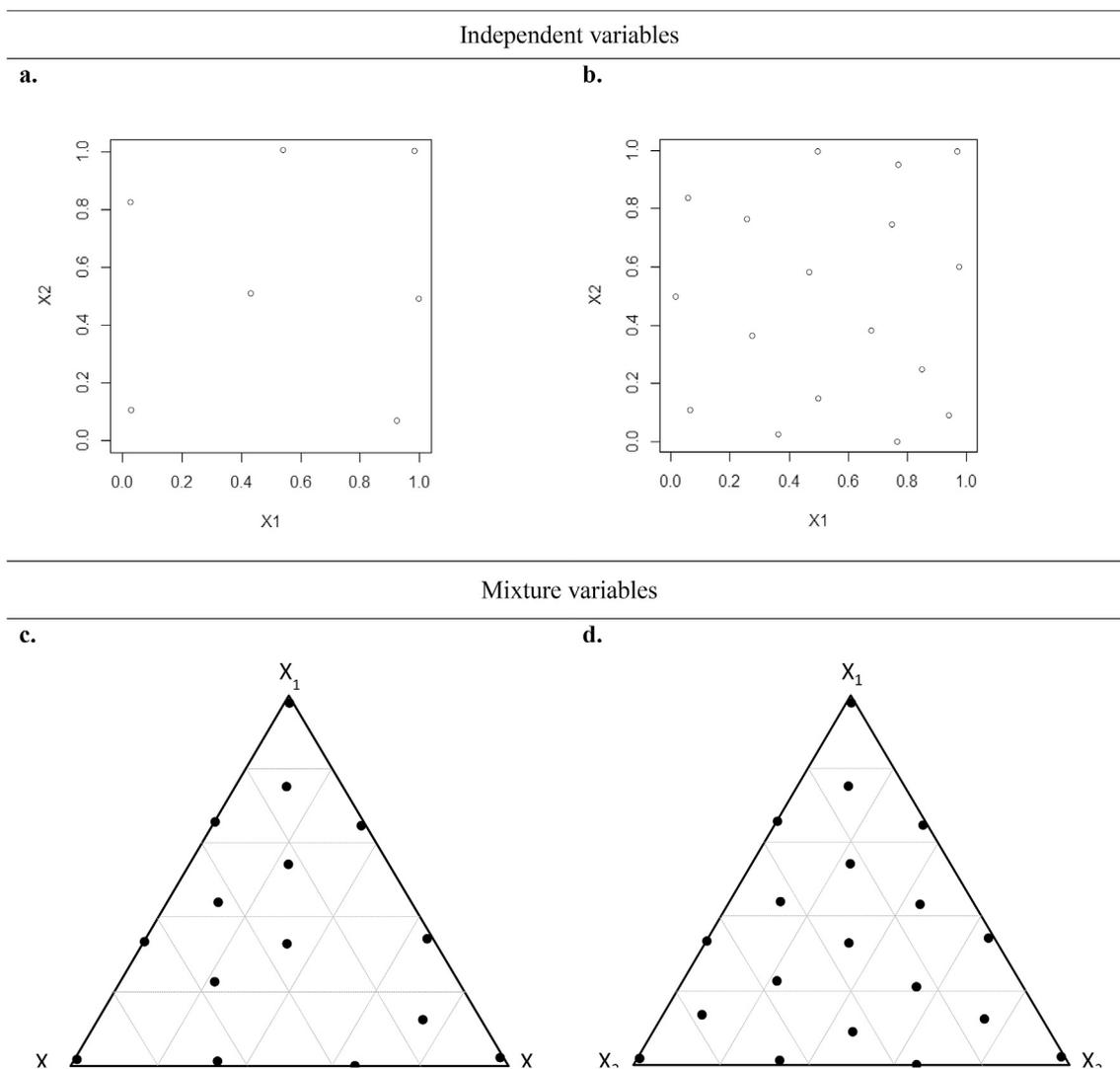


Fig. 1. Kennard and Stone algorithm. a. and b. for independent variables with $N = 7$ points and $N = 17$ points respectively. Solutions for mixture variables c. with $N = 15$ points, d. with $N = 20$ points.

$$Cov = \frac{1}{\bar{\gamma}} \left(\frac{1}{N} \sum_{i=1}^N (\gamma_i - \bar{\gamma})^2 \right)^{1/2} \quad (3)$$

where, $\gamma_i = \min_{j \neq i} \text{dist}(x_i, x_j)$ and $\bar{\gamma} = \frac{1}{N} \sum_{i=1}^N \gamma_i$.

This value provides information on the homogeneity of the minimal distances between the points in the design and is a reflection of the dispersal of the minimal distances. Thus, when a set of points has a high *coverage* value, the distribution of the minimal distances is highly heterogeneous, indicating that some areas in the space are more densely filled than others.

- The **Standard Deviation** of the minimal distances (independent of the number of points in the ensemble).

$$\sigma = \left(\frac{1}{N} \sum_{i=1}^N (\gamma_i - \bar{\gamma})^2 \right)^{\frac{1}{2}} \quad (4)$$

with $\gamma_i = \min_{j \neq i} \text{dist}(x_i, x_j)$ and $\bar{\gamma} = \frac{1}{N} \sum_{i=1}^N \gamma_i$

- The **Audze-Eglais** criterion (AE) [14] is a repulsion criterion, which assimilates the points in a distribution to charged particles. AE is calculated on the basis of the square of the Euclidean distance between points, which is inversely proportional to the magnitude of the forces of repulsion:

$$AE = \sum_{i=1}^{N-1} \sum_{j=i+1}^N \frac{1}{\text{dist}(x_i, x_j)^2} \quad (5)$$

where $\text{dist}(x_i, x_j)$ is the Euclidean distance between any two points x_i and x_j .

This criterion must be minimized to make the potential energy of the structure as low as possible.

In addition to these classical criteria, Borkowski and Piepel [15] added three other distance-based criteria specific to the study of mixtures, with or without constraints. To guarantee that all the components are considered identically, i.e., independently of their range of variation, these criteria are calculated by first normalizing the components relative to their range of variation ($\mathbf{b} - \mathbf{a}$), where \mathbf{a} and \mathbf{b} are the lower and upper bounds, respectively.

These criteria involve a set of M virtual points $\mathbf{u} \{u_1, u_2, \dots, u_M\}$ in the space of interest \mathcal{R} from which we define a minimal Euclidean distance between a point u_m in \mathcal{R} and the points x_j from the design for mixture X with N points

$$d(\mathbf{u}, X) = \min_j \text{dist}_E \left(\frac{u_m - a}{b - a}, \frac{x_j - a}{b - a} \right) \quad (6)$$

where $m = 1, 2, \dots, M$ and $j = 1, 2, \dots, N$.

A uniform design for a mixture will be considered of good quality if the points are uniformly distributed in the space \mathcal{R} . The criteria proposed are as follows:

- The square root of the mean of the square of the minimal distances:

$$rmsd(X) = \sqrt{\frac{1}{M} \sum_{m=1}^M (d(u_m, X))^2} \quad (7)$$

where $m = 1, 2, \dots, M$, as above.

- The mean of the minimal distances:

$$ad(X) = \frac{1}{M} \sum_{m=1}^M d(u_m, X) \quad (8)$$

- The maximum of the minimal distances:

$$md(X) = \max_m d(u_m, X) \quad (9)$$

Analysis of the various criteria provide information on the quality of points distribution in terms of uniformity. For example, a low *Mindist* value and a high *MeanMin* value indicate that at least two points are very closely-spaced, but that the other points are well distributed in the variable space. If the *Mindist* and *MeanMin* values are equivalent, and both low, the smallest distances between points are equivalent, i.e., all the points are closely-spaced and form a single cluster. In contrast, if a distribution presents high values for *Mindist* and *MeanMin*, it will not be uniform. In addition, low values for the *rmsd(X)*, *ad(X)* and *md(X)* criteria indicate that the points in the space \mathcal{R} are close to the X plane. In contrast, a high value for *md(X)* indicates that there exists a region in \mathcal{R} where the points are not close to X .

3. Space-filling designs for mixtures

Numerous algorithms can be used to construct uniform experimental designs for independent variables. We will present some of these algorithms grouped on the basis of their construction principles and we will highlight the parameters that must be optimized, and their appropriateness for use in analyzing variables in mixtures.

3.1. Selection algorithms

Selection algorithms aim to extract a sub-group of N points from a set of Nc candidate points in D dimensions. We have selected the following two selection algorithms:

3.1.1. Kennard and Stone (KS) algorithm [6]

At each iteration, this sequential selection algorithm selects the point furthest from those already selected.

The algorithm can be described as follows:

Kennard and Stone algorithm

- **Step 1:** Consider a group of N_c candidate points randomly generated in a space with D dimensions.
- **Step 2:** Calculate the matrix of the Euclidean distances for the set of candidate points:

$$d_{ij} = \|x_i - x_j\| = \sqrt{\sum_{r=1}^D (x_{ir} - x_{jr})^2} = \text{Euclidean distance between points } i \text{ and } j$$

- **Step 3:** Choose points I and J such that: $d_{IJ} = \max(d_{ij})$

UNTIL $N = N_c$ where $N = \text{Desired number of points}$

Step 4:

- Calculate the distances for the $(N_c - N)$ remaining points relative to the N selected points and retain their minimal value:

$$\Delta_i(N) = \min\{d_{i1}, d_{i2}, \dots, d_{iN}\}$$

Where $\Delta_i(N)$, is the distance from the candidate point i - not yet included in the set of selected points - to the nearest point among the selected points.

- For the $(N+1)^{\text{th}}$ point in the matrix, choose among the $(N_c - N)$ remaining candidate points the point for which:

$$\Delta(N + 1) = \max\{\Delta_i(N)\}$$

so that the new point is the furthest from the previously-selected points.

END

Two solutions obtained using the KS algorithm on a candidate 100-points set in two dimensions are presented Fig. 1a and b. If we select $N = 7$ points, the KS algorithm produces a non-uniform final solution, whereas if N is increased to 17 points, the points are better distributed throughout the space.

This example illustrates the main weakness of this algorithm: it produces more or less uniform solutions depending on the number of points to be selected, N .

To adapt this algorithm to the study of mixtures, the candidate points must belong to the simplex delimiting the variables space for mixtures.

From Fig. 1c et 1d, we can draw the same conclusions as for the KS algorithm in a space of independent variables.

3.1.2. Wootton, sergent, phan-tan-luu (WSP) algorithm [7,16–20]

This selection algorithm will select N points from an initial set of N_c points. Point selected will be spaced at least a pre-determined distance of d_{\min} from the other points already included in the design. Preliminary studies have been showed that the type of the initial distribution and the choice of the initial point O have no impact on the intrinsic quality of the design after the selection algorithm. Therefore, we consider for the candidate points a random distribution, and the initial point has been chosen to be the nearest of the gravity center of the simplex.

The WSP algorithm can be summarized as follows:

WSP algorithm

- **Step 1:** Consider a set of N_C candidate points randomly generated in a space with D dimensions
- **Step 2:** Calculate the matrix of the Euclidean distances for the set of N_C candidate points:

$$d_{ij} = \|x_i - x_j\| = \sqrt{\sum_{r=1}^D (x_{ir} - x_{jr})^2} = \text{distance between points } i \text{ and } j$$

- **Step 3:** Choose an initial point O in the set of candidate points near to the gravity center of the simplex
- **Step 4:** Set a d_{min} value

UNTIL all the candidate points have been selected or eliminated

Step 5:

- Eliminate from the candidate set all the points I such that $d_{OI} < d_{min}$
- Place the point O in the final sub-group
- Eliminate the point O from the set of candidate points and replace it with its closest neighbor from among the remaining points

END

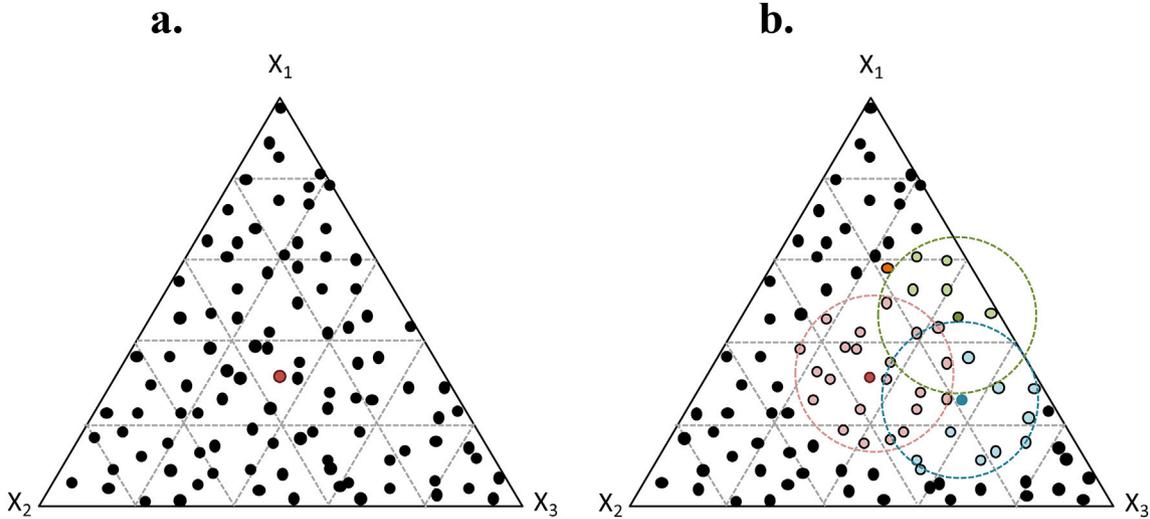


Fig. 2. WSP algorithm applied to mixtures. **a.** Candidate points, $N = 100$. The red point corresponds to the initial point O . **b.** Step-by-step approach of WSP algorithm. Step 1: the red colored points inside the red circle ($d_{OI} < d_{min}$) are eliminated. Step 2: the closest point (blue point) is chosen to become the new point O and points inside the blue circle are eliminated. Step 3: the closest point (green point) is chosen and points for which the distance $d_{OI} < d_{min}$ are eliminated. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

The number of points selected by the WSP algorithm is directly linked to the d_{min} value. Indeed, the smaller the d_{min} value, the larger the number of points selected. If the number of points to be selected is chosen beforehand, the d_{min} value will be adjusted by iteration until the desired number of points (or close to this number) have been selected from the candidate set. The main characteristic of the WSP algorithm is that it sets a minimal distance between all the selected points. This criterion

guarantees a uniform distribution of the final sub-group, whatever the value of N . The WSP algorithm guarantees that solutions will uniformly cover the experimental space (Fig. 2).

Like with the KS algorithm, the WSP algorithm can be applied to mixtures if the set of candidate points is generated in the simplex of mixture variables (Fig. 3).

It seems interesting to precise that the geometric criteria can be

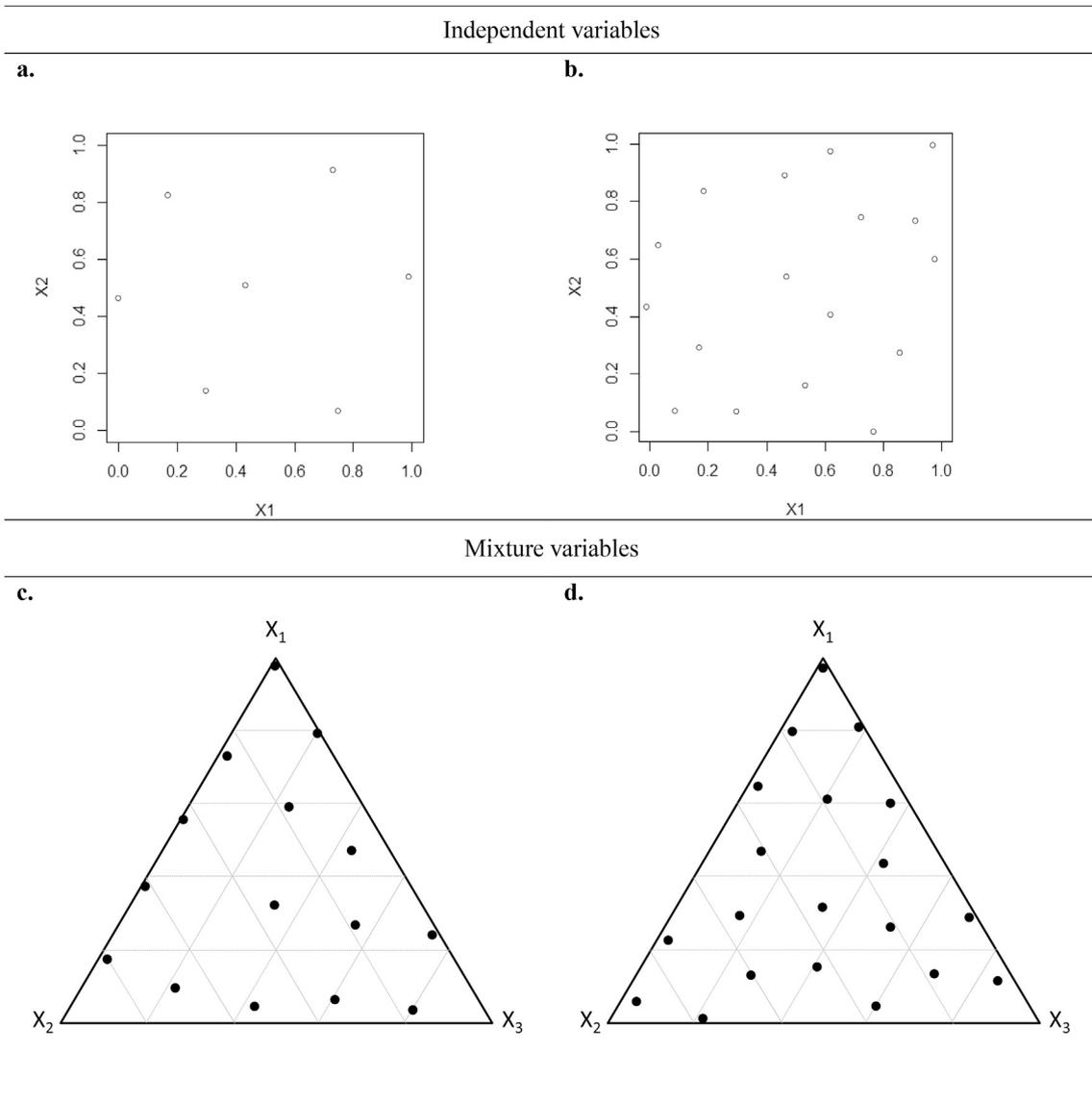


Fig. 3. WSP algorithm. **a.** and **b.** for independent variables with $N = 7$ points and $N = 17$ points respectively. Solutions for mixture variables **c.** with $N = 15$ points, **d.** with $N = 20$ points.

modified in order to take into account experimental constraints. In this case, it is not the closest neighbor of the remaining points which is selected [Fig. 2](#) but another close neighbor which presents more favorable criteria. This selection can be done using desirability functions [\[21\]](#).

3.2. Repulsion algorithms

The WSP and KS algorithms are selection algorithms which require a set of candidate points from which the considered algorithm extracts a

sub-group. Designs can also be constructed starting with a randomly-selected set of points which will then be reorganized according to a principle of repulsion. We will investigate two algorithms: the Strauss algorithm and the maximal entropy algorithm (Dmax).

3.2.1. Strauss algorithm [\[9\]](#)

This algorithm starts with a distribution of N random points then applies the Metropolis-Hastings algorithm (MH) [\[22,23\]](#) which first proposes to a state change differing only by a single point. The second step consists in accepting or rejecting this change of state.

Strauss algorithm

- **Step 1:** Set the values of N , of the radius R (radius of interaction) and the number of iterations NMC
- **Step 2:** Generate a random matrix \mathbf{V} with N points and D dimensions

FOR I varying between 1 and NMC

FOR k varying between 1 and N

Step 3:

- Randomly choose a point x_i
- Replace the point x_i in the matrix \mathbf{V} by the randomly-chosen x_j point, to obtain the matrix $\mathbf{V1}$
- Accept or reject the replacement of point x_i by point x_j based on the Metropolis-Hastings algorithm by calculating the probability:

$$a(V, V1) = \min \left(1, \frac{\pi(V1)}{\pi(V)} \right) \text{ with } \pi(V) \text{ and } \pi(V1), \text{ probability functions}$$

IF $a(V, V1) >$ random probability

THEN the change is accepted, the matrix \mathbf{V} is replaced by the matrix $\mathbf{V1}$

END

END

Selecting the values of the parameters R and NMC for a desired number of points N constitutes the main difficulty with this algorithm. For a set number of N points, selection of a value of R that is too low results in empty zones; an overly-large value produces clusters of points (Fig. 4a and b.).

The number of NMC iterations for the Strauss algorithm must be large to allow the Metropolis-Hastings calculation to converge. In general, in repulsion algorithms, the larger N the greater the required number of iterations. This relation implies long calculation times when the number of dimensions increases. This algorithm also relies on numerous stochastic processes: selection of the initial matrix, selection of the candidate point for substitution and the probability of acceptance or rejection of the change in position. When the algorithm is re-applied with identical parameter values, these random processes can lead to different final solutions which can be of variable quality.

To adapt the Strauss algorithm to the study of mixtures, the step generating the random matrix must be modified. The random matrix must be in the mixture space, and at each step, the point x_i must also be in the simplex. Use of the Strauss algorithm for mixtures presents the same difficulties as noted above for independent variables. Indeed, the value of

R selected can lead to a poor distribution of points in the space for mixtures with empty zones (Fig. 4a) or clusters of points (Fig. 4b).

It should be noted that if the Strauss algorithm is re-applied with the same parameter values, solutions of very different quality may be obtained.

3.2.2. Maximal entropy designs (d_{max})

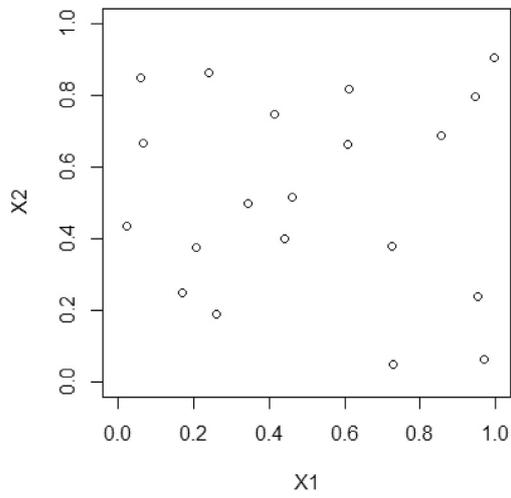
The algorithm to produce maximal entropy designs fills the space by distributing points according to a spatial correlation matrix. Shewry and Wynn [8] showed that the entropy H depends on $\ln(\det(C(X)))$ where $C(X)$ is the matrix for spatial correlation, constructed according to a model with a spherical variogram $\gamma(d)$ which depends of the range parameter a .

$$\begin{cases} \gamma(d) = 1.5 \frac{d}{a} - 0.5 \left(\frac{d}{a} \right)^3, & d < a \\ \gamma(d) = 1, & d \geq a \end{cases} \quad (10)$$

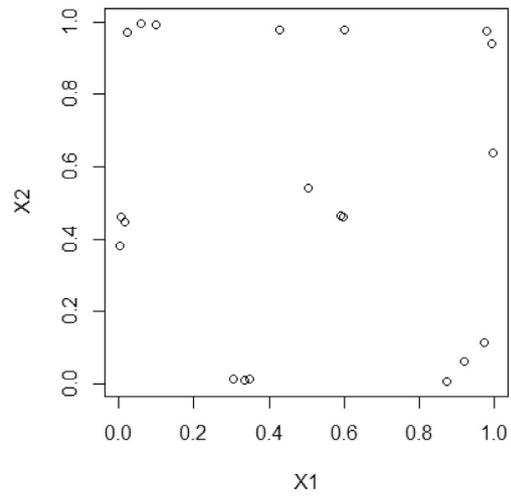
Thus, the entropy H will be maximal when the determinant for the matrix $C(X)$ is large.

Independent variables

a. $R=0.1$ and $NMC=500$

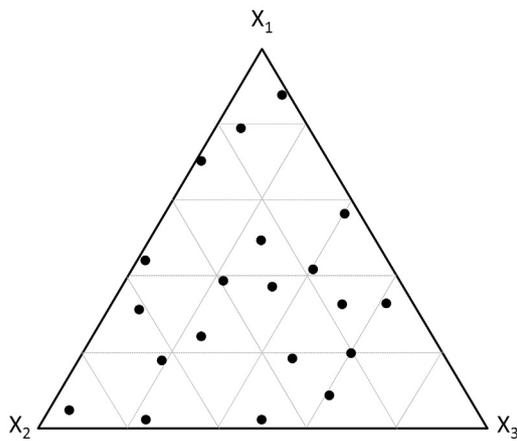


b. $R=0.5$ and $NMC=500$



Mixture variables

c. $R=0.15$, $NMC = 500$



d. $R=0.40$, $NMC = 500$

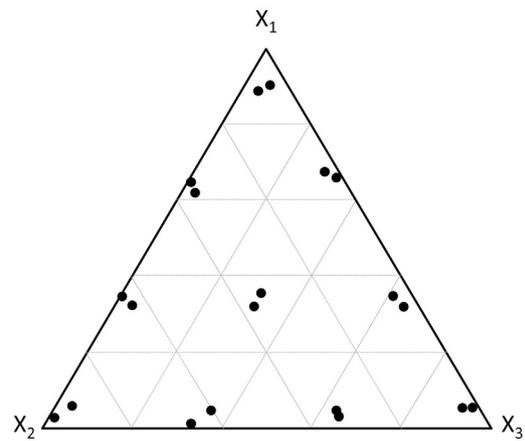


Fig. 4. Strauss algorithm. a. and b. Solutions for independent variables with $N=20$ points. c. and d. Solutions for mixture variables with $N=20$ points for different values of R .

Algorithm for Dmax designs

- **Step 1:** Set the values N , a (extent of the variogram) and the maximal number of iterations NMC
- **Step 2:** Generate a random matrix \mathbf{V} with N points in D dimensions
- **Step 3:** Calculate the spatial correlation matrix $C(V)$ such that:

$$C(V) = \sigma^2 \begin{bmatrix} 1 & & \rho_{ij} \\ & \ddots & \\ \rho_{ij} & & 1 \end{bmatrix}$$

where the values ρ_{ij} depend on the distance d_{ij} between points i and j in the spherical variogram $\gamma(d)$:

$$\begin{cases} \rho_{ii} = 1 \\ \rho_{ij} = 1 - \gamma(d_{ij}) , \text{ if } d_{ij} < a \\ \rho_{ij} = 0, \text{ if } d_{ij} \geq a \end{cases}$$

- **Step 4:** Calculate the determinant of the $C(V)$ matrix

UNTIL NMC is reached

Step 5:

- Randomly choose a point x_i
 - Replace the point x_i in the matrix \mathbf{V} by the randomly-chosen point x_j , to produce the matrix $\mathbf{V1}$
 - Calculate the spatial correlation matrix $C(V1)$ and its determinant
- IF $\det(C(V1)) > \det(C(V))$
 THEN the change is accepted, the matrix \mathbf{V} is replaced by the matrix $\mathbf{V1}$

END

Like with the Strauss algorithm, when constructing Dmax designs parameters must be set. Indeed, the use of a model with a spherical variogram requires the distance a to be set to define the values of the off-diagonal terms in the spatial correlation matrix $C(X)$. The distance a must be carefully chosen to avoid “skewing” the algorithm. Indeed, a very low value of the distance a leads to numerous null off-diagonal terms, characterized by a very large determinant, and thus a large entropy, but does not guaranteeing a uniform distribution of the points throughout the variable space.

To apply this Dmax algorithm to mixtures, the random matrix \mathbf{V} must belong to the mixture space of the mixture, and for each step the point x_i must also belong to the simplex.

3.3. Step-by-Step algorithm

The **Step-by-Step (Sbs)** algorithm was initially proposed by Franco [24] to construct uniform experimental designs in the independent variables space. This algorithm randomly chooses an initial point and then, by iteration, adds points spaced a distance R from the points already

present in the design. Whatever the number of points N making up the experimental design, this construction method ensures that all the points will be placed at the same minimal distance from each other.

To use Franco's algorithm, two parameters must be set: the number of points N and the value of the radius R . However, by setting both the radius R and the number of points N , it remains possible that the radius could be ill-adapted to the number of points and that, consequently, the space could be poorly filled. Indeed, if the radius R is too small, the N points retained by the algorithm will not cover the whole space (Fig. 5) and for a radius R that is too large, the algorithm cannot distribute N points in a space with D dimensions.

To overcome this disadvantage, we will retain the main idea from this algorithm in the construction of the design, but we will only set the distance R between a point and its nearest neighbor in a space with D dimensions; the number of points N will therefore be implicitly determined by this value R and by the random selections made at each iteration.

The algorithm that we propose is therefore as follows:

SbS algorithm

- **Step 1:** Set: the radius R , $iter$, $iter_max$, $Rejet_max$
- **Step 2:** Generate an initial point x_0 in the space $[0,1]^D$ and add it to design \mathbf{X}
- **Step 3:** Generate a point x_j on the sphere with center x_0 and radius R

WHILE ($x_j \notin [0,1]^D$)

| $number\ of\ rejections = number\ of\ rejections + 1$

END

Step 4: The point x_j is selected and added to the \mathbf{X} design

WHILE ($number\ of\ rejections < Rejet_max$)

$iter = 0$

Step 5: Randomly choose a point in design \mathbf{X} , noted x_k

WHILE ($iter < iter_max$)

Step 6: Generate a point x_m on the sphere with center x_k and radius R

IF ($x_m \notin [0,1]^D$)

| $iter = iter + 1$

| $number\ of\ rejections = number\ of\ rejections + 1$

ELSE

| Calculate the distance d between the point x_m and all the points in design \mathbf{X}

| **IF** ($\min(d) \geq R$)

| | **Step 7:** x_m is added to the design \mathbf{X}

| **ELSE**

| | $iter = iter + 1$

| | $number\ of\ rejections = number\ of\ rejections + 1$

| **END**

END

END

END

As the algorithm makes use of random processes, for a single value of R , several solutions with equivalent properties could be obtained. It is important that the maximum number of rejections be sufficiently large to guarantee that no other point could be added to the space of interest. Thus, we know that the solutions obtained effectively fill the whole of the space.

To apply this SbS algorithm to the variables in a mixture, it must be modified, but the difficulty resides in the generation of the points to be added. Indeed, the initial point must be part of the simplex - this condition poses no particular difficulty - but the other points to be added must be at a distance R from the points already selected and must be part of this simplex. When we wish to add a point x_m at a defined distance R from the point x_k this comes down to placing it in a hypersphere of radius R and center x_k . It is easy to generate a point on a hypersphere, but the addition of the mixtures constraint is equivalent to limiting the possible points to the intersection of the hypersphere with the simplex specific to mixtures. This mathematical equation cannot be generalized, whatever the number q of components. We propose to avoid this difficulty by applying the approach described by Cornell [25] where the points from a plane with $(q-1)$ independent variables are transformed to a mixture

plane with q components surrounding a point of interest. To do this, both the point of interest $x_0 = (x_{01}, x_{02}, \dots, x_{0q})$ and the variations in the components surrounding it must be defined to determine the ellipsoid of interest. By applying this method with the SbS algorithm, a point x_m can be generated in the mixtures plane around the point of interest x_k (chosen at steps 3 or 6 of the algorithm) by setting a variation step equal to R for all the components. This approach is equivalent to generating a hypersphere of interest (Fig. 6).

The mathematical transformation applied to the classical independent variables matrix (\mathbf{W}) to produce the mixture design (\mathbf{X}) around the point x_0 is as follows:

$$\mathbf{X} = c\mathbf{W}\mathbf{T}'\mathbf{H} + x_0 \quad (11)$$

where $c = \frac{\rho^*}{\sqrt{q-1}}$ and $\rho^* = \min_i \rho_i$, with $i = 1, \dots, \rho_i = x_{0i} \left\{ \frac{1}{h_i^2} + \frac{1}{a-h_i^2} \right\}^{1/2}$

where $a = \sum_{i=1}^q h_i^2$

\mathbf{H} is the diagonal matrix containing the variations h_i around the point of interest $\mathbf{H} = \text{diag}(h_1, h_2, \dots, h_q)$.

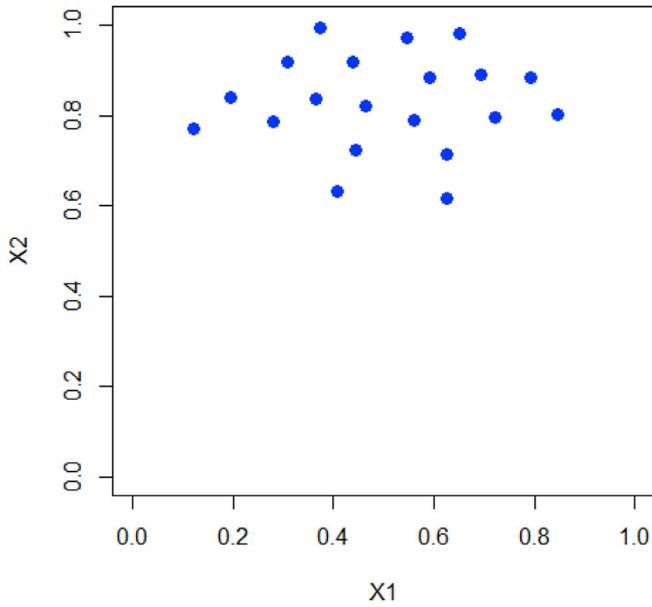


Fig. 5. Construction of SbS designs in a space with orthogonal variables by setting a number of points N equal to 20 and the radius $R = 0.1$. The radius is ill-adapted to the number of points and therefore the points do not fill the whole space.

The matrix T_1 has a dimension of $q \times (q - 1)$ can be constructed as follows. If we denote t_{ij} the element on the i -th line and in the j -th column in the matrix T_1 before its normalization, the matrix T_1 can be calculated by applying the following relations:

$$\begin{array}{cccccc}
 t_{11} = -h_2 & t_{12} = -h_1 h_3 & t_{13} = -h_1 h_4 & t_{14} = -h_1 h_5 & \dots & \\
 t_{21} = h_1 & t_{22} = -h_2 h_3 & t_{23} = -h_2 h_4 & t_{24} = -h_2 h_5 & \dots & \\
 t_{i1} = 0, i > 2 & t_{32} = h_1^2 + h_2^2 & t_{33} = -h_3 h_4 & t_{34} = -h_3 h_5 & \dots & \\
 t_{i2} = 0, i > 3 & t_{43} = h_1^2 + h_2^2 + h_3^2 & t_{44} = -h_4 h_5 & & & \\
 & t_{i3} = 0, i > 4 & t_{54} = h_1^2 + h_2^2 + h_3^2 + h_4^2 & & & \\
 & & t_{64} = 0, i > 5 & & & \dots
 \end{array}$$

To construct the matrix T_1 each term t_{ij} is normalized by the term: $(\sum_{k=1}^q t_{ik}^2)^{1/2}$ where $k = 1, \dots, q - 1$.

To apply the SbS algorithm to studies of mixtures, the initial point x_0 must be chosen in the simplex, and at each step while generating the points x_j and x_m , the Cornell transformation must be applied.

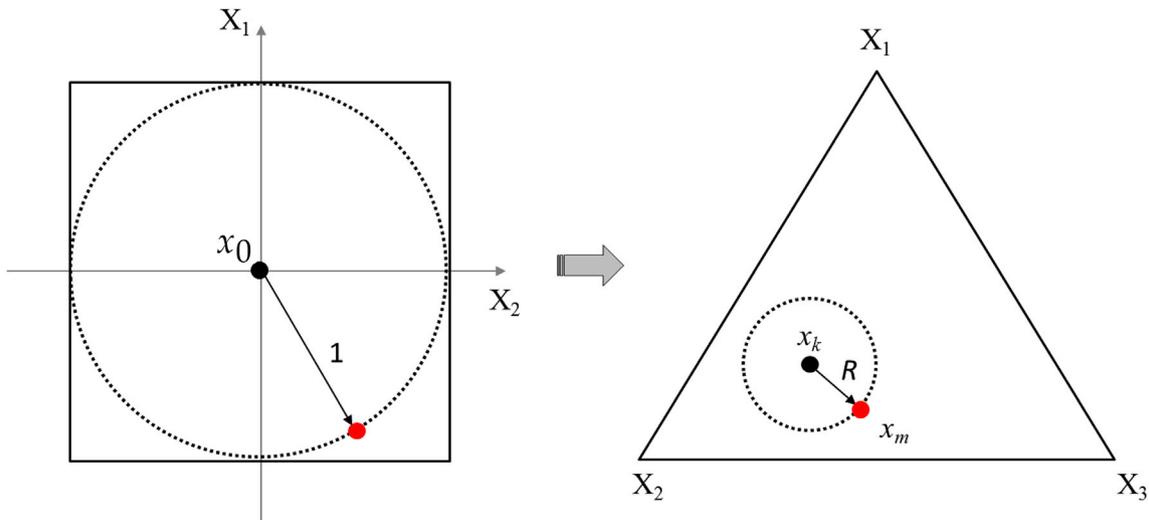


Fig. 6. Illustration of the Cornell transformation adapted to the SbS algorithm which makes it possible to transform point x_m generated in a space with two orthogonal variables onto a 3-component mixture plane.

Like with independent variables, for the same distance R , non-identical solutions sharing equivalent properties can be obtained from different iterations with the same parameters (Fig. 7).

If the number of points N is fixed, the distance R allowing us to meet this objective must be determined. To do this, we repeated the SbS algorithm several times to discover how N changes as a function of R . For example, if a 3-components mixture is considered and the distance R varies between 0.1 and 0.38, repeating the algorithm 100 times for each value of R reveals that the solutions obtained with the SbS algorithm contain between 6 and 88 points (Fig. 8).

From this curve, we can determine the value of R producing a design with N points. Thus, to obtain around 15 points, R should be set to 0.25.

4. Comparison of space-filling designs for mixtures

The various methods available to construct uniform designs of experiments for mixtures all have advantages and disadvantages. To compare them, we chose to study the intrinsic criteria presented in part 2.

For this study, we used designs with 10 and 20 points for a 3-components mixture. Whatever the value of N , the KS and WSP selection algorithms were applied to the same set of candidate points with $N_c = 2500$ points. For the repulsion algorithms, Strauss and Dmax, the same initial random matrix was chosen for a given value of N .

4.1. Experimental designs with $N = 10$ points

Before comparing the quantitative criteria, the distributions of points obtained by applying each of the algorithms tested are graphically represented Fig. 9.

4.1.1. Distribution of points

• Kennard and Stone designs

The KS algorithm places the points in the periphery of the space and leads to a design equivalent to a Scheffé simplex-lattice (Fig. 9a.). In this design, the interior of the experimental space contains very few experimental points.

• WSP designs

For the WSP algorithm (Fig. 9b.), the center of gravity or its nearest neighbor was selected as the initial point; selection of a different initial

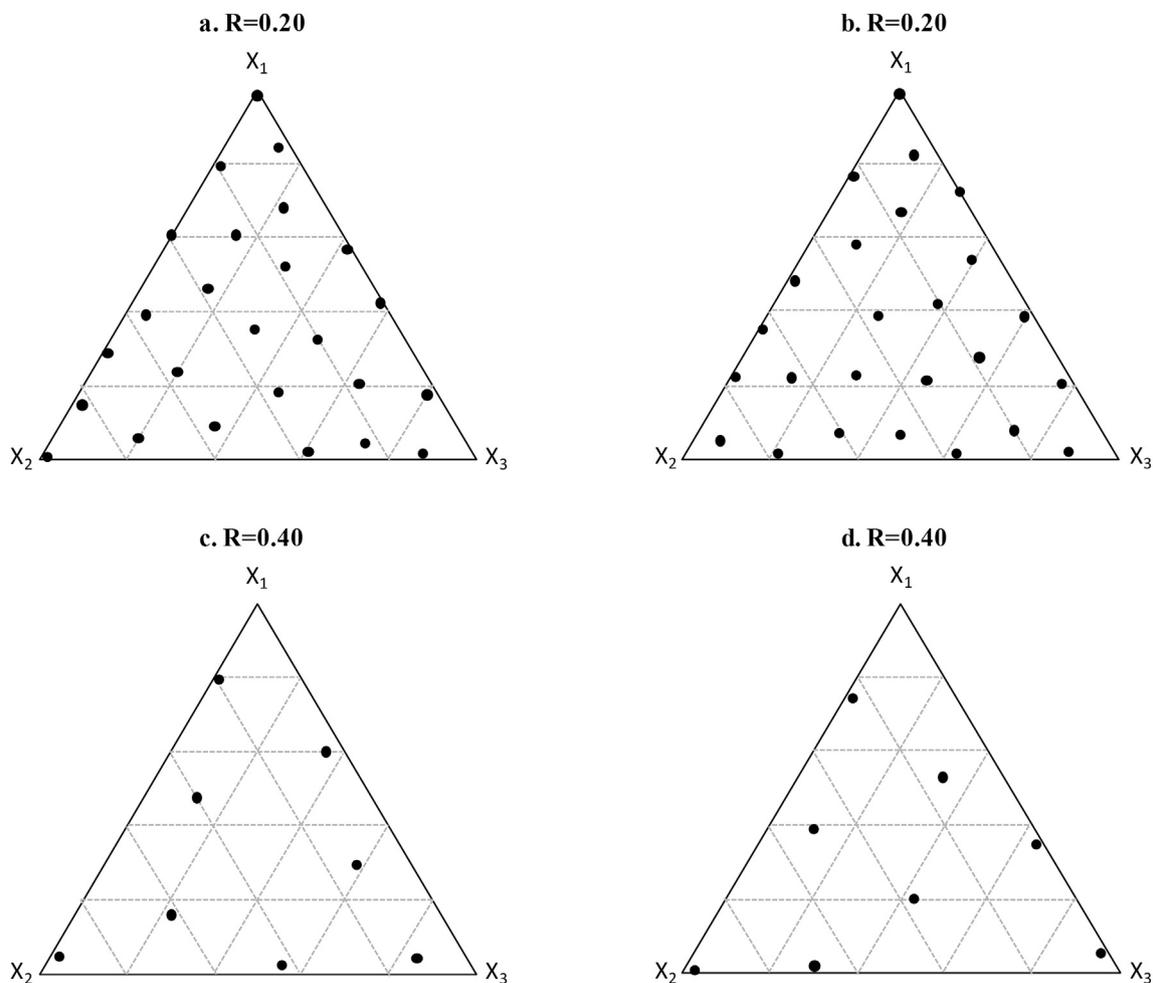


Fig. 7. SbS solutions for 3-component mixtures. For $R = 0.20$, SbS designs contain 25 points (a. and b.) and for $R = 0.40$ the designs contain 8 points (c. and d.).

point could have produced a different solution.

• Strauss designs

For the purposes of this study, the number of iterations NMC was set to 500, which was considered sufficiently large for the algorithm to converge while also providing a reasonable calculation time. As the relation between N and R is unknown, we proceeded by iteration,

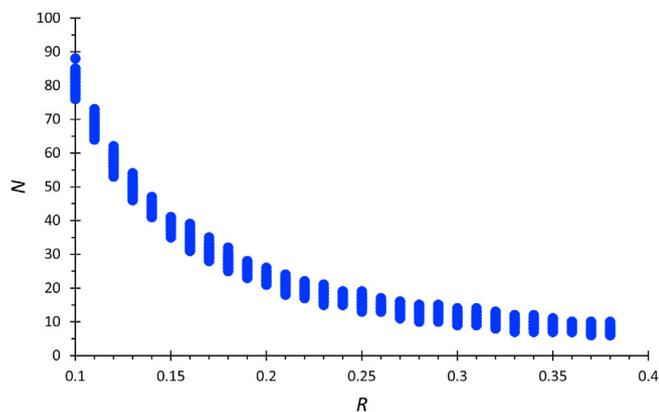


Fig. 8. Relationship between the number of points N and the distance R used when constructing SbS experiment designs for mixtures with three components. For each value of R , the algorithm was repeated 100 times.

varying R from 0.2 to 0.5 and repeating the algorithm several times. Fig. 9c shows extreme cases from among the different solutions obtained.

These data suggest that $R = 0.2$ is too low, as numerous empty zones are present in the design. If R is increased to 0.3, the points are better distributed throughout the space. Increasing the distance R to 0.4 produces two types of distributions: in the first solution, the points are uniformly distributed, whereas in the second solution (represented Fig. 9c in the right)) two points are found very close together. This phenomenon illustrates the fact that the radius R can be too large. This conclusion is confirmed with $R = 0.5$. In these conditions, all of the iterations led to clusters of points. Faced with this phenomenon, we conclude that $R = 0.4$ is a limit value and we recommend that this value be refined between 0.3 and 0.4 to allow the Strauss algorithm to produce the most uniform solution possible.

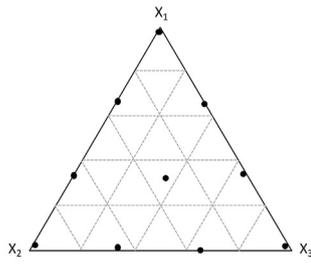
• Dmax designs

For designs with maximal entropy, we varied the limit a of the spherical variogram so as to measure the impact of this parameter on the uniformity of the distribution. For each solution, the determinant of the correlation matrix that we are seeking to maximize was calculated.

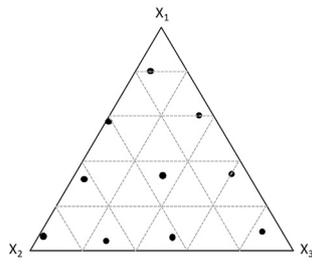
The distributions produced (Fig. 9d.) were observed to vary considerably depending on the limit, in addition, the determinant criterion is probably not sufficient to characterize uniformity. Indeed, a value close to 1 for the determinant does not systematically guarantee a uniform distribution.

Selection algorithms with $N_c=2500$ candidate points

a.



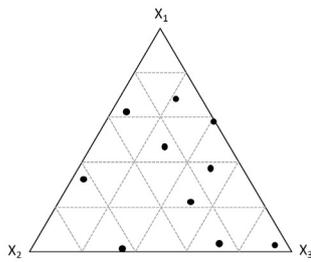
b.



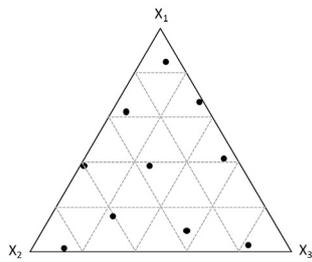
Repulsion algorithms

c.

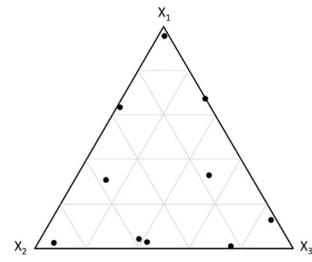
$R = 0.2$



$R = 0.3$

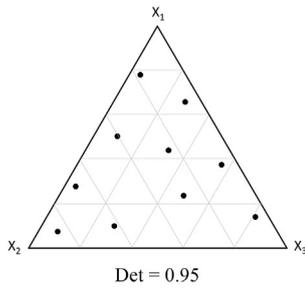


$R = 0.4$

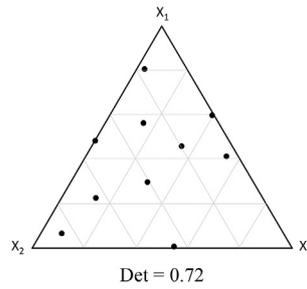


d.

$a=0.3$



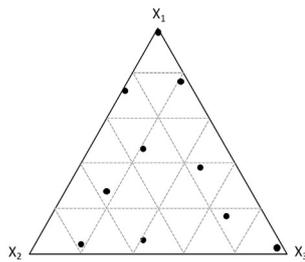
$a=0.4$



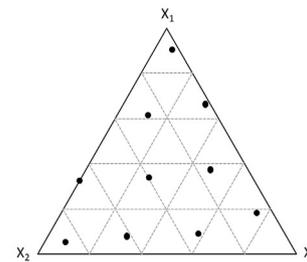
Step-by-Step algorithm

e.

$R = 0.32$



$R = 0.34$



$R = 0.36$

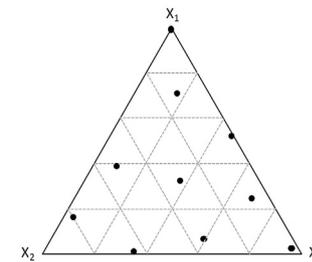


Fig. 9. Solution with $N = 10$ points: a. Obtained from a candidate matrix $N_c = 2500$ points upon application of KS algorithm, b. Obtained from a candidate matrix $N_c = 2500$ points upon application of WSP algorithm, c. Obtained using Strauss algorithm with 500 iterations, d. for Dmax designs with different limit values, e. Obtained using Sbs algorithm for different values of R .

Table 1

Assessment criteria for designs with $N = 10$ points obtained by applying the different algorithms to 3-components mixture.

	Mindist	Maxdist	MeanMin	Coverage	Standard Deviation	AE	rmsd	ad	md
KS	0.432	0.461	0.443	0.022	0.010	117.15	0.170	0.160	0.269
	0.432	0.447	0.440	0.012	0.005	118.47	0.168	0.157	0.265
WSP	0.350	0.358	0.352	0.006	0.002	167.07	0.141	0.130	0.256
	0.340	0.350	0.342	0.010	0.003	170.79	0.146	0.135	0.356
Strauss $R = 0.4$	0.349	0.389	0.372	0.036	0.014	143.22	0.157	0.146	0.290
	0.334	0.402	0.372	0.060	0.022	153.20	0.143	0.133	0.254
	0.294	0.451	0.359	0.128	0.046	143.32	0.16	0.150	0.288
	0.352	0.388	0.369	0.031	0.012	137.52	0.161	0.148	0.310
	0.223	0.445	0.350	0.203	0.071	166.82	0.146	0.134	0.263
Dmax $a = 0.3$	0.221	0.456	0.287	0.224	0.064	259.08	0.143	0.130	0.372
	0.215	0.364	0.261	0.154	0.040	244.21	0.152	0.136	0.440
	0.263	0.39	0.315	0.146	0.046	204.34	0.140	0.128	0.317
	0.232	0.391	0.293	0.165	0.049	234.78	0.147	0.133	0.429
	0.232	0.379	0.278	0.146	0.041	268.25	0.153	0.133	0.380
SbS $R = 0.34$	0.340	0.340	0.340	0.000	0.000	179.20	0.133	0.124	0.247
	0.340	0.340	0.340	0.000	0.000	184.08	0.136	0.126	0.303
	0.340	0.340	0.340	0.000	0.000	172.49	0.148	0.137	0.284
	0.340	0.340	0.340	0.000	0.000	166.03	0.148	0.137	0.294
	0.340	0.340	0.340	0.000	0.000	170.24	0.148	0.136	0.290

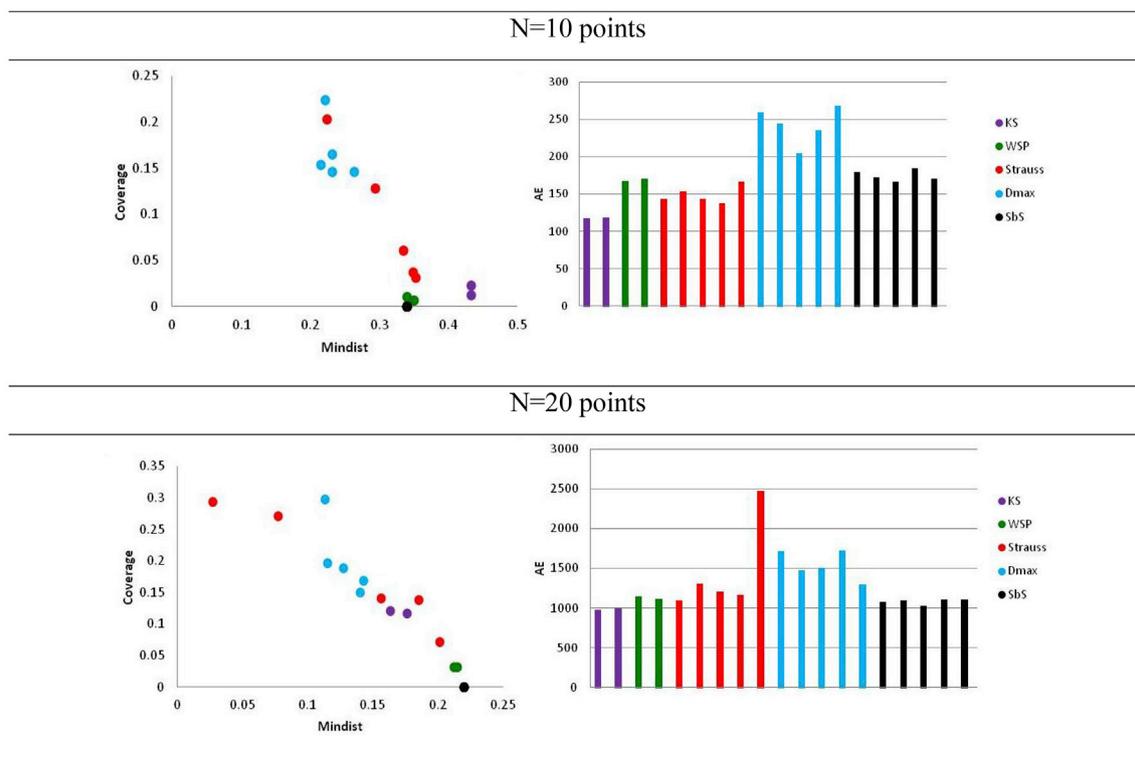


Fig. 10. Comparison of the intrinsic quality of designs for mixtures with $N = 10$ points and $N = 20$ points. Left: Mindist values; Right: AE values for the designs produced by the different algorithms.

• **SbS designs**

When constructing SbS designs, the value of R (radius of the sphere) was varied between 0.32 and 0.36, as recommended by the data presented in Fig. 8. Some of the solutions obtained for different R values are shown in Fig. 9e.

In these constructions, we obtained solutions for which all the points are located at the same distance from neighboring points. For a given number of points N and a carefully selected R value, repeated application of the SbS algorithm produces solutions with a different structure, but in which distribution of the points is guaranteed to be uniform.

4.1.2. *Calculation of the intrinsic criteria*

For this study with $N = 10$ points, all of the criteria presented in part 2 were calculated for each of the designs. For the selection algorithms (KS and WSP), the algorithm's initial point was varied. For the repulsion algorithms, the radius R (Strauss) or the limit a (Dmax) was defined and five repetitions were performed. For the SbS algorithm, the radius R was defined and five repetitions were also performed. The values for the criteria are listed in Table 1.

To compare the performance of the different designs in terms of uniformity, we propose to use the graphical representation of coverage = f (Mindist) values, knowing that the ideal zone is defined by

Table 2

Assessment criteria for designs with $N = 20$ points obtained by applying the different algorithms to 3-components mixtures.

	Mindist	Maxdist	MeanMin	Coverage	Standard Deviation	AE	rmsd	ad	md
KS	0.163	0.274	0.243	0.120	0.029	975.91	0.099	0.093	0.168
	0.176	0.271	0.242	0.117	0.028	994.89	0.098	0.091	0.176
WSP	0.212	0.239	0.217	0.032	0.007	1143.16	0.095	0.088	0.207
	0.215	0.245	0.219	0.032	0.007	1111.76	0.096	0.089	0.207
Strauss	0.185	0.311	0.220	0.138	0.030	1092.28	0.103	0.094	0.297
	0.077	0.312	0.210	0.271	0.057	1301.73	0.098	0.090	0.210
	0.156	0.295	0.203	0.140	0.028	1206.28	0.103	0.093	0.326
	0.201	0.266	0.214	0.072	0.015	1164.37	0.097	0.090	0.235
	0.027	0.235	0.191	0.293	0.056	2467.48	0.101	0.092	0.276
Dmax	0.115	0.227	0.171	0.197	0.034	1718.43	0.103	0.092	0.352
	0.113	0.316	0.183	0.297	0.054	1471.48	0.100	0.092	0.208
	0.143	0.230	0.175	0.169	0.029	1509.39	0.108	0.096	0.363
	0.127	0.250	0.173	0.188	0.032	1721.08	0.105	0.094	0.258
	0.140	0.254	0.195	0.150	0.029	1290.74	0.101	0.092	0.340
SbS	0.220	0.220	0.220	0.000	0.000	1072.41	0.098	0.091	0.205
	0.220	0.220	0.220	0.000	0.000	1093.00	0.095	0.088	0.197
	0.220	0.220	0.220	0.000	0.000	1030.67	0.101	0.093	0.212
	0.220	0.220	0.220	0.000	0.000	1107.79	0.095	0.088	0.197
	0.220	0.220	0.220	0.000	0.000	1108.12	0.097	0.090	0.216

Table 3

Characteristics of the different algorithms used to construct uniform designs for mixtures.

Algorithm	Parameters to be defined	Initial matrix	Stochastic process	Stop iteration of the algorithm
KS	- N - initial point	Candidate matrix with N_c points	NO	When N points are contained in the design
WSP	- d_{min} - initial point	Candidate matrix with N_c points	NO	When all the points in the candidate matrix have been considered
Strauss	- R (radius of interaction) - number of iterations (NMC)	Random matrix with N points	YES	Number of iterations
Dmax	- a (extent) - number of iterations	Random matrix with N points	YES	Number of iterations
SbS	- R (radius of the sphere)	A random point	YES	Number of rejections

high *Mindist* values and low *coverage* values. The AE criterion can also be considered, it also relies on the calculation of Euclidean distances which we seek to minimize to limit the potential energy of the structure as far as possible.

This analysis reveals that the designs for mixtures produced by the Strauss and Dmax algorithms present greater heterogeneity of minimal distances than the designs produced by the KS, WSP and SbS designs, where the minimal distances were practically equal for all repetitions (Fig. 10 and Table 1).

Representation of the AE criterion as a histogram reveals slightly lower values for the KS design, whereas designs produced by the WSP, Strauss and SbS algorithms present larger, equivalent values. Higher values still were obtained for designs with maximal entropy, which suggests that these distributions were less uniform (Fig. 10).

According to the Borkowski and Piepel *ad* criterion, the mean of the minimal distances (Table 1), only the KS designs are distinguished from the other designs due to larger values. All the other designs present similar *ad* values. In this case, the *ad* criterion cannot discriminate between the different designs.

4.2. Designs with $N = 20$ points

In this comparison (Table 2), the *Mindist* values were low and different from the *MeanMin* values for the Strauss, Dmax and KS designs, indicating that at least two points were very closely-spaced. In addition, these designs present very variable *coverage* values from one repetition to the next, which confirms the strong heterogeneity of the minimal distances.

The WSP and SbS designs were characterized by high *Mindist* and *MeanMin* values and very low *coverage* values, indicating a uniform distribution (Fig. 10). In comparison to designs with $N = 10$ points, the KS designs with $N = 20$ points appear to be of lower quality with higher *coverage* values.

For the AE criterion (Fig. 10), the KS, WSP and SbS algorithms give the lowest values. For the Strauss designs, this criterion shows a poor repeatability of the algorithm, leading to very different solutions, with variations in the AE value between 1092.28 and 2467.48. However, using the *Mindist* criterion we demonstrated that two designs had very closely-spaced points (repetition 2, *Mindist* = 0.076 and repetition 5, *Mindist* = 0.027) whereas the AE criterion only flagged design 5. It therefore appears that this criterion, although very informative, cannot distinguish between designs with very closely-spaced points as it is a criterion assessing overall uniformity.

4.3. Comparison of the algorithms for 3 components

These algorithms were modifications of algorithms developed to produce designs with independent variables. Table 3 summarizes the main characteristics of these algorithms in terms of the parameters to adjust, criteria for stopping iteration, etc.

This table shows differences in the number of parameters to be adjusted which can sometimes require optimization. In addition, some algorithms use stochastic processes, potentially resulting in heterogeneity in the designs, and which require the algorithm to be repeated for the same value of N and any other parameters to be defined. To compare the performance of these algorithms, we considered the criteria for designs produced in the case of a 3-components mixture. The *Mindist*, *MeanMin* and *coverage* criteria can be used to characterize the uniformity of the distribution by detecting the presence of clusters or gaps. The AE criterion is informative, but remains an overall assessment criterion which can discriminate between a very poorly-conditioned design and a relatively poorly-conditioned design, but cannot identify an almost uniform

Table 4

Assessment criteria for designs with $N = 20$ to 100 points obtained by applying the different algorithms to 5-components mixtures.

		Mindist	Maxdist	MeanMin	Coverage	Standard Deviation	AE	rmsd	ad	md
N = 20	KS	0,343	0,508	0,398	0,148	0,059	529,96	0,201	0,193	0,352
	WSP	0,335	0,369	0,338	0,022	0,007	808,54	0,179	0,173	0,361
	SbS	0,370	0,370	0,370	0,000	0,000	605,26	0,201	0,195	0,352
N = 30	KS	0,308	0,434	0,330	0,072	0,024	1527,89	0,173	0,168	0,297
	WSP	0,290	0,338	0,296	0,040	0,012	1981,14	0,161	0,156	0,313
	SbS	0,310	0,310	0,310	0,000	0,000	1585,20	0,176	0,170	0,293
N = 40	KS	0,255	0,338	0,285	0,100	0,028	3097,56	0,159	0,154	0,266
	WSP	0,255	0,285	0,260	0,034	0,009	3744,55	0,145	0,140	0,297
	SbS	0,290	0,290	0,290	0,000	0,000	2878,06	0,163	0,158	0,267
N = 50	KS	0,243	0,327	0,260	0,081	0,021	5230,42	0,147	0,142	0,278
	WSP	0,242	0,292	0,247	0,042	0,010	6010,38	0,137	0,132	0,244
	SbS	0,260	0,260	0,260	0,000	0,000	5046,34	0,151	0,147	0,253
N = 100	KS	0,184	0,259	0,202	0,074	0,015	25809,82	0,116	0,112	0,217
	WSP	0,187	0,230	0,191	0,035	0,007	27885,44	0,112	0,108	0,198
	SbS	0,210	0,210	0,210	0,000	0,000	22884,40	0,122	0,118	0,214

Table 5

Assessment criteria for designs with $N = 30$ to 200 points obtained by applying the SbS algorithm to 10-components mixtures. The Coverage and the Standard deviation criteria are equal to zero and are not been reported in the table.

	Mindist	Maxdist	MeanMin	AE	rmsd	ad	md
N = 30	0,280	0,280	0,280	2638,64	0,210	0,206	0,370
N = 50	0,250	0,250	0,250	7784,98	0,193	0,191	0,354
N = 70	0,240	0,240	0,240	15090,24	0,194	0,191	0,325
N = 90	0,209	0,209	0,209	28822,72	0,172	0,169	0,395
N = 110	0,219	0,219	0,219	37699,47	0,167	0,165	0,305
N = 130	0,212	0,212	0,212	55378,52	0,164	0,162	0,271
N = 150	0,208	0,208	0,208	69904,15	0,163	0,161	0,333
N = 170	0,203	0,203	0,203	96750,61	0,159	0,157	0,288
N = 200	0,200	0,200	0,200	138029,61	0,154	0,152	0,273

distribution in which two points are very closely-spaced. Interpretation based on the AE alone could therefore lead to erroneous conclusions. The criteria proposed by Borkowski and Piepel do not make it possible to rank designs based on their intrinsic qualities. Indeed, these criteria, although easy and rapid to calculate, measure distances between the design to be assessed and a random distribution of 10000 virtual points. Because of this approach, for the same design, different values can be obtained for this criterion at each calculation. As a result of the data presented here, we recommend retaining the *Mindist*, *MeanMin* and *coverage* criteria and completing them with the Audze-Eglais criterion.

4.4. Construction of designs with more than 3 components

4.4.1. Mixtures with 5 components

Considering the results obtained with 3 components, a study with 5 components has been achieved. Only three algorithms KS, WSP and SbS have been tested taking into account that Strauss and Dmax algorithms would be difficult to perform because of the required optimization of the parameters.

All the criteria have been calculated and values are reported Table 4 for $N = 20$ points to 100 points.

It is obvious that the three algorithms lead to very similar results. Nevertheless, we can observe that WSP designs always present AE values higher than KS and SbS designs that means that the potential energy of the structure is higher. Moreover, the *Mindist* values of SbS designs is slightly higher than those obtained with KS algorithm.

4.4.2. Mixtures with 10 components

The same behavior has been observed for 10 components. Only the criteria values for SbS algorithms have been reported Table 5 for $N = 30$ points to 200 points.

5. Conclusion

In this study, we presented algorithms for the construction of uniform designs for mixture variables. With regard to the construction of designs, our results demonstrate that some algorithms, such as the Strauss and Dmax algorithms, require optimization of their essential parameters to guarantee a uniform distribution. It is this optimization step that can be long and difficult, particularly with high dimensions. The KS algorithm tends to place points in the periphery of the space, and depending on the value of N , does not produce a uniform distribution throughout the space. The WSP algorithm produces uniform distributions, but requires adjustment of the *dmin* value depending on the desired number of points N , and on the structure of the set of candidate points. The new construction algorithm, SbS, has the advantage of not being a selection algorithm and therefore it is independent of the quality of the matrix of candidate points. This algorithm is powerful, guaranteeing good uniformity of distribution of the points throughout the space. Only the radius of the sphere must be adjusted to produce the desired number of points.

This study has been performed for “classic” experimental domain of interest which is a simplex for 3 components or a hypersimplex for more than 3 components. It would be interesting to extend this study to the reduced domain (hyperpolyhedron) generated by individual or relational constraints on the components. All the algorithms could be easily modified in order to take into account these constraints. Thus, for the selection algorithms as KS and WSP, it is sufficient to modify the step 1 of the algorithm by randomly generating the candidate points in the hyperpolyhedron. For the repulsion algorithms as Strauss and Dmax, it is the same principle. The step 2 of the algorithms must be changed and the matrix V is generated with N points in the space of constraints. For the last method, SbS algorithm, for each iteration, (Step 2 and 4), we have to test if the new point belongs to the hyperpolyhedron.

References

- [1] H. Scheffé, Experiments with mixtures, *J. Roy. Stat. Soc. B Methodol.* 20 (1958) 344–360.
- [2] H. Scheffé, The simplex-centroid design for experiments with mixtures, *J. Roy. Stat. Soc. B Methodol.* 25 (1963) 235–263.
- [3] K.T. Fang, D.K.J. Lin, P. Winker, Y. Zhang, Uniform design: theory and application, *Technometrics* 42 (2000) 237–248.
- [4] K.T. Fang, R. Li, A. Sudjianto, Design and Modeling for Computer Experiments, Chapman & Hall/CRC, 2006.
- [5] T.J. Santner, B.J. Williams, W.I. Notz, The Design and Analysis of Computer Experiments, Springer, New-York, 2003.
- [6] R.W. Kennard, L.A. Stone, Computer aided design of experiments, *Technometrics* 11 (1969) 137–148.
- [7] 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) 26–31.
- [8] M.C. Shewry, H.P. Wynn, Maximum entropy sampling, *J. Appl. Stat.* 14 (1987) 165–170.
- [9] J. Franco, X. Bay, D. Dupuy, B. Corre, Planification d'expériences numériques à partir du processus ponctuel de Strauss, 2008. <http://dice.emse.fr/communications> (accessed April 15, 2013).
- [10] M.E. Johnson, L.M. Moore, D. Ylvisaker, Minimax and maximin distance designs, *J. Stat. Plann. Inference* 26 (1990) 131–148.
- [11] V.C.P. Chen, K.L. Tsui, R.R. Barton, M. Meckesheimer, A review on design, modeling and applications of computer experiments, *IIE Trans.* 38 (2006) 273–291.
- [12] M.W. Trosset, Approximate maximin distance designs, in: *Proceedings of the Section on Physical and Engineering Sciences*, 1999, pp. 223–227.
- [13] M. Gunzburger, J. Burkhardt, Uniformity Measures for Point Samples in Hypercubes, 2004. <https://People.Sc.Fsu.Edu/~jburkhardt/Pdf/Ptmeas.Pdf>.
- [14] P. Audze, V. Eglais, New approach for planning out of experiments, *Probl. Dynam. Strengths* 35 (1977) 104–107.
- [15] J.J. Borkowski, G.F. Piepel, Uniform designs for highly constrained mixture experiments, *J. Qual. Technol.: Quart. J. Method. Appl. Related Top.* 41 (2009) 35–47.
- [16] M. Sergent, Contribution de la Méthodologie de la Recherche Expérimentale à l'élaboration de matrices uniformes : Application aux effets de solvants et de substituants, Thesis, Marseille, 1989.
- [17] M. Sergent, R. Phan-Tan-Luu, J. Elguero, Statistical analysis of solvent scales, part 1, *Anales de Química Int. Ed.* 93 (1997) 71–75.
- [18] M. Sergent, R. Phan-Tan-Luu, R. Faure, J. Elguero, Statistical analysis of solvents scales, part 2, *Anales de Química Int. Ed.* 93 (1997) 295–300.
- [19] J. Santiago, Développement de nouveaux plans d'expériences uniformes adaptés à la simulation numérique en grande dimension, Thesis, Aix-Marseille, 2013.
- [20] A. Beal, M. Claeys-Bruno, M. Sergent, Constructing space-filling designs using an adaptive WSP algorithm for spaces with constraints, *Chemometr. Intell. Lab. Syst.* 133 (2014) 84–91.
- [21] A. Benazzouz, L. Moity, C. Pierlot, M. Sergent, V. Molinier, J.-M. Aubry, Selection of a greener set of solvents evenly spread in the hansen space by space-filling design, *Ind. Eng. Chem. Res.* 52 (2013) 16585–16597.
- [22] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, Equation of state calculations by fast computing machines, *J. Chem. Phys.* 21 (1953) 1087–1092.
- [23] W.K. Hastings, Monte Carlo sampling methods using Markov chains and their applications, *Biometrika* 57 (1970) 97–109, <https://doi.org/10.1093/biomet/57.1.97>.
- [24] J. Franco, X. Bay, D. Dupuy, B. Corre, Planification d'expériences numériques à partir du processus ponctuel de Strauss, 2008.
- [25] J.A. Cornell, The use of independent variables, in: *Experiments with Mixtures*, John Wiley & Sons, Inc., 2002, pp. 96–131.