

# Principle of a fractional factorial design for qualitative and quantitative factors: application to the production of *Bradyrhizobium japonicum* in culture and inoculation media

S Cliquet <sup>1†</sup>, C Durier <sup>2\*</sup>, A Kobilinsky <sup>2</sup>

<sup>1</sup> INRA, laboratoire de microbiologie des sols, F21034 Dijon cedex;

<sup>2</sup> INRA, laboratoire de biométrie, F78026 Versailles cedex, France

(Received 13 August 1993; accepted 8 November 1994)

**Summary** — To improve the production of *Bradyrhizobium japonicum* in liquid culture media, different carbon and nitrogen substrates at different concentrations were tested. In order to study simultaneously these qualitative and quantitative factors, a suitable experimental design was necessary. We develop here the principle leading to such fractional factorial designs. The specific design used allowed us to decrease the theoretical number of treatments from 1 024 to 128 and to get estimates of factorial main effects and 2-factor interactions. Five of the 7 tested factors, pH, carbon and organic nitrogen sources, yeast extract concentration and organic nitrogen concentration, were found to have a significant effect on optical density. They also were found to interact with each other. The design allowed us to select 2 media that produced more than 10<sup>10</sup> bacteria/ml.

***Bradyrhizobium japonicum* / carbon and nitrogen substrate / pH / factorial fractional design / microbial production**

**Résumé** — Principe d'un plan factoriel fractionné intégrant des facteurs qualitatifs et quantitatifs. Application à la production d'un milieu de culture et d'inoculation pour *Bradyrhizobium japonicum*. Afin d'améliorer la production de *Bradyrhizobium japonicum* en milieu de culture liquide, nous avons comparé les performances de plusieurs substrats carbonés et azotés à différentes concentrations. La prise en compte simultanée de ces facteurs qualitatifs et quantitatifs a nécessité l'utilisation de plans factoriels adaptés, dont nous développons ici le principe de construction. Le plan factoriel fractionnaire retenu a permis de réduire le nombre de traitements étudiés de 1 024 à 128, tout en autorisant l'estimation des effets principaux et des interactions entre 2 facteurs. L'analyse a montré que, sur les 7 facteurs testés, 5 d'entre eux : le pH, la source de carbone, la source d'azote organique, la dose d'extrait de levure et la dose d'azote organique, ont un effet principal significatif sur la densité optique du milieu et présentent également des interactions significatives. Le plan d'expérience a permis de sélectionner 2 milieux de culture produisant plus de 10<sup>10</sup> germes/ml.

***Bradyrhizobium japonicum* / substrat carboné et azoté / pH / plan factoriel fractionnaire / production microbienne**

\* Correspondence and reprints

† Present address: NCAUR-USDA, 1815 North University, Peoria, 61604 IL, USA.

## INTRODUCTION

Soybean yield in France requires the inoculation of a symbiotic bacterium, *Bradyrhizobium japonicum*, which fixes atmospheric nitrogen. Inoculation is usually performed with sterile peat containing a culture of *B japonicum*, and mixed with soybean seeds on sowing. This technique is expensive however (peat grinding, packaging, and sterilization) and judged as dull by farmers. Bacterial survival can also be severely reduced by desiccation in the hours following inoculation.

A high bacterial density is thus required during growth and storage. Therefore it is necessary to use a culture medium that allows for a high bacterial density. To achieve this, the carbon and nitrogen substrates metabolized by *B japonicum* should be carefully tested and a large number of culture media experimented. Several studies (Duménil *et al*, 1975; Phan-Tan-Luu *et al*, 1979; Fannin *et al*, 1981) concerned fractional factorial designs taking into account several factors simultaneously with possible interactions between them. These designs involved only quantitative factors and some were applied to the optimization of fermentation (Deshayes, 1980; De Méo *et al*, 1985).

We introduce here the principle of fractional designs for qualitative and quantitative factors at 2 and 4 levels. The design built along these principles allowed us to decrease the theoretical number of treatments from 1 024 to 128 in a 1/8th fractional design. It is also shown how an efficient 1/16th fractional design, with 64 units only, could have been obtained had the number of available experimental units been more restricted.

## MATERIAL AND METHODS

### Strain and culture medium

The strain of *B japonicum* G49 (Indian Agricultural Research Institute (IARI) SB16, New-Delhi) used is the current strain for inoculation of soybean in France. The basal medium (Bergersen, 1961), modified from Sherwood (1972) and Balatti and Mazza (1978) contained (per l): NaCl, 0.1 g; MgSO<sub>4</sub>•7H<sub>2</sub>O, 0.492 g; CaCl<sub>2</sub>•2H<sub>2</sub>O, 0.147 g; FeCl<sub>3</sub> at 28%, 30 µl.

Nitrogen, carbon substrates and addition of a growth factor (yeast extract) were the factors studied in the experimental design. The pH was adjusted to 6 or 7 with KH<sub>2</sub>PO<sub>4</sub> and K<sub>2</sub>HPO<sub>4</sub> at 10 mM. The media containing carbon, nitrogen sources and yeast extract

were autoclaved at 120°C except glucose which was sterilized by filtration.

### Growth conditions

Because numerous runs were made, bacteria were grown in tubes (18 by 180 mm) previously matched for optical density (OD) readings. Tubes were placed on a rotary shaker (200 rpm) and incubated at 28°C for 6 d. In these conditions, bacterial production in flasks was 6 x 10<sup>9</sup> bacteria/ml (standard deviation: 2 x 10<sup>9</sup>) and of the same order of magnitude as microbial production in tubes (5 x 10<sup>9</sup> bacteria/ml; standard deviation: 3 x 10<sup>9</sup>).

### Analysis

The optical density was read at 600 nm after growth. On some cultures, viable *B japonicum* were counted by plate dilution on Bergersen medium (Bergersen, 1961) using the Spiral system.

### Choice of factors and their levels

The choice of factors and their levels was determined by preliminary trials (Cliquet, 1990) and from bibliographic data. Four carbon sources were selected: glucose, gluconate, mannitol and glycerol (Burton, 1967; Elkan and Kwik, 1968; Keele *et al*, 1969, 1970). It is believed that bacterial growth is not limited by the carbon dose, so studying it at 2 levels seemed to be sufficient. On the other hand, yeast extract, an important growth factor, might show toxicity at high concentrations (Date, 1972; Sherwood, 1972).

Yeast extract concentration therefore had to be at least a 3-level factor to detect the curvature implied by its toxic effect. Construction of fractional designs is however easier for mixtures of 2-level and 4-level factors than for 2-level and 3-level factors (Kobilinsky and Monod, 1991). Four doses were therefore chosen, 1, 2, 3 and 4 g/l for yeast extract, and 0.1, 0.2, 0.3 and 0.4 g/l for organic nitrogen dose, as either Na glutamate or casein hydrolysate (Elkan and Kwik, 1968; Chakrabarti *et al*, 1981).

As simultaneous mineral and organic nitrogen might also benefit growth (Elkan and Kwik, 1968), NH<sub>4</sub>Cl was sometimes added. Finally, initial pH was adjusted to 6 or 7, the usual range for this type of culture (Jordan, 1982).

There were thus 7 factors in this factorial design, the levels of which are given in table I. The number of combinations of levels of the 7 factors was 4 x 4 x 4 x 2 x 2 x 2 x 2, that is 1 024 combinations also called the treatments. Only a fraction of these, made up of 128 treatments, was tested. The method we used to choose the fraction will be explained, using a plain 2-factor example.

**Table I.** Factors and levels in the experimental design.

Symbol	Factor	Number of levels	Levels (coded levels)			
A	Carbon source	4	Mannitol	Glycerol	Gluconate	Glucose
B	Organic nitrogen dose (g/l)	4	0.1 (-3)	0.2 (-1)	0.3 (+1)	0.4 (+3)
C	Yeast extract dose (g/l)	4	1 (-3)	2 (-1)	3 (+1)	4 (+3)
D	Organic nitrogen source	2		Casein hydrolysate (-1)	Sodium glutamate (+1)	
E	Carbon dose (g/l)	2		4 (-1)	6 (+1)	
F	Ammonium chloride dose (g/l)	2		0 (-1)	0.1 (+1)	
G	pH	2		6 (-1)	7 (+1)	

## PRINCIPLE OF THE EXPERIMENTAL DESIGN

### Quantitative factor case: the polynomial effects

Consider the case in which 2 quantitative factors are to be studied:

Factor	Number of levels	Coded levels
A	4	-3, -1, 1, 3
B	2	-1, 1

The quantity  $Y$  measuring the microbial growth is assumed to follow the model:

$$Y = f(A, B) + \varepsilon$$

In this context,  $Y$  is often called the experimental response or just the response. It differs from  $f(A, B)$ , the theoretical response (Phan-Tan-Luu *et al.*, 1983), because of the error  $\varepsilon$ .

Quantities called polynomial factorial effects, or more simply polynomial effects, which characterize the influence of the various factors and

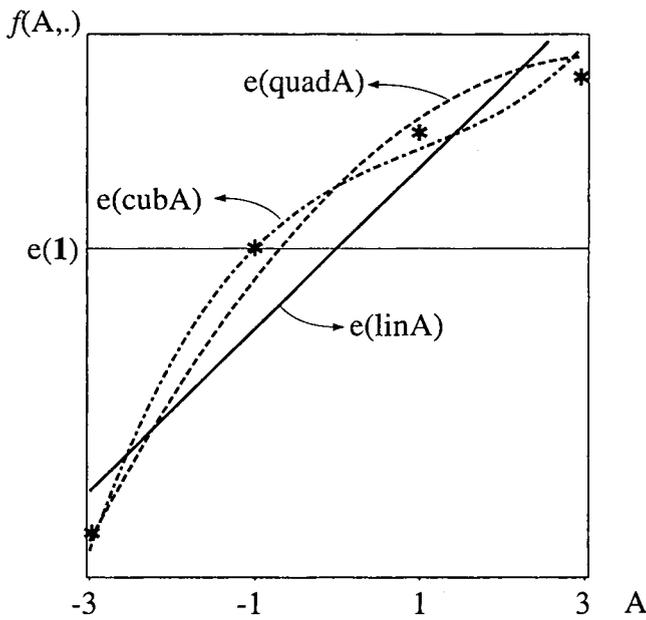
their interactions on the response, are deduced from the theoretical response.

More precisely, these effects are defined from the values of the theoretical response  $f$  on the grid of  $4 \times 2$  points associated with the different combinations of levels of the factors A and B. Thus, the mean factorial effect stands for the mean of this set of 8 values. It is denoted by  $e(1)$ .

Consider a given factor, say A. For each of its levels, we can compute the corresponding mean of the values of  $f$ , i.e. in this case the mean over the 2 levels of B. If A is used to denote the level (as well as the factor), the associated mean is denoted by  $f(A, \bullet)$ :

$$f(A, \bullet) = (f(A, 1) + f(A, -1))/2$$

The slope of the best fitting straight line between these means and the factor A is called the linear factorial effect of factor A and denoted  $e(\text{lin}A)$  (John, 1971). The coefficient of  $A^2$  in the best fitting parabola (fig 1) is called the quadratic factorial effect and denoted  $e(\text{quad}A)$  (John, 1971). Finally, if the curve representing the theoretical mean response  $f(A, \bullet)$  as a function of A exhibits an inflection point, one has to consider the cubic effect  $e(\text{cub}A)$ , defined as the coefficient of  $A^3$  in the cubic that best fits the 4 points considered on that curve (fig 1).



**Fig 1.** Polynomial factorial effects: \* represents the theoretical values of  $f(A, \bullet)$ ; the arrows indicate the effect deduced for each best fitting curve: the ordinate of the horizontal line — is the mean effect,  $e(1)$ ; the slope of the straight line — is the linear effect,  $e(\text{lin}A)$ ; the curvature of the quadratic curve — — is the quadratic effect,  $e(\text{quad}A)$ ; the coefficient of  $A^3$  in the cubic curve — — is the cubic effect,  $e(\text{cub}A)$ .

Since factor B has only 2 levels, the only polynomial effect of B which can be considered is the linear effect  $e(\text{lin}B)$ . To study the combined influence of the 2 factors A and B, interaction terms denoted by  $e(\text{lin}A \cdot \text{lin}B)$ ,  $e(\text{quad}A \cdot \text{lin}B)$  and  $e(\text{cub}A \cdot \text{lin}B)$  are introduced. With each interaction term is associated a polynomial degree which is defined as the sum of the degrees in each effect. The degree is 2 for  $e(\text{lin}A \cdot \text{lin}B)$ , 3 for  $e(\text{quad}A \cdot \text{lin}B)$ , and 4 for  $e(\text{cub}A \cdot \text{lin}B)$ .

To define, for instance,  $e(\text{lin}A \cdot \text{lin}B)$ , one looks first at the variation of the theoretical response in function of A for each level of B and deduces, as previously, a linear effect of A for this level of B, denoted by  $e(\text{lin}A(B))$  (fig 2). The variation of this effect as a function of B then yields  $e(\text{lin}A \cdot \text{lin}B)$ , which is here the slope of the straight line through the points associated with  $B = -1$  and  $B = +1$  in figure 3. The effects obtained in this way are generally normalized by multiplication by an appropriate constant in order to make them comparable with each other.

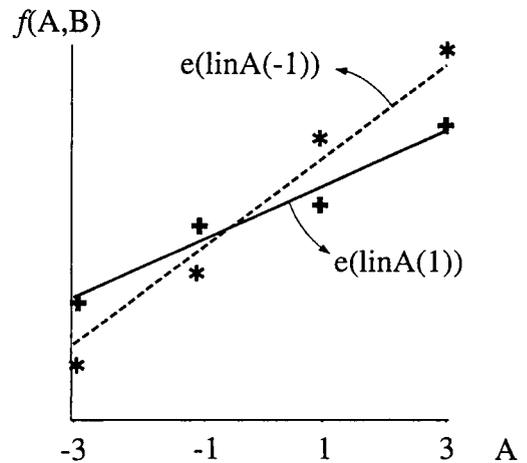
There is a very simple way of obtaining these normalized effects from the values of the theoretical response  $f$  on the 8 points of the grid. One introduces the orthogonal polynomials in A and B denoted by *mean*, *linA*, *quadA*, *cubA* and *linB* whose expressions as functions of A and B are given in the *Appendix*. These orthogonal poly-

nomials, which have long since been used to simplify the computations in polynomial regressions, are defined and tabulated in Fisher and Yates (1957) and Pearson and Hartley (1976). Although their interest for calculation has now disappeared, their use is still highly recommended in the general context of polynomial regression to obtain meaningful parameters which are the least correlated possible (Durier and Kobilinsky, 1989; Kobilinsky, 1989).

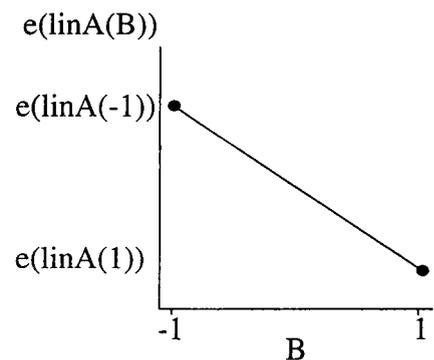
The values of the orthogonal polynomials and their products  $\text{lin}A \cdot \text{lin}B$ , ...,  $\text{cub}A \cdot \text{lin}B$  calculated on the 8 grid points are given in table II. These are the coefficients to apply to the 8 responses  $f(-3, -1)$ , ...,  $f(3, 1)$  to obtain the corresponding effects (up to normalization coefficients given in the last row). For instance:

$$e(\text{lin}A) = \langle \text{lin}A, f \rangle / 8\sqrt{5} = 1 / 8\sqrt{5} [-3 f(-3, -1) - f(-1, -1) + f(1, -1) + f(3, -1) - 3 f(-3, 1) - f(-1, 1) + f(1, 1) + f(3, 1)]$$

The normalization coefficients are chosen in order to obtain the same variance for all the columns in table II.



**Fig 2.** Polynomial factorial effects. The points stand for the theoretical values  $f(A, B)$ : \* for  $B = -1$  and + for  $B = 1$ .



**Fig 3.** Polynomial factorial effects. The slope is the  $e(\text{lin}A \cdot \text{lin}B)$  interaction.

**Table II.** Orthogonal polynomials defined for factor A at 4 levels and factor B at 2 levels and product orthogonal polynomials.

<i>f</i> (A,B)	A	B	Mean	<i>linA</i>	<i>quadA</i>	<i>cubA</i>	<i>linB</i>	<i>linA</i> • <i>linB</i>	<i>quadA</i> • <i>linB</i>	<i>cubA</i> • <i>linB</i>
<i>f</i> (-3, -1)	-3	-1	1	-3	1	-1	-1	3	-1	1
<i>f</i> (-1, -1)	-1	-1	1	-1	-1	3	-1	1	1	-3
<i>f</i> (1, -1)	1	-1	1	1	-1	-3	-1	-1	1	3
<i>f</i> (3, -1)	3	-1	1	3	1	1	-1	-3	-1	-1
<i>f</i> (-3, 1)	-3	1	1	-3	1	-1	1	-3	1	-1
<i>f</i> (-1, 1)	-1	1	1	-1	-1	3	1	-1	-1	3
<i>f</i> (1, 1)	1	1	1	1	-1	-3	1	1	-1	-3
<i>f</i> (3, 1)	3	1	1	3	1	1	1	3	1	1
Normalization coefficients			1	1/√5	1	1/√5	1	1/√5	1	1/√5
Polynomial effect			e(1)	e( <i>linA</i> )	e( <i>quadA</i> )	e( <i>cubA</i> )	e( <i>linB</i> )	e( <i>linA</i> • <i>linB</i> )	e( <i>quadA</i> • <i>linB</i> )	e( <i>cubA</i> • <i>linB</i> )

Their scalar products with the *f*(A,B)-column divided by 8 give the corresponding polynomial effects (multiplied by a normalization coefficient).

The polynomial effects are defined from the theoretical responses on the 8 initial grid points only. To interpolate between these points, *f* is generally assumed to have a polynomial form, for instance, a polynomial of degree 2 like:

$$f(A, B) = a_0 + a_1A + a_2A^2 + b_1B + c_1AB$$

This function can be written as a function of the orthogonal polynomials:

$$f(A,B) = \alpha_0 + \alpha_1 \text{lin}A + \alpha_2 \text{quad}A + \beta_1 \text{lin}B + \gamma_1 \text{lin}A \cdot \text{lin}B$$

The absence of B<sup>2</sup> and quadB indicates that B has no quadratic effect, an assumption which is coherent with the choice of only 2 levels for B. The coefficients  $\alpha_0, \alpha_1, \alpha_2, \beta_1, \gamma_1$  appearing in the latter form are then precisely the effects e(1), e(*linA*), e(*quadA*), e(*linB*), and e(*linA*•*linB*). The other effects e(*cubA*), e(*quadA*•*linB*), e(*cubA*•*linB*), associated with terms that are excluded from the model, are assumed to be zero.

Since the number of parameters in the model for *f* is less than the number of points in the grid, the polynomial effects can be estimated from a fraction of the 8 points instead of the whole set. This fraction must have at least as many points as the number of parameters in the model, if all effects are to be estimable. It must also be carefully chosen to make the estimation of effects as accurate as possible. The plain technique

described thereafter leads to very efficient fractions for situations involving a mixture of 2- or 4-level qualitative and quantitative factors.

**Decomposition in pseudofactors of a 4-level factor. Definition of (pseudo)factorial effects**

Fractional factorial design for 2-level factors are well known (Finney, 1945). They can be easily adapted to mixtures of 2-level and 4-level factors, by decomposing each 4-level factor into two 2-level pseudofactors. Moreover if a 4-level factor is quantitative, an appropriate decomposition leads to simple relations between the factorial effects associated with the pseudofactors and the polynomial effects, which make sense in that case. These relationships can be used to select a fraction well suited to the polynomial form assumed for the theoretical response.

In the example, the 4-level factor A is decomposed into two 2-level pseudofactors A<sub>1</sub> and A<sub>2</sub>. The correspondence between the quantitative levels of A and the levels of the pseudofactors A<sub>1</sub> and A<sub>2</sub> is given in table III. In this table we have also reported the levels of the product pseudo-factor A<sub>1</sub>A<sub>2</sub>, which are the products of the levels of A<sub>1</sub> and A<sub>2</sub>, and the corresponding values of the orthogonal polynomials of degree 1, 2 and 3 in A: *linA*, *quadA* and *cubA*.

The pseudofactors A<sub>1</sub>, A<sub>2</sub> and A<sub>1</sub>A<sub>2</sub> have no special meaning and are only used as tools to

build the design. The theoretical response  $f(A,B)$  can be expressed as a function  $g(A_1,A_2,B)$ , of the levels of the pseudofactors  $A_1$   $A_2$  and the factor B. The factorial effects (also called pseudofactorial effects) of  $A_1$ ,  $A_2$ , B and their interactions can be obtained from table IV in the same way as the previously introduced polynomial effects.

For instance, the factorial effect of  $A_1$ , denoted by  $e(A_1)$  is defined as the half difference between the means of  $g$  for  $A_1 = 1$   $g(1,*,*)$ , and for  $A_1 = -1$ ,  $g(-1,*,*)$ :

$$\begin{aligned}
 e(A_1) &= g(1,*,*) - g(-1,*,*) / 2 = 1/2 ([g(1,1,-1) \\
 &+ g(1,-1,-1) + g(1,1,1) + g(1,-1,1)] / 4) \\
 &- 1/2 ([g(-1,1,-1) + g(-1,-1,-1) \\
 &+ g(-1,1,1) + g(-1,-1,1)] / 4) \\
 &= 1 / 8 [g(1,1,-1) - g(-1,1,-1) + g(1,-1,-1) \\
 &- g(-1,-1,-1) + g(1,1,1) - g(-1,1,1) \\
 &+ g(1,-1,1) - g(-1,-1,1)] \quad [E1]
 \end{aligned}$$

Similarly, the effect of interaction between  $A_1$  and  $A_2$ , denoted by  $e(A_1A_2)$  is defined by:

$$\begin{aligned}
 e(A_1A_2) &= 1/2 ((\text{effect of } A_1 \text{ when } A_2 = 1) \\
 &- (\text{effect of } A_1 \text{ when } A_2 = -1)) = 1/2 \{([g(1,1,*) \\
 &- g(-1,1,*)] / 2 - ([g(1,-1,*) - g(-1,-1,*)] / 2)\} \\
 &= 1 / 8 [g(1, 1, -1) - g(-1, 1, -1) - g(1, -1, -1) \\
 &+ g(-1, -1, -1) + g(1, 1, 1) - g(-1, 1, 1) \\
 &- g(1, -1, 1) + g(-1, -1, 1)] \quad [E2]
 \end{aligned}$$

**Relationship between polynomial factors and pseudofactors and induced relationships between effects**

It is easily checked in table III that the following relationships between the polynomial factors  $linA$ ,  $quadA$ ,  $cubA$  on one side the pseudofactors  $A_1,A_2$  and  $A_1A_2$  on the other exist:

$$\begin{aligned}
 lin A &= (2A_1 + A_2) / \sqrt{5}; A_1 = (2linA - cubA) / \sqrt{5} \\
 quadA &= A_1A_2; A_1A_2 = quadA \\
 cubA &= (-A_1 + 2A_2) / \sqrt{5}; A_2 = (linA + 2cubA) / \sqrt{5} \quad [E3]
 \end{aligned}$$

**Table III.** Correspondence between levels of factor A, of pseudofactors  $A_1$ ,  $A_2$  and  $A_1A_2$ , and orthogonal polynomials of degree 1, 2 and 3 ( $linA$ ,  $quadA$  and  $cubA$ ).

A	$A_1$	$A_2$	$A_1A_2$	$linA$	$quadA$	$cubA$
-3	-1	-1	1	-3	1	-1
-1	-1	1	-1	-1	-1	3
1	1	-1	-1	1	-1	-3
3	1	1	1	3	1	1
				$1/\sqrt{5}$ <sup>a</sup>	$1$ <sup>a</sup>	$1/\sqrt{5}$ <sup>a</sup>

<sup>a</sup> In the last row, normalization coefficients for the orthogonal polynomials.

**Table IV.** Defining matrix for pseudo-factorial effects in the complete design.

$f(A,B) =$	$g(A_1,A_2,B)$	1	$A_1$	$A_2$	B	$A_1A_2B$	$A_1A_2$	$A_1B$	$A_2B$
$f(-3,-1) =$	$g(-1,-1,-1)$	1	-1	-1	-1	-1	1	1	1
$f(-1,-1) =$	$g(-1, 1,-1)$	1	-1	1	-1	1	-1	1	-1
$f(1,-1) =$	$g( 1,-1,-1)$	1	1	-1	-1	1	-1	-1	1
$f(3, -1) =$	$g( 1, 1,-1)$	1	1	1	-1	-1	1	-1	-1
$f(-3,1) =$	$g(-1,-1, 1)$	1	-1	-1	1	1	1	-1	-1
$f(-1,1) =$	$g(-1, 1, 1)$	1	-1	1	1	-1	-1	-1	1
$f(1,1) =$	$g( 1, -1, 1)$	1	1	-1	1	-1	-1	1	-1
$f(3,1) =$	$g( 1, 1, 1)$	1	1	1	1	1	1	1	1

This matrix is denoted by X in section II.5.

They induce the following similar relations between effects, for instance

$$e(\text{lin}A) = (2e(A_1) + e(A_2)) / \sqrt{5}$$

$$\text{and } e(A_1) = (2e(\text{lin}A) - e(\text{cub}A)) / \sqrt{5}$$

and they are easily extended to pseudofactorial effects built as products, for instance

$$e(A_1A_2B) = e(\text{quad}A \cdot \text{lin}B)$$

$$e(A_1B) = (2e(\text{lin}A \cdot \text{lin}B) - e(\text{cub}A \cdot \text{lin}B)) / \sqrt{5}$$

**Qualitative four-level factors**

Each qualitative factor A is also decomposed into 2 pseudofactors A<sub>1</sub> and A<sub>2</sub> but in that case the correspondence between levels is not relevant and can be arbitrarily chosen. For instance if a<sub>1</sub>, a<sub>2</sub>, a<sub>3</sub> and a<sub>4</sub> are the 4 levels of A, a possible choice is:

A	A <sub>1</sub>	A <sub>2</sub>	A <sub>1</sub> A <sub>2</sub>
a <sub>2</sub>	1	1	1
a <sub>3</sub>	-1	1	-1
a <sub>1</sub>	1	-1	-1
a <sub>4</sub>	-1	-1	1

This choice induces a decomposition of the main effect of A into 3 effects e(A<sub>1</sub>), e(A<sub>2</sub>) and e(A<sub>1</sub>A<sub>2</sub>) and similarly a decomposition of the interaction between A and B into 3 effects e(A<sub>1</sub>B), e(A<sub>2</sub>B) and e(A<sub>1</sub>A<sub>2</sub>B).

**Construction and properties of a fractional design**

The columns A<sub>1</sub>, A<sub>2</sub> and B of table IV gives the 8 treatments of a complete factorial design. Suppose we can only afford to experiment 4 of them. A simple way to select them is to chose those satisfying the equation A<sub>1</sub>A<sub>2</sub>B = 1, which is called a defining relation (Box *et al*, 1978).

**Table V.** Defining matrix for pseudo-factorial effects in the half fraction design.

	1	A <sub>1</sub>	A <sub>2</sub>	B	A <sub>1</sub> A <sub>2</sub> B	A <sub>1</sub> A <sub>2</sub>	A <sub>1</sub> B	A <sub>2</sub> B	
g(-1, 1,-1)	1	-1	1	-1	1	-1	1	-1	2
g( 1,-1,-1)	1	1	-1	-1	1	-1	-1	1	3
g(-1,-1, 1)	1	-1	-1	1	1	1	-1	-1	5
g( 1, 1, 1)	1	1	1	1	1	1	1	1	8

The fractional design thus defined, which is a half fraction, includes the treatments 2, 3, 5 and 8 of table IV which are singled out in table V.

Some elementary results on matrices are used in the following development. They are recalled in the *Appendix* for the use of the reader.

The general mean effect, denoted by e(1), the main effects of factors A<sub>1</sub>, A<sub>2</sub> and B and their interaction effects are defined from the theoretical responses on the complete factorial design by equalities similar to (E2) and (E1). If X' is the transpose of matrix X appearing in table IV, these definitions can be presented in matrix form as follows:

$$\begin{pmatrix} e(1) \\ e(A_1) \\ e(A_2) \\ e(B) \\ e(A_1A_2B) \\ e(A_1A_2) \\ e(A_1B) \\ e(A_2B) \end{pmatrix} = \frac{1}{8} \underbrace{\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 \\ -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \\ -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & -1 & 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \end{pmatrix}}_{X' \text{ matrix}} \times \begin{pmatrix} g(-1,-1,-1) \\ g(-1,1,-1) \\ g(1,-1,-1) \\ g(1,1,-1) \\ g(-1,-1,1) \\ g(-1,1,1) \\ g(1,-1,1) \\ g(1,1,1) \end{pmatrix}$$

Let e and g be the vectors appearing on the left and the right of this matricial expression respectively, that is the vector of pseudofactorial effects and the vector of theoretical responses. We can write more concisely:

$$e = 1/8 (X' g)$$

It is easy to check that XX' = 8I where I is the identity matrix. Premultiplication by of the 2 terms of the above equality then yields the equality:

$$g = Xe$$

which expresses the theoretical responses as functions of the pseudofactorial effects and can be written in the developed form as

$$\begin{pmatrix} g(-1,-1,-1) \\ g(1,-1,-1) \\ g(1,-1,1) \\ g(-1,-1,1) \\ g(-1,1,1) \\ g(1,-1,1) \\ g(1,1,1) \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & -1 & -1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & 1 & -1 & -1 & -1 & 1 \\ 1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}}_{X \text{ matrix}} \times \begin{pmatrix} e(1) \\ e(A_1) \\ e(A_2) \\ e(B) \\ e(A_1A_2B) \\ e(A_1A_2) \\ e(A_1B) \\ e(A_2B) \end{pmatrix}$$

The only responses estimable from the 4 points of the fractional design are:

$$g(-1, 1, -1), g(1, -1, -1), g(-1, -1, 1), g(1, 1, 1)$$

The 4 corresponding equations can be written in the following manner:

$$\begin{pmatrix} g(-1, 1, -1) \\ g(1, -1, -1) \\ g(-1, -1, 1) \\ g(1, 1, 1) \end{pmatrix} = \begin{pmatrix} 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix} \times \begin{pmatrix} e(1) + e(A_1A_2B) \\ e(A_1) + e(A_2B) \\ e(A_2) + e(A_1B) \\ e(B) + e(A_1A_2) \end{pmatrix}$$

From the estimations of the 4 responses, it is impossible to get separate estimates of the 2 factorial effects in 1 of the sums on the right, like  $e(1) + e(A_1A_2B)$ , or  $e(B) + e(A_1A_2)$ . If one of them was increased and the other decreased by the same amount and the observations are not affected. The 2 effects thus associated are said to be confounded or aliased.

The confounding of effects is due to the equality of the corresponding columns in table V:

$$1 = A_1A_2B, B = A_1A_2, A_1 = A_2B, A_2 = A_1B$$

These equalities can be directly obtained by multiplying the defining relationship  $1 = A_1A_2B$  of the fraction successively by  $B, A_1, A_2$ , since  $B^2 = 1$  for each of its levels, 1 or  $-1$ , and similarly  $A_1^2 = A_2^2 = 1$ .

The confounding of effects renders the interpretation of the statistical analysis very difficult, unless 1 of the 2 confounded effects is negligible. Thus if  $e(B) + e(A_1A_2)$  is found significantly greater than 0, it is impossible to determine without further information which of the 2 effects is greater than 0. However when  $A$  is quantitative,  $e(A_1A_2) = e(\text{quad}A)$ . Since quadratic effects can often be assumed to be far smaller than linear ones, the sum  $e(B) + e(A_1A_2) \approx e(B)$  can often be assimilated to the linear effect of  $B$ .

The unambiguous interpretation of all 4 sums of aliased effects in this simple example requires the strong assumption that the 3 degrees of freedom of the interaction between  $A$  and  $B$ , i.e.  $e(A_1A_2B)$ ,  $e(A_1B)$ , and  $e(A_2B)$ , are negligible as well as the quadratic effect of  $A$ .

The same method of fractionation leads to more interesting fractions when the number of factors is larger. It is then possible to confound important effects with interactions of 3 factors or more, or with effects of degree greater than or equal to 3. It is then quite realistic to assume that these last effects are negligible and this entails that all important effects can be well estimated.

### Complementary half-fraction. Fraction 1/4

It is possible in the example to select the other half of the design defined by  $A_1A_2B = -1$ . The confounded effects are the same but it is their difference which is estimated instead of their sum. For instance  $B = -A_1A_2$ , and so it is the difference  $e(B) - e(A_1A_2)$ , which can be estimated instead of the sum  $e(B) + e(A_1A_2)$ .

If a further fraction of the half-fraction defined by  $A_1A_2B = 1$  is required, a possible choice is to select the 2 treatments, out of the 4 of the half-fraction, which satisfy:

$$A_1A_2 = 1$$

This gives the 1/4 fraction of the complete factorial design defined by the 2 defining relations  $A_1A_2B = 1$  and  $A_1A_2 = 1$ . Making the product of these 2 defining relation gives:

$$\begin{aligned} A_1A_2B \times A_1A_2 &= 1 \text{ hence } B = 1 \\ \text{since } A_1^2 &= 1, A_2^2 = 1 \end{aligned}$$

Thus on this 1/4 fraction  $A_1A_2B = A_1A_2 = B = 1$ . Multiplying these equalities by  $A_1$  then gives  $A_2B = A_2 = A_1B = A_1$ . These equalities imply, exactly as in the half fraction, that the corresponding effects are aliased. More precisely from the two responses it is only possible to estimate the 2 sums  $e(A_1A_2B) + e(A_1A_2) + e(B) + e(1)$  and  $e(A_2B) + e(A_2) + e(A_1B) + e(A_1)$ . This example, in which the factor  $B$  remains constant, has clearly no practical interest. But a similar kind of fractionation going beyond the half fraction is very helpful when the factors are more numerous.

**GENERALIZATION: THE 1/8 FRACTION FOR IMPROVING THE CULTURE OF *B JAPONICUM***

**Construction**

We now present the 1/8th fractional design, which was obtained by the previously described method for the 7 factors in table I.

Two pseudofactors  $A_1$  and  $A_2$  are associated with the factor carbon source (denoted by A), with the following arbitrarily chosen correspondence between levels.

A	$A_1$	$A_2$
mannitol	-1	-1
glycerol	-1	1
gluconate	1	-1
glucose	1	1

The organic nitrogen concentration (B) and yeast extract concentration (C) are also decomposed into two-level pseudofactors  $B_1$  and  $B_2$  on one side,  $C_1$ ,  $C_2$  on the other. These factors are quantitative, and so the correspondence between levels is chosen as indicated in table III for the factor A of the previous example. The defining relations are:

$$A_1 B_1 B_2 C_1 C_2 E = A_1 B_2 C_2 D F$$

$$= -A_2 B_1 C_1 C_2 D G$$

$$= 1$$

The reasons for this choice will appear more clearly in the following sections. Premultiplications by E, F, G respectively give:

$$E = A_1 B_1 B_2 C_1 C_2 = A_1 B_2 C_2 D E F$$

$$= -A_2 B_1 C_1 C_2 D E G$$

$$F = A_1 B_1 B_2 C_1 C_2 E F = A_1 B_2 C_2 D$$

$$= -A_2 B_1 C_1 C_2 D F G$$

$$G = A_1 B_1 B_2 C_1 C_2 E G = A_1 B_2 C_2 D F G$$

$$= -A_2 B_1 C_1 C_2 D$$

To obtain the design, the 128 combinations of levels of the 7 pseudofactors  $A_1 B_1 B_2 C_1 C_2 D$  are first written down and for each of them the levels of E, F, G are deduced using the terms underlined in the above equalities.

Multiplying the 3 defining relations between these yields:

$$1 = A_1 B_1 B_2 C_1 C_2 E = A_1 B_2 C_2 D F = -A_2 B_1 C_1 C_2 D G$$

$$= B_1 C_1 D E F = -A_1 A_2 B_2 D E G = -A_1 A_2 B_1 B_2 C_1 F G$$

$$= -A_2 C_2 E F G \quad [E4]$$

For instance,  $A_1 B_1 B_2 C_1 C_2 E \times A_1 B_2 C_2 D F = 1$ , which implies  $B_1 C_1 D E F = 1$  since  $A_1^2 = B_2^2 = C_2^2 = 1$ .

The effects associated with the 8 equal products in [E4] are confounded. They are said to be confounded with the general mean  $e(1)$ . Consider now any product which does not appear in [E4], for instance  $A_1 B_1 B_2$ . Multiplying it by the equalities in [E4] gives:

$$A_1 B_1 B_2 = C_1 C_2 E = B_1 C_2 D F = -A_1 A_2 B_2 C_1 C_2 D G$$

$$= A_1 B_2 C_1 D E F = -A_2 B_1 D E G = -A_2 C_1 F G$$

$$= -A_1 A_2 B_1 B_2 C_2 E F H$$

The corresponding effects  $e(C_1 C_2 E) \dots e(A_1 A_2 B_1 B_2 C_2 E F G)$  are the effects confounded with  $e(A_1 B_1 B_2)$ . The choice of the defining relationships [E4] relies on the respective importance of the polynomial effects, which we now consider.

**Classification of polynomial effects**

The meaningful effects are the polynomial effects such as  $e(A_1 \cdot \text{quad} B)$ ,  $e(A_1 A_2 \cdot \text{lin} B)$  and  $e(\text{lin} B \cdot \text{lin} C \cdot D)$ , which are defined and computed as explained in the simpler example in the previous section. Two criteria may be used to classify these effects: the number of factors appearing in the effects and their degrees. The degree is the sum of the degrees in each factor; the degree attributed to qualitative factors is always 1. These 2 criteria are given below for the 3 polynomial effects just mentioned:

Effect	Factors involved	Number of factors	Degree
$e(A_1 \cdot \text{quad} B)$	A, B	2	$\text{degree}[e(A_1)]^* + \text{degree}[e(\text{quad} B)] = 1 + 2 = 3$
$e(A_1 A_2 \cdot \text{lin} B)$	A, B	2	$\text{degree}[e(A_1 A_2)]^* + \text{degree}[e(B)] = 1 + 1 = 2$
$e(\text{lin} B \cdot \text{lin} C \cdot D)$	B, C, D	3	$\text{degree}[e(\text{lin} B)] + \text{degree}[e(\text{lin} C)] + \text{degree}[e(D)] = 1 + 1 + 1 = 3$

\*  $\text{degree}[e(A_1)] = \text{degree}[e(A_2)] = \text{degree}[e(A_1 A_2)] = 1$  since A is qualitative.

The importance of an effect depends on its degree and on the number of factors in it. The more important effects are those of degree 1:  $e(A_1)$ ,  $e(A_2)$ ,  $e(A_1A_2)$ ,  $e(\text{lin}B)$ ,  $e(\text{lin}C)$ ,  $e(D)$ ,  $e(\text{lin}E)$ ,  $e(\text{lin}F)$  and  $e(\text{lin}G)$ . The effects of high degrees like  $e(A_1 \cdot \text{quad}B \cdot \text{quad}C)$  and  $e(A_1 \cdot D \cdot \text{lin}E \cdot \text{lin}F \cdot \text{lin}G)$ , which are of degree 5, are generally assumed to be negligible. For effects of the same degree, the importance generally decreases with the number of factors involved. For instance  $e(\text{cub}B)$  is considered more important than  $e(\text{lin}B \cdot \text{lin}C \cdot D)$ .

### ***Rationale for choosing the defining relationships and properties of the design***

The above classification can be used in conjunction with the relationships [E3] between pseudo- and polynomial factors to choose appropriate defining relations. There are 2 possible approaches. The approach described in this section (Bailey, 1982; Kobilinsky, 1985) was used to obtain the defining relations of this 1/8th fraction. It only relies on the presence or absence of each term in the relationships linking pseudofactorial and polynomial effects.

Pseudofactorial effects are categorized as negligible or not according to whether their expressions in function of the polynomial effects involve negligible terms only. The defining relationships are chosen in order to estimate all non-negligible pseudofactorial effects, by only confounding them with negligible ones. For instance, assume that the negligible polynomial effects are those of degree 3 or more. The estimability is then required for the 4 pseudofactorial effects  $e(B_1C_1)$ ,  $e(B_1C_2)$ ,  $e(B_2C_1)$  and  $e(B_2C_2)$  carrying information on  $e(\text{lin}A \cdot \text{lin}B)$ . These effects can be confounded with any negligible effect such as  $e(DEF)$ ,  $e(B_1C_1C_2) = (2e(\text{lin}B \cdot \text{quad}C) - e(\text{cub}B \cdot \text{quad}C)) / \sqrt{5} \dots$

The importance of polynomial effects depends on their degree. This suggests defining the degree of a pseudofactorial effect as the minimum degree of the polynomial effects as a function of which it can be expressed. For instance,

$$e(B_1) = (2 e(\text{lin}B) - e(\text{cub}B)) / \sqrt{5} \text{ is of degree 1}$$

$$e(C_1C_2) = e(\text{quad}C) \text{ is of degree 2}$$

$$e(B_1C_1C_2) = (2 e(\text{lin}B \cdot \text{quad}C) - e(\text{cub}B \cdot \text{quad}C)) / \sqrt{5} \\ \text{is of degree 3 (= 1 + 2)}$$

The degree of a pseudofactorial effect is in fact the sum of the degrees of each factor in it. It can be obtained here very quickly as the number of letters different from A plus 1 if it includes the letter A (once or twice).

The pseudofactorial effects can thus be classified as the polynomial effects as a function of their degree and of the number of factors they contain. If polynomial effects of degree greater than or equal to  $n$  and involving more than  $m$  factors are neglected, then so are the pseudofactorial effects of degree greater or equal to  $n$  involving more than  $m$  factors. It is therefore very simple to find the negligible pseudofactorial effects and to check the admissibility of a set of defining relationships.

For instance:

$$e(\text{lin}B \cdot \text{lin}C) = 1/5 (4e(B_1C_1) + 2e(B_1C_2) + 2e(B_2C_1) + e(B_2C_2))$$

The effects confounded with  $e(B_1C_1)$  and  $e(B_1C_2)$ ,  $e(B_2C_1)$  and  $e(B_2C_2)$ , obtained by multiplying the equalities [E4] by the corresponding effects, are interactions between three factors at least. If these interactions are assumed to be zero, these effects are estimable and so is  $e(\text{lin}B \cdot \text{lin}C)$ .

$e(\text{quad}B) = e(B_1B_2)$  is confounded with interactions of 4 factors or more and with 1 interaction  $A_1C_1C_2E$ , involve only 3 factors (of degree 4) that can be safely neglected.

More generally, it can be noticed that all the pseudofactorial effects confounded with the general mean, that is appearing in [E4], are of degree greater or equal to 5 (the design is said to be of polynomial resolution 5). This implies that: (i) the factorial effects of degree 1 are confounded with effects of degree at least 4; and (ii) the factorial effects of degree 2 are confounded with effects of degree at least 3.

Moreover, the main effects of any degree are confounded with interactions of at least 3 factors and at least degree 4, and interactions of 2 factors and degree 2 are confounded only with interactions of 3 factors except  $e(A_1 \cdot \text{lin}E)$ , confounded with a 2-factor interaction  $e(\text{quad}B \cdot \text{quad}C)$ . However, the latter is of degree 4 and can be assumed to be negligible. Interactions of 2 factors and degree 3 are also confounded with interactions of at least 3 factors, apart from  $A_1B_1B_2$ ,  $C_1C_2E$ ,  $A_1C_1C_2$  and  $B_1B_2E$  which are confounded by pairs.

The final properties come from the fact that the effects confounded with the mean are interactions of 5 factors except  $A_1B_1B_2C_1C_2E$  in which

appear only 4 factors. The smallest number of factors in the effects confounded with the general mean is the resolution of the design. The resolution here is only 4. It would have been better of course to use a resolution 5 design but this resolution cannot be obtained by the fractionation techniques considered in this section.

Another approach (Edmondson, 1991) goes 1 step further by considering the respective importance of the coefficients in the relations [E3] linking pseudofactorial and polynomial effects. We will show below how this second approach can still produce an efficient 1/16th fraction with 64 units instead of 128.

## REPETITIONS

In order to get the repeatability of results for fixed culture media, a subfraction of 16 treatments was repeated. This subfraction was obtained by adding the following defining relationships:

$$B_2 = 1, \quad C_2 = 1, \quad A_2B_1G = 1$$

These 16 points made up a resolution 3 fraction of the complete factorial design in which B and C were studied at 2 levels only (0.2–0.4 g/l for B, 2–4 g/l for C). Viable cells were counted for the 2 repetitions of these 16 repeated treatments in order to study the relationship between the optical density (OD) and the number of viable bacteria.

## STATISTICAL ANALYSIS

The polynomial model used for the analysis included all main effects, all 2-factor interactions of degree 2, some degree 3 estimable effects and finally 1 term out of each pair  $\{e(A_1B_1B_2); e(C_1C_2E)\}$ ,  $\{e(A_1C_1C_2); e(B_1B_2E)\}$  of confounded 2-factor interactions. Note that these last terms stand for the corresponding sums of confounded effects.

$$\begin{aligned} \text{OD} = & e(1) + e(A_1) A_1 + e(A_2) A_2 + e(A_1A_2) A_1A_2 \\ & + e(\text{lin}B) \text{lin}B + e(\text{quad}B) \text{quad}B + e(\text{cub}B) \text{cub}B \quad \left. \begin{array}{l} \text{Main effects} \\ \text{+ e(linC) linC + e(quadC) quadC + e(cubC) cubC} \\ \text{+ e(D) D + e(linE) linE + e(linF) linF + e(linG) linG} \end{array} \right\} \\ & + e(A_1 \cdot \text{lin}B) A_1 \cdot \text{lin}B + \dots + e(A_1 \cdot \text{lin}G) A_1 \cdot \text{lin}G \quad \left. \begin{array}{l} \text{Two-factor inter-} \\ \text{actions involving A} \\ \text{+ e(A}_2 \cdot \text{lin}B) A_2 \cdot \text{lin}B + \dots + e(A_2 \cdot \text{lin}G) A_2 \cdot \text{lin}G \\ \text{+ e(A}_1A_2 \cdot \text{lin}B) A_1A_2 \cdot \text{lin}B + \dots + e(A_1A_2 \cdot \text{lin}G) A_1A_2 \cdot \text{lin}G \\ \text{+ e(linB} \cdot \text{lin}C) \text{linB} \cdot \text{lin}C + e(\text{linB} \cdot \text{D}) + \dots + (\text{linF} \cdot \text{lin}G) \text{linF} \cdot \text{lin}G \end{array} \right\} \\ & \text{linear} \cdot \text{linear interactions} \\ & + \text{estimable or confounded degree 3 effects} \end{aligned}$$

The polynomial effects, the coefficients of this model, were estimated by least-squares approximation. From the standard deviations of their estimates, confidence intervals at different levels of confidence were deduced, using the *t*-Student distribution. In this model we selected all the terms that were significantly different from 0, *ie* having a confidence interval not covering 0. We added the terms of degree smaller or equal in each factor to them. For instance, if  $e(A_1 \cdot \text{quad}B)$  was significant, the terms  $\text{lin}B$ ,  $\text{quad}B$ ,  $A_1$  and  $A_1 \cdot \text{lin}B$  were introduced in addition to  $e(A_1 \cdot \text{quad}B)$ . The confidence level chosen was equal to 99% to avoid retaining a lot of terms that were negligible with respect to the most important effects in the model. The submodel thus obtained could be used, if it did not include any of the confounded effects, to predict the OD for any values of the factors within the experimental domain and in particular for each of the 1 024 treatments of the complete factorial design.

## 1/16th FRACTION

If polynomial effects of degree strictly greater than 2 are assumed zero, it is possible to fractionate one step further using Edmondson's method (1991, 1993), thus getting a 1/16 fraction with only 64 experimental units.

Possible defining relations for this 1/16 fraction are:

$$\begin{aligned} A_1B_1B_2C_1D &= A_2B_1B_2C_2E = A_2B_1C_1C_2F \\ &= A_1B_2C_1C_2G = 1 \end{aligned}$$

These defining relations were obtained by an automatic search program Planor (Kobilinsky, 1994). In its simpler form, the input for this program is a pseudofactorial model and the terms which are to be estimated in it. It is also possible to introduce several pseudofactorial models and the corresponding families of terms to be estimated. For this search, we required:

- i) all the main effects to be estimable in a model including all main effects and 2-factor interactions;
- ii) all 2-factor interactions, not including  $B_2$  or  $C_2$ , to be estimable in a model including the main effects and these interaction terms.

The symbolic expression of these requirements, as they appear in the program, was:

$$p: A + B_1 + C_1 + D + E + F + G$$

$$q: A + B + C + D + E + F + G$$

Requirement	Model	Terms to estimate
(i)	$q \bullet q$	$q$
(ii)	$p \bullet p$	$p \bullet p$

Note that a product like  $q \bullet q$  gives, after replacement of  $q$  by its expression, development and omission of redundant terms, a model including all main effects and 2-factor interactions. Requirement (i) thus means that the searched design is a resolution 4 design, and requirement (ii) is an empirical choice justified by the nature of the relationships between pseudofactorial and polynomial effects. This is better explained at the end of this section.

It was possible, in that case, to obtain all the sets of defining relations meeting the requirements. There were 1 152 such sets, which, after investigation, were found to be derived from 6 basic ones by permuting the factors with symmetric roles, that is D,E,F and G first then B and C and finally the 3 pseudo-factors  $A_1, A_2$  and  $A_1A_2$ .

Table VI gives these sets and, for each of them, 3 global efficiency criteria. These global criteria are defined by comparison between the normalized information matrix  $X'X/N$  of the design and  $X'_0X_0/N_0$  of the complete factorial design ( $N = 64, N_0 = 1\ 024$ ). It is convenient to choose the parametrization so that the latter is

the identity:  $X'_0X_0/N_0 = I$ . Then, if  $\lambda_i, i = 1, \dots, 45$  are the eigenvalues of  $X'X/N$ .

$$trace = \frac{45}{\sum_{i=1}^{45} 1/\lambda_i}, \det = (\prod_{i=1}^{45} \lambda_i)^{1/45}, \minval = \min_i \lambda_i.$$

The column No of permutations in table VI gives the number of other sets obtained by permuting the factors with symmetric roles. The best set for the trace and determinant criteria is the one proposed at the beginning of this section. We now study it more thoroughly by the same method as above.

Multiplying the 4 defining relations by each other gives the following 16 which correspond to all the pseudofactorial effects confounded with the general mean.

$$\begin{aligned} 1 &= A_2B_1B_2C_2E = A_1B_1B_2C_1D = A_1A_2C_1C_2DE \\ &= A_1B_2C_1C_2G = A_1A_2B_1C_1EG = B_1C_2DG \\ &= A_2B_2DEG = A_2B_1C_1C_2F = B_2C_1EF \\ &= A_1A_2B_2C_2DF = A_1B_1DEF = A_1A_2B_1B_2FG \\ &= A_1C_2EFG = A_2C_1DFG = B_1B_2C_1C_2DEFG \end{aligned}$$

Multiplication of the above terms by another one, say  $C_2D$ , gives all the effects confounded with  $C_2D$ .

**Table VI.** 1/16th fractions: defining relationships and global efficiency criteria.

Design	Defining relationship	No of permutations	Trace	Det	Minval
1	$A_1B_1B_2C_1D = 1; A_2B_1B_2C_2E = 1$ $A_2B_1C_1C_2F = 1; A_1B_2C_1C_2G = 1$	144	0.948	0.976	0.434
2	$A_1B_1B_2C_1D = 1; A_2B_1B_2C_2E = 1$ $A_1B_1C_1C_2F = 1; A_1A_2B_2C_1C_2G = 1$	144	0.934	0.969	0.460
3	$A_1B_2C_1D = 1; A_2B_2C_2E = 1$ $A_1B_1C_1C_2F = 1; A_1A_2B_1B_2C_1C_2G = 1$	288	0.912	0.958	0.460
4	$A_1B_2C_1D = 1; A_2B_2C_2E = 1$ $A_1B_1C_1C_2F = 1; A_1A_2B_2C_1C_2G = 1$	288	0.900	0.956	0.330
5	$A_1B_2C_1D = 1; A_2B_1C_2E = 1$ $A_1B_1B_2C_2F = 1; A_2B_2C_1C_2G = 1$	144	0.883	0.946	0.400
6	$A_1B_2C_1D = 1; A_1B_1C_2E = 1$ $A_2B_1B_2C_2F = 1; A_1A_2B_2C_1C_2G = 1$	144	0.874	0.944	0.330

$$\begin{aligned}
 C_2D &= A_2B_1B_2DE = A_1B_1B_2C_1C_2 = A_1A_2C_1E \\
 &= A_1B_2C_1DG = A_1A_2B_1C_1C_2DEG = B_1G \\
 &= A_2B_2C_2EG = A_2B_1C_1DF = B_2C_1C_2DEF \\
 &= A_1A_2B_2F = A_1B_1C_2EF = A_1A_2B_1B_2C_2DFG \\
 &= A_1DEFG = A_2C_1C_2FG = B_1B_2C_1EFG
 \end{aligned}$$

If factorial effects of degree strictly greater than 2 are neglected, the set of terms confounded with  $e(C_2D)$  reduces to  $\{e(C_2D); e(B_1G)\}$ . The only estimable linear combination of these 2 effects is their sum  $e(C_2D) + e(B_1G)$ . The estimate  $\hat{e}(C_2D)$  of this sum is  $\langle C_2D, Y \rangle / 64$ . Up to division by 64, this is the linear combination of the responses with the elements  $-1, 1$  of the vector  $C_2D = B_1G$  as coefficients. The variance of this estimate is  $\sigma^2/64$  where  $\sigma^2$  is the residual variance.

It can similarly be shown that the other sets of non-negligible confounded effects are  $\{e(B_2C_1); e(EF)\}$ ,  $\{e(C_2G); e(B_1D)\}$ ,  $\{e(B_1C_2); e(DG)\}$ ,  $\{e(B_2E); e(C_1F)\}$ ,  $\{e(B_1G); e(C_2D)\}$ ,  $\{e(C_1E); e(B_2F)\}$ .

The estimates  $\hat{e}(B_2C_1), \hat{e}(C_2G), \dots$  of the corresponding sums  $e(B_2C_1) + e(EF), e(C_2G) + e(B_1D), \dots$  have the same variance  $\sigma^2/64$  and are uncorrelated. By construction, all main effects are unconfounded, and so are the factorial effects of degree 2 which do not appear in the above sets of confounded effects, for instance,  $e(A_1B_1), e(DE)$ .

The equalities similar to [E3] for B and C and the assumed nullity of the effects of degree strictly greater than 2 lead to equalities of the following type:

$$\begin{aligned}
 e(B_1) &= 2e(\text{lin}B)/\sqrt{5} \\
 e(B_2) &= e(\text{lin}B)/\sqrt{5} \\
 e(B_1F) &= 2e(\text{lin}B \cdot F)/\sqrt{5} \\
 e(B_2F) &= e(\text{lin}B \cdot F)/\sqrt{5} \\
 e(B_1C_1) &= 4e(\text{lin}C \cdot \text{lin}B)/\sqrt{5}
 \end{aligned}$$

The last equality above is a consequence of:

$$\begin{aligned}
 B_1C_1 &= (4\text{lin}B \cdot \text{lin}C - 2\text{lin}B \cdot \text{cub}C - 2\text{cub}B \cdot \text{lin}C \\
 &\quad + \text{cub}B \cdot \text{cub}C)/5
 \end{aligned}$$

and of the nullity of the effects  $e(\text{lin}B \cdot \text{cub}C), \dots$  of degree strictly greater than 2.

Consider the estimation of a particular polynomial effect, say  $e(\text{lin}C \cdot D)$ . The pseudo-factorial effects depending on it are  $e(C_1D)$  and  $e(C_2D)$ . The latter is confounded with  $e(B_1G)$  which depends on  $e(\text{lin}B \cdot G)$  and so the unconfounded pseudofactorial effect  $e(B_2G)$  has to be taken into account too. The estimation of  $e(\text{lin}C \cdot D)$  thus involves the following system of equalities:

$$\begin{aligned}
 e(C_1D) &= (2e(\text{lin}C \cdot D))/\sqrt{5} \\
 e(C_2D) + e(B_1G) &= (e(\text{lin}C \cdot D) + 2e(\text{lin}B \cdot G))/\sqrt{5} \\
 e(B_2G) &= e(\text{lin}B \cdot G)/\sqrt{5}
 \end{aligned}$$

which in matrixial form become

$$\begin{pmatrix} e(C_1D) \\ e(C_2D) + e(B_1G) \\ e(B_2G) \end{pmatrix} = \begin{pmatrix} 1 & 2 & 0 \\ \sqrt{5} & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix} \times \begin{pmatrix} e(\text{lin}C \cdot D) \\ e(\text{lin}B \cdot G) \end{pmatrix}$$

or more concisely

$$e = Xp$$

The 3 coordinates  $\hat{e}(C_1D), \hat{e}(C_2D)$  and  $\hat{e}(B_2G)$  of the estimate  $\hat{e}$  of the left vector  $e$  are uncorrelated and of same variance  $\sigma^2/64$ . It follows (see any book on the linear model, for instance Draper and Smith, 1981) that

$$\hat{p} = (X'X)^{-1}X'\hat{e} \quad \text{with } \text{var}(\hat{p}) = \sigma^2/64 (X'X)^{-1}$$

In particular it is found that

$$\text{var}(\hat{e}(\text{lin}CD)) = 25 \cdot \sigma^2 / 21 \cdot 64$$

which can be compared with the variance that would have been obtained had both  $C_1D$  and  $C_2D$  been unconfounded. The estimate would then have been derived from the following equalities

$$\begin{aligned}
 e(C_1D) &= 2e(\text{lin}C \cdot D)/\sqrt{5} \\
 e(C_2D) &= e(\text{lin}C \cdot D)/\sqrt{5}
 \end{aligned}$$

and its variance would have been  $\sigma^2/64$ . The comparison of these 2 variances gives an efficiency  $21/25=0.84$  to estimate  $e(\text{lin}C \cdot D)$ . The efficiency can be computed similarly for the other polynomial effects. Table VII gives these efficiencies for the design considered above as well as for the following 2 in table VI. The second design could be preferred to the first if the BC interaction was judged *a priori* more important than some interactions between 2-level factors assigned to E and G. Note that in design 3, the 2-factor interaction effects including A are characterized by three efficiencies, called principal efficiencies, associated with the 3 degrees of freedom of each effect. The precise definition of these efficiencies can be found in Kobilinsky (1990). The smallest of the 3 efficiencies is the worst for any comparison between the 4 levels of these interactions.

The method used to obtain the efficiencies makes it clear that factorial effects including the pseudofactors  $B_1$  or  $C_1$  are more important than those including  $B_2$  or  $C_2$  for the estimation of the

**Table VII.** Factors efficiencies for polynomial effects in the 1/16th fraction.

Effect	Efficiency		
	Design 1	Design 2	Design 3
A	1	1	1
linB, linC	1	1	1
quadB, quadC	1	1	1
D, E, F, G	1	1	1
A•linB	1	1	1 0.840 0.800
A•linC	1	1	1 0.960 0.840
AD	1	1	1 1 0.833
AE	1	1	1 1 0.952
AF, AG	1	1	1 1 0.800
linB•linC	0.640	0.960	0.800
linB•D	0.840	0.800	0.840
linB•G	0.840	0.800	0.800
linB•E	0.840	0.800	0.960
linB•F	0.840	0.960	0.800
linC•D	0.840	0.960	0.840
linC•E, linC•F, linC•G	0.840	0.800	0.800
DE	1	0.800	1
DF	1	0.960	1
DG	0.810	0.800	1
EG	1	0.667	0.800
EF	0.810	0.800	0.800
FG	1	0.800	0.667

corresponding polynomial effects. This explains why requirement (ii) leads to good designs: the important 2-factor interactions, those which do not include  $B_2$  or  $C_2$ , cannot be confounded with one another.

The 1/16th design can still efficiently estimate all the polynomial effects of degree 1 or 2 under the assumption that those of degree 3 or more are negligible; this design is a resolution 4 fraction. However, restricting the number of tested treatments to 64 implies the loss of some interesting properties compared with the 1/8th design: variances of the polynomial effects are  $\sigma^2/64$  at the most in the 1/16th fraction, twice their variances in the 1/8th fraction; estimates of the different effects of degree 1 or 2 are not uncorrelated; and robustness, in the sense of main effects unconfounded with 3-factor interactions, is lost.

## RESULTS AND DISCUSSION

### *Reproducibility of the results*

The reproducibility of the results was found to be satisfactory, with a variation coefficient of the optical density of 7.5%.

### *Effects of factors on the optical density after 6 d culture*

There was no significant effect on the OD at the beginning of the culture. The effects found after 6 d culture could therefore be attributed to the influence of the different treatments on the development of the bacteria. The significant effects (at 99%) are reported in table VIII.

All factors, except ammonium chloride dose and carbon dose, had a significant effect on OD and interacted significantly with 1 or 2 other factors. To interpret, the responses predicted by the selected submodel (table VIII) were plotted against the values of the 5 factors kept in the model (pH, carbon source, organic nitrogen dose, yeast extract dose and organic nitrogen source).

In each graph of figure 4, for a given yeast extract dose, the OD is plotted against the different carbon sources. The points are connected by lines of different patterns to represent the different organic nitrogen doses. The pH and organic nitrogen source are fixed. These 4 selected graphs, associated with an initial pH of 6 and glutamate as organic nitrogen source, are those which lead to a maximal predicted OD.

**Table VIII.** Statistical analysis of the optical density after 6 d culture: selection of significant effects at 99%.

Effect		Effect estimate $\pm$ confidence interval at 99%
e(1)	Mean effect	1.77
e(A <sub>1</sub> )		-0.12 $\pm$ 0.031
e(A <sub>2</sub> )	Carbon source	+0.33 $\pm$ 0.033
e(A <sub>1</sub> A <sub>2</sub> )		+0.43 $\pm$ 0.031
e(linB)	Organic nitrogen dose	-0.17 $\pm$ 0.053
e(linC)	Yeast extract dose	-0.33 $\pm$ 0.031
e(D) *	Organic nitrogen source	-0.01 $\pm$ 0.031
e(linG)	pH	-0.26 $\pm$ 0.036
e(A <sub>1</sub> •linB) *		+0.03 $\pm$ 0.032
e(A <sub>1</sub> •linC)		+0.09 $\pm$ 0.035
e(A <sub>2</sub> •linC) *		+0.02 $\pm$ 0.038
e(A <sub>1</sub> A <sub>2</sub> •linC)		+0.19 $\pm$ 0.032
e(A <sub>1</sub> •D) *		-0.01 $\pm$ 0.032
e(A <sub>2</sub> •linG)		-0.17 $\pm$ 0.034
e(linB•D)		-0.06 $\pm$ 0.035
e(linC•linG)		-0.16 $\pm$ 0.035
e(quadC)		-0.17 $\pm$ 0.030
e(A <sub>1</sub> •linB•D)		+0.09 $\pm$ 0.035
e(A <sub>2</sub> •linC•D)		-0.16 $\pm$ 0.033

\* Not significant but included because of significant interactions that involve this effect. Residual standard deviation: 0.13 (30 degrees of freedom); coefficient of variation = 7.5%.

### Yeast extract dose

The linear effect of the yeast extract dose (C) was found to be important and negative: the OD decreases as the dose increases. This clearly appears in figure 4.

According to Elkan and Kwik (1968), the yeast extract at a relatively low concentration stimulates the growth of most strains. However, an excess of it may inhibit the growth of some *Rhizobium* (Date, 1972), and may even induce a cellular distortion (Skinner *et al*, 1977).

The level of response to the yeast extract appeared to depend on the strain: thus strain G49 studied here was found affected by doses of 3 and 4 g/l, while Sherwood's culture medium (1972) includes 5 g/l of yeast extract. We also detected a significant interaction with carbon source, examined in the next section.

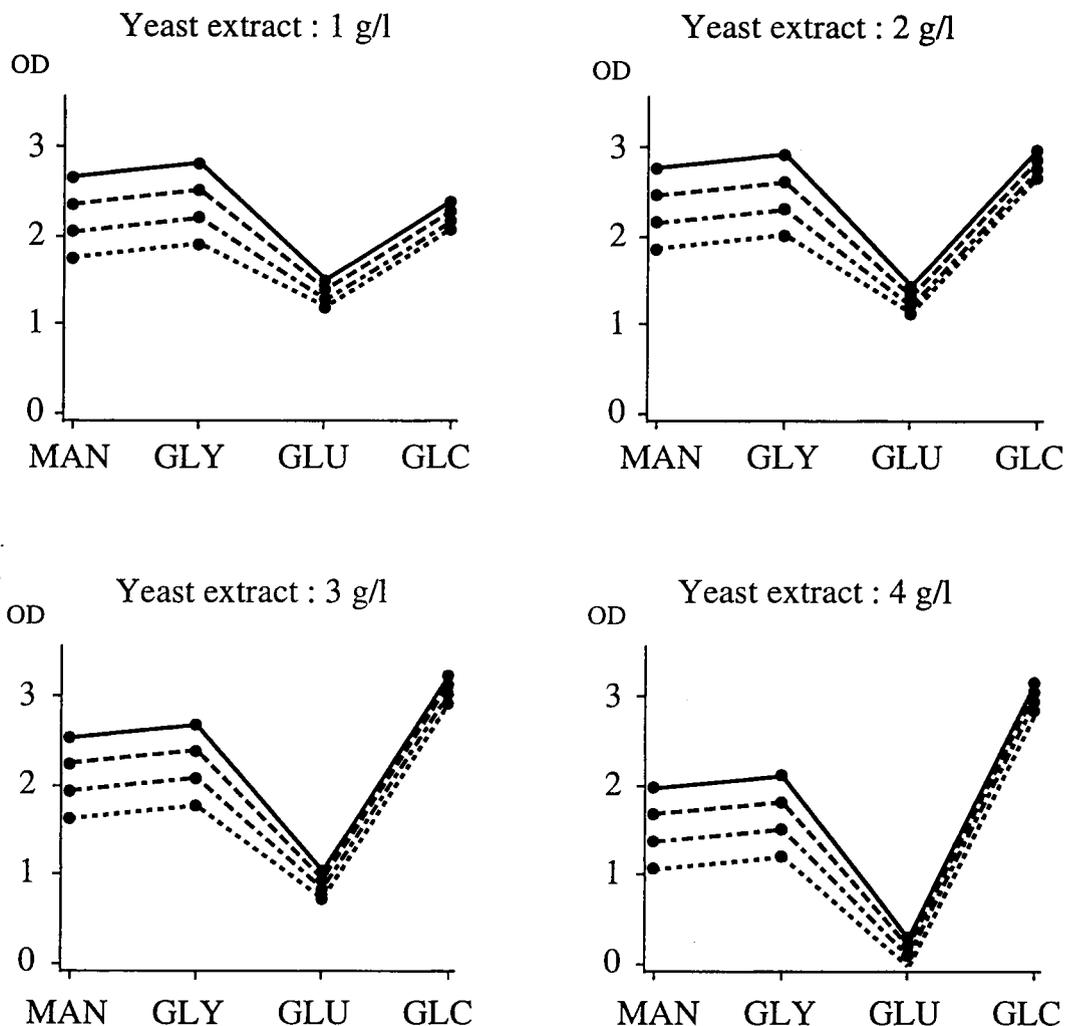
### Carbon source and yeast extract dose

With mannitol or glycerol, increasing the yeast extract dose from 1 to 3 g/l does not influence the response a great deal. However, the optical

density obviously decreases for these 2 substrates at 4 g/l. With glucose, the predicted response behaves differently; it first increases as the yeast extract varies from 1 to 2 g/l, and then remains the same even when the yeast extract dose reaches 4 g/l.

With gluconate as carbon source, the OD is always lower and also affected more by an increase in the yeast extract dose. The low OD in that case could be due to the important alkalization observed after 6 d culture in these media.

According to the literature, glycerol can be used by most strains (Elkan and Kwik, 1968); assimilation of glucose and manitol depends on the strain (Elkan and Kwik, 1968; Chakrabarti *et al*, 1981). Gluconate can also be recommended for the culture of *B japonicum* (Martinez de Drets and Arias, 1972). However, all these carbon sources were studied with a fixed amount of vitamin. The corresponding results did not take into account the interactions. They could have been very different had the carbon source been used with another substrate or with different compositions of the culture media.



**Fig 4.** Predicted values of OD after 6 d culture against the carbon source, organic nitrogen dose and yeast extract dose. pH = 6; organic nitrogen source = sodium glutamate; organic nitrogen dose: — 0.1 g/l; - - - 0.2 g/l; - · - · - 0.3 g/l; · · · · 0.4 g/l. Carbon source: MAN: mannitol; GLY = glycerol; GLU = gluconate; GLC = glucose.

### Source and dose of organic nitrogen

The main effects and interactions of the source (D) and dose (B) of organic nitrogen are relatively small. Though casein hydrolysate is the most used nitrogen source (Elkan and Kwik, 1968; Chakrabarti *et al*, 1981), OD is slightly greater in our results with glutamate.

The variation of OD due to the organic nitrogen dose is very small, much smaller than that generated by the increase of the yeast extract dose. This could be explained by the amount of organic nitrogen brought by the yeast extract (about 0.1 g nitrogen/g of yeast extract) which seemed to cover the nitrogen needs of the strain. This could also explain why mineral nitrogen from the ammonium chloride (0–26 mg/l of nitrogen) had no significant effect on the response.

It could be possible to stimulate bacterial growth by increasing organic and mineral nitrogen concentrations while lowering the yeast extract dose: Kwik (1973) recommends a medium including 0.1 g/l (10 mg of nitrogen/l) of yeast extract, with ammonium chloride (130 mg of nitrogen/l) and possibly glutamic acid (130 mg of nitrogen/l) as a source of organic nitrogen.

### Effect of the pH

In the graphs in figure 4, we have only reported the responses for an initial pH of 6. The important negative effect of pH (–0.26) comes from a general decrease of the OD at a pH of 7. The negative interaction CG in table VIII (–0.16) shows that this effect is still strengthened by an increase of the yeast extract concentration.

## Checking the results

Two culture media associated with the maximal turbidities at the 6th day were selected:

- 1) glycerol 10 g/l; sodium glutamate 0.1 g/l; yeast extract 2 g/l;
- 2) Glucose 10 g/l; sodium glutamate 0.1 g/l; yeast extract 2 g/l;

In both cases, the initial pH is taken equal to 6. The bacterial productions after 6 d culture were then  $1.5 \times 10^{10}$  cells/ml for medium 1 and  $2 \times 10^{10}$  cells/ml for medium 2, *ie* about 10 times the production in a classical culture media (Burton, 1967).

## On the choice of the growth measurement

The OD provided a very quick measurement allowing comparison of a great number of media. However the observed values of OD were greater than those (from 0.3 to 1) which are usually assumed to give a linear and reliable relation between the number of bacteria and the OD.

The composition of media, which can induce the secretion of several metabolites, or the production of polysaccharides, depending on the carbon substrate (Tully, 1985), could also explain the important variability leading to differences in the OD even when the numbers of bacteria are very similar.

In spite of this drawback, the results obtained in this study with the OD have been validated through confirmatory experiments. During additional experiments not reported here, the preferred method of measurement was the count of viable *B japonicum* in each medium tested (Cliquet, 1990).

## CONCLUSION

To cope with the variety of substrates metabolized by *B japonicum* and quoted in the literature, a great number of media were compared in a simple fractional design including different kinds of substrates at different concentrations.

This approach was well suited to simultaneously produce a comprehension of the effects of these factors and enough accuracy to study each of them. An experiment including 144 treatments handled simultaneously is however large and has high contamination risks, the more so as the growth of *B japonicum* is rather slow. So a more tractable solution for this kind of study could be to

divide the experiment into several blocks, each block including a smaller number of treatment to compare at the same time. The block effect is then taken into account in the analysis and does not increase the residual variance. Another solution is to perform sequentially smaller self-interpretable fractions like the one described above.

The results demonstrate once again the immense advantage of the old Fisherian principle (Fisher, 1935) over the one factor at a time approach, which unfortunately is still used too often nowadays. The interest and originality here is the integration of quantitative as well as qualitative factors in a fractional design which allowed the selection of a culture medium providing a high production.

## REFERENCES

- Balatti AP, Mazza LA (1978) Production of inoculants of legumes. *Rev Latinoam Microb* 20, 87-93
- Bailey RA (1982) The decomposition of treatment degrees of freedom in quantitative factorial experiments. *J R Statist Soc B* 44, 63-70
- Bergersen FJ (1961) The growth of *Rhizobium* in synthetic media. *Aust J Biol Sci* 14, 349-360
- Box GEP, Hunter WG, Hunter JS (1978) *Statistics for Experimenters*. John Wiley & Sons Inc, New York, USA
- Burton JC (1967) *Rhizobium* culture and use. In: *Microbial Technology*. Pepler and Rheingold, New York, USA, 1-33
- Chakrabarti S, Lee SM, Gibson HA (1981) Diversity in the nutritional requirements of strains of various *Rhizobium* species. *Soil Biol Biochem* 13, 349-354
- Cliquet S (1990) Production d'inoculum liquides de *Bradyrhizobium japonicum* : mise au point de milieu de culture, résistance à la dessiccation et efficacité des inoculum produits à noduler le soja. Thèse de doctorat, université Aix-Marseille II, France
- Date RA (1972) Sources and quantities of yeast extract for growth of *Rhizobia*. *J Appl Bact* 35, 379-387
- De Méo M, Laget M, Phan-Tan-Luu R, Mathieu D, Duménil G (1985) Application des plans d'expériences à l'optimisation des milieux et des conditions de culture en fermentation. *Biosciences* 4, 99-102
- Deshayes CMP (1980) Utilisation de modèles mathématiques pour l'optimisation en fermentation. Application aux transformations par les microorganismes. *Bull Soc Chim Fr* 1, 24-34
- Draper NR, Smith H (1981) *Applied Regression Analysis*. Second Edition. John Wiley & Sons Inc, New York, USA
- Duménil G, Cremieux A, Phan-Tan-Luu R, Combet M (1975) Bioconversion from DL-homoserine to L-thre-

- online. I. Application of factorial design. *Eur J Appl Microb* 1, 213
- Durier C, Kobilinsky A (1989) Apolor: une nouvelle vision de la régression polynomiale. *Actes XXI Jour Stat – ASU 93-97*
- Edmondson RN (1991) Agricultural response surface experiments based on four-level factorial designs. *Biometrics* 47, 1435-1448
- Edmondson RN (1993) Systematic row-and-column designs balanced for low order polynomial interactions between rows and columns. *J R Statist Soc B* 55, 707-723
- Elkan GH, Kwik IEM (1968) Nitrogen, energy and vitamin nutrition of *Rhizobium japonicum*. *J Appl Bact* 31, 399-404
- Fannin TE, Marcus MD, Anderson DA, Bergman HL (1981) Use of a fractional factorial design to evaluate interactions of environmental factors affecting biodegradation rates. *Appl and Envir Microb* 42, 6, 936-943
- Finney DJ (1945) The fractional replication of factorial arrangements. *Ann Eugen* 12, 291-301
- Fisher RA (1935) *The Design of Experiments*. Oliver & Boyd, Edinburgh, UK
- Fisher RA, Yates F (1957) *Statistical Tables for Biological, Agricultural and Medical Research*. Oliver & Boyd, Edinburgh, UK
- John PWM (1971) *Statistical Design and Analysis of Experiments*. Macmillan Company, London, UK
- Jordan C (1982) Family III - Rhizobiaceae. In: *Bergey's Manual of Systematic Bacteriology*, 1, Krieg, Williams and Wilkins, Baltimore, USA, 234-244
- Keele B, Hamilton P, Elkan G (1969) Glucose catabolism in *Rhizobium japonicum*. *J Bact* 97, 1184-1191
- Keele B, Hamilton P, Elkan G (1970) Gluconate catabolism in *Rhizobium japonicum*. *J Bact* 101, 698-704
- Kobilinsky A (1985) Orthogonal factorial designs for quantitative factors. *Statistics and Decisions, Suppl* 2, 275-285
- Kobilinsky A (1989) Tactiques en analyse de variance et en régression. *Rev Modulad* 1, 25-58
- Kobilinsky A (1990) Complex linear models and cyclic designs. *Linear Algebra Appl* 127, 227-282
- Kobilinsky A, Monod H (1991) Experimental design generated by group morphisms: an introduction. *Scand J Stat* 18, 119-134
- Kobilinsky A (1994) PLANOR: programme de génération automatique de plans d'expériences réguliers. Document interne, Laboratoire de Biométrie, INRA Versailles, France
- Kwik IEM (1973) Energy source, carbon and nitrogen nutrition of *Rhizobium japonicum* and their taxonomic significance. *Diss Abstr*, 33, 12-22
- Martinez de Drets, Arias A (1972) Enzymatic basis for differentiation of *Rhizobium* into fast- and slow-growing groups. *J Bact* 109, 467-470
- Pearson ES, Hartley HO (1976) *Biometrika Tables for Statisticians*. Charles Griffin and Company Ltd for the Biometrika Trust, London, UK

- Phan-Tan-Luu R, Bazard D, Flayeux R *et al* (1979) Applications de la méthodologie de la recherche expérimentale. Plans d'expériences factoriels et Simplex séquentiels. Exemples. *Ind Alim* 96, 441-451
- Phan-Tan-Luu R, Feneuille D, Mathieu D (1983) *Méthodologie de la recherche expérimentale. Étude des surfaces de réponse*. Laboratoire de prospective réactionnelle et d'analyse de l'information, IPSOI, Marseille, France
- Sherwood M (1972) Inhibition of *Rhizobium trifolii* by yeast extracts or glycine is prevented by calcium. *J Gen Microb* 71, 351-358
- Skinner FA, Roughley RJ, Chandler MR (1977) Effect of yeast extract concentration on viability and cell distortion in *Rhizobium* spp. *J Appl Bac* 43, 287-297
- Tully R (1985) New culture media to suppress exopolysaccharide production by *Rhizobium japonicum*. *Appl Microb Biotechnol* 21, 252-254

## APPENDIX: ORTHOGONAL POLYNOMIALS AND ELEMENTS OF MATRIX CALCULUS

For a 4-level quantitative factor, A and a 2-level quantitative factor, B with coded levels centered around 0, the orthogonal polynomials are defined (Kobilinsky, 1989) as:

$$\text{mean} = 1 \quad \text{lin}A = A/\sqrt{m_2} \quad \text{quad}A = \frac{A^2 - \frac{m_2}{m_0}}{\sqrt{m_4 - \frac{m_2 m_2}{m_0}}}$$

$$\text{cub}A = \frac{A^3 - \frac{m_4}{m_2} A}{\sqrt{m_6 - \frac{m_4 m_4}{m_2}}}$$

$$\text{lin}B = B / \sqrt{m'_2}$$

where

$$m_0 = 1,$$

$$m_2 = \frac{(-3)^2 + (-1)^2 + 1^2 + 3^2}{4} = 5,$$

$$m_4 = \frac{(-3)^4 + (-1)^4 + 1^4 + 3^4}{4} = 41,$$

$$m_6 = \frac{(-3)^6 + (-1)^6 + 1^6 + 3^6}{4} = 365$$

$$\text{and } m'_2 = \frac{(-1)^2 + 1^2}{2} = 1$$

A matrix is a table of numbers such as

$$\mathbf{M} = \begin{pmatrix} 1 & -2 & 1 \\ 3 & 6 & 4 \end{pmatrix}$$

$\mathbf{M}$  is a 2-rowed, 3-columned matrix.

The following system of equations

$$\begin{cases} y_1 = x_1 - 2x_2 + x_3 \\ y_2 = 3x_1 + 6x_2 + 4x_3 \end{cases}$$

can also be put under the following matricial form:

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 1 & -2 & 1 \\ 3 & 6 & 4 \end{pmatrix} \times \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

$$\mathbf{y} = \mathbf{M} \times \mathbf{x}$$

We have put a symbol  $\mathbf{x}$ , though it is usually omitted, to represent the matricial product

(whose element in row  $i$ , column  $j$  is by definition the sum of products between corresponding elements in the  $i$ th row of the left matrix and the  $j$ th column of the right).

The product of a matrix by a scalar (real number) is obtained by multiplying each element in it by this scalar:

$$3 \times \mathbf{M} = \begin{pmatrix} 3 & -6 & 3 \\ 9 & 18 & 12 \end{pmatrix}$$

The transpose of  $\mathbf{M}$  is denoted by  $\mathbf{M}'$ . It is the matrix whose columns are the rows of  $\mathbf{M}$ . Thus the transpose of  $\mathbf{M}$  is  $\mathbf{M}'$ .

$$\mathbf{M}' = \begin{pmatrix} 1 & 3 \\ -2 & 6 \\ 1 & 4 \end{pmatrix}$$

The identity matrix is a square matrix (as many rows as columns) whose diagonal elements are 1 and other elements 0. For instance the 3 x 3 identity matrix is:

$$\mathbf{I} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

For any matrix  $\mathbf{A}$  having the same number of rows and columns as the identity matrix  $\mathbf{I}$ ,  $\mathbf{A} \times \mathbf{I} = \mathbf{I} \times \mathbf{A} = \mathbf{A}$ .