Design of experiments: a powerful tool in food microbiology

A. Bevilacqua¹,2, M.R. Corbo¹,2 and M. Sinigaglia¹,2

¹Department of Food Science, Faculty of Agricultural Science, University of Foggia, Via Napoli 25, 71122 Foggia, Italy
²Food Quality and Health Research Center (BIOAGROMED), University of Foggia, Via Napoli 52, 71122 Foggia, Italy

The Design of Experiments (DoE) is an efficient procedure able to combine two or more variables and point out the correlation amongst them. There are several kinds of DoE and the choice of one of them relies on the purpose of the research (comparative aim, screening or building a model). In this chapter we propose a step-by-step description of the theory of DoE, then focusing on the $2^k$ design, Central Composite Design and mixture approach. For each procedure, there are some details on how researchers can use it for microbiological purposes.

Keywords food microbiology; design of experiments; multivariate analysis

1. Design of Experiments: an overview

What is the Design of Experiments (DoE)? This is an open question with some possible answers. Concerning the biological sciences, in particular for food microbiology, we can define DoE as a structured, systematic and rigorous approach to problem solving that applies principles and techniques at the data collection stage, so as to ensure the generation of valid, defensible and supportable conclusions [1]. In other words, it is a method used to determine the relationship between the different factors affecting a process ($X_s$) and the output of that process ($Y$).

There are at least 4 general fields for DoE application:

1. **comparative**: researcher is interested in assessing whether a change in a single factor (called also variable) could result in a change (positive or negative) to the output of the process;
2. **screening/characterizing**: in this second case, researcher wants to assess what is the most important factor affecting the output of the process;
3. **modeling**: researcher wants to build a simple mathematical model, able to assure a good estimate of the output of the process as a function of the factors;
4. **optimizing**: researcher wants to assess the optimal setting of process parameters, i.e. determining the level of each factor of the design that optimizes process response.

The methodology of DoE was introduced by Fisher in 1935 [2]; as an example, he described how to test the hypothesis that a lady could distinguish by flavour alone whether the milk or the tea was first placed in the cup. This example could sound like a frivolous application, but it allowed Fisher to illustrate the 6 basic principles of DoE:

1. **comparison**, i.e. the use of a control acting as a baseline;
2. **replication**: the use of multiple measurements for each experiment is a good mean for the estimation of the casual variation around the results;
3. **blocking**, i.e. the arrangement of experiments into groups (called also blocks). Blocking reduces sources of variation, thus allowing a more precise estimate of the output of the process;
4. **randomization**: there are different meanings of the term randomization. In general, randomization means arrange something in random order [3]. In DoE theory, this word is correlated with "blocking": when a set of experiments is divided into different groups or blocks, randomization gives each experiments the same choice of being assigned to any of the groups [4];
5. **orthogonality**: in mathematics, two vectors are orthogonal if they are perpendicular. Otherwise, in statistics two experiments are orthogonal if they are unrelated, finally, in programming language, orthogonality means that a relatively small set of primitive constructs can be combined in a relatively small number of ways to build the control and data structures of the language. In DoE theory, orthogonality means that there are several set of experiments or some levels of the factors that are independent of one another and can be combined to derive all the combinations of the design;
6. **factorial experiments**, i.e. the use of significant combinations (factorial experiments) amongst the different factors, instead of evaluating one factor at a time. Factorial experiments are useful for the assessment of some possible interactions amongst the factors.

A systematic description of the step-by-step procedure for an effective DoE approach can be found in the book of Box et al. [5]. Briefly, a robust DoE approach for food microbiology can be obtained through a step-by-step procedure, reported in Fig. 1.
There are several kinds of designs, depending on the purpose and the number of factors to be tested; a classification of the different DoE approaches was proposed by Box et al. [5], later modified by Bevilacqua and Sinigaglia [6]. For the purpose of food microbiology, it can be summarized as reported in Table 1.

<table>
<thead>
<tr>
<th>Number of factors</th>
<th>Purpose</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-4</td>
<td>Screening/studying the effects of the factors</td>
<td>Building a model/optimizing the output of a process</td>
</tr>
<tr>
<td></td>
<td>Full Factorial Designs</td>
<td>Central Composite Design</td>
</tr>
<tr>
<td></td>
<td>Fractional Factorial Designs</td>
<td>Box-Behnken Design</td>
</tr>
<tr>
<td>5 or more</td>
<td>Fractional Factorial Designs</td>
<td>Reduce the number of variables</td>
</tr>
<tr>
<td></td>
<td>Plackett-Burman Design</td>
<td></td>
</tr>
</tbody>
</table>

Due to the complexity of DoE theory and the peculiarities of each approach, in this chapter we cannot describe all the designs; moreover, many designs are not currently used in food microbiology, but in social sciences as well as in engineering and computational sciences. Based on a quick insight on the literature over the past decade, we can propose a kind of decision tree for the use of DoE in food microbiology, containing both the convenient approaches to be used and the outputs of the design (Fig. 2).
Fig. 2 Decision tree for DoE approach in food Microbiology: purpose, design and output.

Apart from this decision tree, there is the approach of the mixture design (e.g. the centroid); it has been proposed by some authors for some benefits (few combinations, possibility of weighting each factor of the design).
In the following paragraphs, there is a brief overview of the method reported in the decision tree, as well as a description of centroid approach.

2. Screening: the $2^k$-factorial design

2.1 The design

$2^k$ factorial designs (or two-level factorial design) can be considered as the simplest approach of DoE for food microbiology; briefly, it consists of $k$ variables or factors, set at two different levels, indicated as "-1" (the minimum) and "+1" (the maximum). The factors of this kind of design can be either quantitative and qualitative; thus two levels of a quantitative variable, two different temperatures or two different concentrations of an antimicrobial; for qualitative factors they might be two kind of compounds, or the presence/absence of a chemical.

The two-level factorial designs are of special importance for different reasons [7]:

1. they require relatively few runs per factor studied;
2. when the factors are quantitative, although unable to explore a wide region in the factor space, they often determine a promising direction for further experiments;
3. they form the basis of two-level factorial-fractional designs, in which only significant combinations amongst the factors are performed;
4. factorial designs fit naturally into a sequential strategy.

Writing a $2^k$ factorial design is very simple and can be described as a step-by-step procedure:

1. define the factors (k) (step 1);
2. define the level of the factors (defined as -1 and +1, or - and +, or 0 and 1) (step 2);
3. the number of the combinations is given by the equation $2^k$, that means that if I have 3 factors, the number of combinations will be $2^3$, i.e. 8. If the factors are 4, the combinations will be 16 and so on (step 3);
4. write the combinations amongst the factors following a scheme based on $2$ to (k-1) power (step 4);

We can try to make clear this procedure by an example. *Alicyclobacillus acidoterrestris* is an emerging microorganisms, because it is able to spoil juices and other acidic drinks [8]; the growth can occur for pHs and temperatures of 3.5-5.5 and 35-55°C, respectively. In addition, oxygen is another key-element for its growth: alicyclobacilli are aerobic (21% of oxygen); however, they can grow also under micro-aerobic conditions (0.1%) [9].

Therefore, the variables under investigations are pH, temperature and O$_2$ (step 1); the levels of the variables can be set at: 3.5 (-1) and 5.5 (+1) for the pH; 35 (-1) and 55°C (+1) for the temperature; 0.1 (-1) and 21% of O$_2$ (+1) (step 2); the number of combinations is 8 (step 3); we can write the combinations as reported in the Table 2 (step 4).
Table 2  

<table>
<thead>
<tr>
<th>Procedure</th>
<th>$2^{k-1}$, where k is the number of the factors</th>
<th>$2^1$</th>
<th>$2^0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Combinations</td>
<td>Write 4 times &quot;-1&quot; and 4 times &quot;+1&quot;</td>
<td>Write 2 times &quot;-1&quot; and 2 times &quot;+1&quot;</td>
<td>Write 1 time &quot;-1&quot; and 1 time &quot;+1&quot;</td>
</tr>
<tr>
<td>1</td>
<td>3.5 (-1)</td>
<td>35°C (-1)</td>
<td>0.1% (-1)</td>
</tr>
<tr>
<td>2</td>
<td>3.5 (-1)</td>
<td>35°C (-1)</td>
<td>21% (+1)</td>
</tr>
<tr>
<td>3</td>
<td>3.5 (-1)</td>
<td>55°C (+1)</td>
<td>0.1% (-1)</td>
</tr>
<tr>
<td>4</td>
<td>3.5 (-1)</td>
<td>55°C (+1)</td>
<td>21% (+1)</td>
</tr>
<tr>
<td>5</td>
<td>5.5 (+1)</td>
<td>35°C (-1)</td>
<td>0.1% (-1)</td>
</tr>
<tr>
<td>6</td>
<td>5.5 (+1)</td>
<td>35°C (+1)</td>
<td>21% (+1)</td>
</tr>
<tr>
<td>7</td>
<td>5.5 (+1)</td>
<td>55°C (+1)</td>
<td>0.1% (-1)</td>
</tr>
<tr>
<td>8</td>
<td>5.5 (+1)</td>
<td>55°C (+1)</td>
<td>21% (+1)</td>
</tr>
</tbody>
</table>

*aIn the brackets there is the level (coded value) of the factors.

2.2 Results and Pareto chart of standardized effects

The analysis of the data of a $2^k$ design is based on the ANOVA (analysis of variance) and the method of Yates. Briefly, for the example of A. acidoterrestris (pH, temperature and oxygen), a computational software estimates the effects of each factor (individual effect) and the influence of the interaction amongst them (interactive effects, i.e. pH*temperature, pH*oxygen, temperature*oxygen and pH*temperature*oxygen). These effects, called standardized effects, can be represented in the Pareto chart.

The standardized Pareto Chart contains a bar for each effect, sorted from the most significant to the least significant one. The length of each bar is proportional to the standardized effect, which equals the magnitude of the t-statistic that would be used to test the statistical significance of that effect. A vertical line is drawn at the location of the 0.05 critical value for Student’s t-test. Any bars that extend to the right of that line indicate effects that are statistically significant at the 95% significance level.

As an example figure 3 shows the Pareto chart resulting from a $2^k$ design for the evaluation of the effects of pH, temperature and oxygen on the growth of A. acidoterrestris.

![Pareto chart of the standardized effects of pH, temperature and oxygen on the growth of A. acidoterrestris.](image)

**Fig. 3** Pareto chart of the standardized effects of pH, temperature and oxygen on the growth of A. acidoterrestris.

What are the conclusions/ideas of this figure? Briefly, temperature, pH and oxygen are significant as individual and positive terms (i.e. an increase of one factor at a time results in an increase of microbial growth); moreover, the statistical analysis highlights a significant effects of the interactions pH*temperature and temperature*oxygen.
3. Building a model: the Central Composite Design

3.1 Combinations of the design

An interesting tool in food microbiology is the Box-Wilson Central Composite Design, usually called Central Composite Design. It involves five different levels for each factor and contains a factorial design, added with some combinations containing the central point (i.e. the mean value of each factor) and some star points (i.e. the minimum and the maximum of the range) to allow a good estimation of the curvature of the output.

There are several kinds of Box-Wilson Design (Circumscribed, Inscribed and Face-Centered), depending on the distance of the levels from the average value of each factor; the most used approach in food science is the Circumscribed Design, usually referred to as CCD (Central Composite Design).

In a CCD the distance from the centre of the design space to a factorial point is ±1 unit for each factor, the distance from the centre of the design space to a star point is ±α with \( \alpha > 1 \). The precise value of \( \alpha \) depends on certain properties desired for the design and on the number of factors involved. If the design is rotatable (i.e. the levels can be put on a circumference and central point is the centre of this circumference), the value of \( \alpha \) can be calculated as follows (eq. 1):

\[
\alpha = \left[ 2^k \right]^{\frac{1}{4}}
\]

where \( k \) is the number of factors of the design, that means that \( \alpha \) is 1.414 for 2-factor design and 1.682 when the factors are 3; however, any value between 1.5 and 2.0 gives reasonably satisfactory design [10]. Finally, \( \alpha \) is 2 for 4-factor design.

Writing a CCD is very simple, if you follow a step-by-step procedure:

1. Define the variables (factors).
2. Define the levels of the variables; it is a common practice choose a range for each factor, setting the minimum and the maximum value of this range as "-2" and "+2", respectively; then, the mean value is considered the central point of the range (coded value 0). Finally, you should estimate the mean values of the ranges -2/0 and 0/+2, that are referred to as the levels "-1" and "+1". For example, the main purpose of many researches is to study the combined effect of pH and Aw on \( A. acidoterrestris \) growth; I know that alicyclobacilli can grow in a range of pH of 3.5-5.5. Therefore, 3.5 and 5.5 will be referred to as the coded levels "-2" and "+2", respectively; then, the mean value of this range (i.e. 4.5) is the central point of my design (level "0"). Finally, I have to calculate the mean values of the ranges 3.5/4.5 and 4.5/5.5, that are 4.0 and 5.0; these values are referred to as "-1" and "+1".
3. Define the number of combinations of the design: a 3-factor CCD requires at least 15 combinations (normally 17), whereas the combinations of a 4-factor design are at least 25 (normally 30).
4. Write the combinations, adding to the design the so called cube points, the star and the central points. The cube points are the combinations of the design in which all the factors are set to the levels "-1" or "+1", following a binary procedure, as reported for a \( 2^k \) design; a central point is a kind of control of the CCD, because all the factors are set to the mean value of the range, i.e. the level "0". Otherwise, the star points contain a factor at the minimum or the maximum (i.e. the levels "-2" and "+2") and the other set to the central value. A central composite design at 3 factors can be represented as a cube; the vertices of the cube are the cube points, whereas the centre of the solid is the central point of the design. If we imagine that the cube is inscribed in a sphere, the star points define the ray of the sphere. A graphical representation of CCD is reported in the figure 4.

![Graphical representation of a 3-factor Central Composite Design](image)
How write a CCD? Write the cube points following the scheme of $2^k$ design, using the levels "-1" and "+1"; then you should write a central point (all the factors at the level "0") and the star points. Finally add some other central points. Table 3 shows a typical scheme for a 3-factor CCD.

<table>
<thead>
<tr>
<th>Cubes points</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>2</td>
<td>-1</td>
<td>-1</td>
<td>+1</td>
</tr>
<tr>
<td>3</td>
<td>-1</td>
<td>+1</td>
<td>-1</td>
</tr>
<tr>
<td>4</td>
<td>-1</td>
<td>+1</td>
<td>+1</td>
</tr>
<tr>
<td>5</td>
<td>+1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>6</td>
<td>+1</td>
<td>-1</td>
<td>+1</td>
</tr>
<tr>
<td>7</td>
<td>+1</td>
<td>+1</td>
<td>-1</td>
</tr>
<tr>
<td>8</td>
<td>+1</td>
<td>+1</td>
<td>+1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Central point</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Star points</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0</td>
<td>0</td>
<td>-2</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>0</td>
<td>+2</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>-2</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>+2</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>-2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>+2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Central points</td>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>16</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>17</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

### 3.2 Polynomial equation

A CCD supports the building of a polynomial equation which takes into account the individual, interactive and quadratic terms and basically reads as follows (eq. 2) [11]:

$$y = \beta_0 + \sum_{j=1}^{n} \beta_j \cdot X_j + \sum_{j=1}^{n} \sum_{k=j+1}^{n} \beta_{jk} \cdot X_j \cdot X_k + \sum_{j=1}^{n} \beta_{jj} \cdot X_j^2$$

(2)

where $\sum_{j=1}^{n} \beta_j \cdot X_j$ is the individual effect of each factor; $\sum_{j=1}^{n} \sum_{k=j+1}^{n} \beta_{jk} \cdot X_j \cdot X_k$ indicates the interactions amongst the variables; finally, the term $\sum_{j=1}^{n} \beta_{jj} \cdot X_j^2$ takes into account a possible non-linear/quadratic effect of some factors.

What does it mean "quadratic effect"? It indicates a possible curvature and non-linear correlation between a factor and the output of the equation and the existence of a possible minimum and/or maximum, when the individual and the quadratic terms show different sign (e.g. the individual effect is positive and the quadratic negative, thus it indicates that the output of the process increase with increasing of the factor up to a critical threshold or break-point of the factor itself, after which a further increase results in a decrease of the output).

An example of a polynomial equation, built on the results of a CCD, is that reported by Bevilacqua [12], showing the effects of glucose (0.0-4.0%), NaCl (0.0-6.0%), $p$-coumaric (0.0-0.4%) and vanillic (0.0-0.4%) acids on the growth of *Lactobacillus plantarum* strains isolated from Italian table olives; the equation reads as follows (table 4):

| Polynomial equation showing the individual and interactive effects of glucose (glu), NaCl, p-coumaric (coum) and vanillic acides (van) on the growth of *L. plantarum* |
|---|---|---|---|
| Equation | SE | F | $R^2$ |
| $N = 0.84 \cdot glu + 1.02 \cdot NaCl + 8.00 \cdot coum - 0.15 \cdot glu^2 - 0.10 \cdot NaCl^2 - 13.41 \cdot coum + 0.35$ | 90.63 | 0.978 |
| $- 5.69 \cdot van - 0.11 \cdot glu \cdot NaCl - 1.86 \cdot NaCl \cdot coum$ |  |  |

*SE, standard error of the model; F, Fisher's test value; $R^2$, multiple regression coefficient*
Trying to give a biological meaning to this polynomial equation, we could obtain the following details:

1. The effect of glucose in not strictly linear, as the equation contains both a positive individual effect (first order term) and a negative quadratic coefficient. Therefore, an increase of the glucose in the medium determines an increase of cell number of lactic acid bacteria; however, the negative quadratic terms highlights that there is a critical concentration of the glucose, after which it begins to be toxic for cells, thus reducing cell count.

2. A similar trend can be deduced also for salt, with a critical concentration decreasing cell number.

3. Another interesting idea can be inferred from the interactive terms. Concerning the interaction glucose/NaCl, the mathematical coefficient is negative (-0.11), thus it suggests that glucose and NaCl act in a negative way each other, i.e. if I want to maintain high level of cell number, I should keep the glucose at high levels and use low concentration of NaCl (or do the opposite).

The polynomial equations, built on a CCD, as well as all the polynomial models show some benefits and some limits [13]. The benefits are:

1. the use of a simple regression procedure for model building;
2. no knowledge needed.

Opposite to these advantages, there are some limits, i.e.

1. a polynomial equation has not a theoretical foundation and the coefficients could not a have a biological meaning;
2. the equation can be applied only for the situation for which it has been developed;
3. it does not contribute to investigate the mechanisms giving those results.

Despite these limits, a polynomial equation resulting from a CCD approach remains one of the most useful tool for food microbiology, for example for the optimization of a preserving treatment or a technological process, the improvement of a laboratory medium or pointing out a critical dose of an antimicrobial.

### 3.3 Surface response plots and statistical indices

Along with the polynomial equation, the main outputs of a CCD are some statistical parameters, which give a measure of the goodness of fitting of the proposed equation, i.e.:

a. the regression coefficients and the adjusted regression coefficient (a regression coefficient corrected taking into account that the model is based on many independent variables);

b. the lack of fit and the standard error of the model, which gives a measure of the probable variation of the data around the prediction;

Fisher test value (F-value), which is an indirect measure of the significance of the model (see table 4).

Finally, the polynomial equation are graphically represented by a surface response plot, i.e. a 3D plot reporting the interaction of two factors and their effect on the output of the equation. How build a 3D plot for a CCD with more than 2 factors (for example in the case of 3 independent variables)? It's a common practice to impose a constant value to a factor (generally the coded level "0"), thus studying the interaction between the others.

For example, Sinigaglia et al [14] studied the effects of pH (3.5-5.5), temperature (35-45°C) and A_w (0.960-0.992) on the germination of *A. acidoterrestris* spores. After building a polynomial equation, they drew the surface response plots as follows:

- interaction pH vs temperature, setting A_w at 0.976;
- interaction pH vs A_w, setting the temperature at 45°C;
- interaction A_w vs temperature, setting pH at 4.5.

This kind of approach appears as useful and gives some details on the effects of two factors, when the other (or the others in the case of a 4-factor CCD) is set at the optimal point. A modification of this approach has been proposed by Bevilacqua et al. [15], setting the third factor at the extreme (levels "-2" or "+2") with the aim of studying the effect of the interaction under sub-optimal conditions.

A modification of the surface response plot is the contour plot, reporting the interactions in a 2D figure (figure 5).
4. Optimizing a process: saddle point and desirability function

A new frontier in the field of food microbiology is the approach of process optimization, that can be achieved either by the method of Taguchi and by response surface methodology and CCD procedure.

The term Taguchi Methods refers to a collection of principles which make up the framework of a continually evolving approach to quality. This system of quality engineering takes its name (at least in the United States) from Genichi Taguchi, who along with Deming, Juran and Ishikawa, is considered a pioneer of the modern quality movement.

Taguchi defines quality in terms of a loss to society caused by a product failure. For instance, such a loss might be a loss of product function, time, property, a negative environmental effect, or more significantly, a financial loss. Thus, he introduced the concept of quality loss function, as well as the use of orthogonal arrays.

Despite its effectiveness, Taguchi method is not used currently in food microbiology, as process optimization is achieved through Central Composite Design, by evaluating the saddle point and the desirability function.

4.1 Saddle point

There are at least two different definitions of saddle point (a mathematical and a statistical ones). In the field of DoE, a saddle point is the point of maximum and minimum response; it is the solution of the following equation (equation 3):

\[
\frac{\partial Y}{\partial x_1} = \frac{\partial Y}{\partial x_2} = \ldots = 0
\]  

What is the impact of saddle point for process optimization? It represents the combination of the factor achieving the desired result (for example the maximal production of biomass [16] or the maximal inactivation of a spoiling microorganism [15]).

4.2 Desirability function

The desirability function could represent the last step of model building through the CCD approach; it is based on the prediction profiles of each factor obtained through the polynomial equation and convert these values in a dimensionless parameter, i.e. the desirability, ranging between 0, if the value of the output is completely unacceptable, and 1, if we obtain the desired target.

The desirability function can be evaluated as follows (equation 4) [17]:

\[
\begin{align*}
0, & \quad y < y_{\text{min}} \\
\frac{y - y_{\text{min}}}{y_{\text{max}} - y_{\text{min}}}, & \quad y_{\text{min}} \leq y \leq y_{\text{max}} \\
1, & \quad y > y_{\text{max}}
\end{align*}
\]  

Fig. 5 Interaction of the factors A (3.5-5.5-arbitrary units) and B (0-160-arbitrary units) on the output of the process C (arbitrary units). a) surface response plot; b) contour plot.
In this equation, $y$ is the output of the process and $y_{\text{max}}$ and $y_{\text{min}}$ the maximal and the minimal response, respectively. The desirability approach represents an useful tool to point out a good compromise between the optimal response process (i.e. the saddle point) and the costs of a process.

What is its impact? Researcher could impose a minimal result to be achieved (for example a reduction of a spoiling microorganism of 50%) and the desirability function highlights the condition that can achieve that result.

Some examples of desirability function in food microbiology are reported by Bevilacqua et al. [15, 17], in order to attain the combinations of pH/cinnamaldehyde/eugenol/thermal treatment resulting in a significant reduction of *A. acidoterrestris* spores, and by Li et al. [16], for the optimization of the composition of medium used for the production of glutamine by *Corynebacterium glutamicum*.

### 5. Mixture designs

A mixture experiment is a particular kind of DoE in which the factors are ingredients or components of a mixture. Mixtures are different from other types of experimental design because the proportions of the constituents must add up to 100%. Increasing the level of one constituent necessarily reduces the level of the others. Types of mixture design include [18]:

1. augmented lattice designs
2. extreme vertex designs
3. lattice designs
4. L simplex designs
5. simplex centroid designs.

The most simple mixture design is the simplex-lattice design, which contains some combinations with one pure factor (i.e. a factor set at the level 100% and the other at 0) and some other combinations with two components at 0.5 (i.e. 50% of two components in the mixture, the other at 0) [18].

If we represent this design with an equilater triangle, whose side is 1, the coordinates of combinations are: (1, 0, 0), (0, 1, 0), (0, 0, 1), (0.5, 0.5, 0), (0.5, 0, 0.5) and (0, 0.5, 0.5); this design is characterized by a step of 1/2, because each side of the triangle is divided into two parts of equal length [18] (figure 6).

<table>
<thead>
<tr>
<th>combination</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.5</td>
<td>0</td>
<td>0.5</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0.5</td>
<td>0.5</td>
<td>0</td>
</tr>
</tbody>
</table>

![Fig. 6](image-url) Simplex-lattice design. a) combinations; b) graphical representation.

A more complex lattice design is characterized by a step of 1/3 because it contains some combinations like 0.33-0.67-0 and 0.33-0.33-0.33.

A central point can be added systematically to a simplex-lattice design to obtain a simplex centroid design; thus a simplex centroid contains:

- Pure products
- Fifty-fifty mixtures of pure products
- A combination with an equal amount of the 3 pure products (central point) (figure 7) [18].
For microbiological purpose, as well as for building a simple polynomial equation, some new combinations should be added located at the centre of gravity of simplexes (the small individual triangle composing the big one) (figure 8).

As reported for CCD and $2^k$ designs, the output of a mixture designs are a Pareto chart of standardized effects and a polynomial equation, along with a graphical representation of the interactions amongst the factors (triangular plots) [18].

6. Conclusions

DoE represents an important tool for food microbiology, due to the different approaches available and the flexibility of the theory. Its application allows a reliable estimation of the individual effects of factors, as well as the optimization of a process.

However, like all the statistical approaches, DoE is only a tool, not a golden remedy; therefore, there are situations or systems where its use is convenient and other where it should not be used. In conclusion, it’s a tool: it's up on the researcher choose when use and when not, keeping in mind the biological phenomenon under investigation and the final goal of the experiment.
References