FRACTIONAL FACTORIALS IN A CASE STUDY NUTRITION EXPERIMENT WITH BANANA TREES

Paulo César Moraes RIBEIRO[1]
Matheus Pena CAMPOS[2]
Leila Aparecida Salles PIO[2]
Júlio Sílvio de Sousa BUENO FILHO[1]

■ ABSTRACT: In this paper we study combining designs concatenating levels from a full factorial for some factors with screening alternatives for the others. This was done to deal with a practical situation in plant nutrition experiments. The original problem was a study design for 14 potential factors in banana tree nutrition, and researchers imagined four full factorials were needed to test their hypothesis, being two from the 3^3 and two of the 3^4 series. As this would demand at least 216 experimental units and facing limited resources we seek for a different planning strategy. The idea was to combine in the same experiment four instances of DSD (*Definitive Screening Designs*) for 10 three-level factors, each in a different block, with a fraction of the full factorial of the 3^4 series. A central point treatment, with average level for all factors, was present in all blocks. Interchange algorithms were used to concatenate the factor levels. Resulting optimized design was compared to the designs sampled following the same principle. Design comparison criterion was the expected average variance of the estimates for factors (A_r optimality). Optimization reduced 4.02% of the average values of the criterion in a reference population of sampled designs. It was possible to show that the variance for linear and quadratic effects in the full factorial were higher than in the optimized plan. As an example, the analysis of an actual field trial is presented. Authors recommend the use of fractional factorial strategy including DSD designs in agronomic trials, specially in the screening phase.

■ KEYWORDS: Interchange algorithm; combining designs; screening designs.

1 Introduction

A common problem in many sector of economic activity, including agriculture, is the search for process variables that has greatest potential to change production levels. Given that dozens of candidate factors can manifest themselves at different expression levels, combinations of all possible factors' levels make experimentation unfeasible. Since the 1950's, however, some important concepts of experimentation strategy has been developed to tackle specific problems, as opposed to use standard designs for all kind of experiments (MEAD et al, 2012).

In general, we expect from most of the factors to be relatively irrelevant and a few to be the important ones to tune the production process. In this sense, repeating levels of the non important factors are simply helping inference on the relevant ones. Screening designs had been proposed to help finding those causal variables using sensible fractions of full factorials.

In many cases, screening designs are saturated or supersaturated, as they have more study factors than possible runs. Some examples are resolution II designs and those proposed by Plackett and Burman (1946). Those designs use only two levels of each factor, frequently added by a central point. This makes second order effects and factor interactions confounded or non-estimable.

To deal with this, Jones and Nachtcheim (2011) proposed a new class of trial designs, the so called *definitive screening designs* (DSD). Such plans require only $2m + 1$ runs for m factors. There is no need of orthogonality between pairs of main effects, as required in Cheng and Wu (2001). Tsai *et al.* (2000) presented essentially the same properties, but did not emphasize the patters in resulting designs. The construction method from Jones and Nachtsheim (2011) is slightly different, although also based on an optimality criterion and a search algorithm.

DSDs has been investigated since this seminal paper and there are many proposals for extensions for blocking and different numbers of factors as well as discussions on algorithms, and combinatorics of DSDs (JONES and NACHTSHEIM, 2011a, 2013; LIN, 2015; WANG et al., 2015; YANG et al. 2017). In the case of experiments with more than five factors, DSD estimate the quadratic model with no interactions with a high efficiency, due to small aliasing among first order interactions and quadractic terms. For some designs it is also possible to estimate some first order interactions and this is associated with better estimation of main effects.

As it is always the case in a fractional factorial analysis, conditioning on the absence of causal main effects or quadratic terms, other interactions benefit from smaller aliasing in the new design matrix.

An additional problem to make those designs used in agronomy is the traditional reliance in full factorial experiments. To convince practical experimenters, it still need to be shown with actual examples that fractional factorials are viable, with as powerful an inference, a better local control and for smaller cost. To achieve such aim we present a study design to infer on nutrition factors in banana trees.

In this study, 14 potential nutrition factors were tested. The initial proposal was to test research hipothesys with four (4) full factorial designs, being two of the

3^3 and two of the 3^4 series. This would need at least 216 experimental units (e.u.). This would make the experiment unfeasible due to restrictions in area and seedlings available.

We decided to plan for the limited resources combining DSD ideas for 10 of the factors and a full factorial of the 3^4 series for the remaining four. Experimental area was divided in four blocks, each receiving a DSD in which one treatment has both central point of Full factorial and DSD, but remaining e.u. receive a combination of non-central points from a DSD and the full factorial.

The methodology to combine DSD and full factorial is presented in next section. It follows the structure of resulting combined design and a brief description of the search algorithm. Resulting distribution for the optimality criterion in the population of designs is compared with values found after searching. Some efficiency comparisons with factors from the full factorial are also presented. The analysis of a continuous variable from the actual experiment that used this design is given as an example.

2   Methodology

For the results in this paper we implemented an interchange algorithm to combine DSD (fractional factorials of the 3^{10-6} series) with the full factorial (3^4). The computer search has a criterion that minimize the variance of estimates for main effects and quadratic terms of all factors in DSD and for main effects, quadratic and first order interaction between pairs of factors in the full factorial. Basic code presented in the Apendix was developed in R (R CORE TEAM, 2018). Additional code and explanation, if needed, can be shared by the authors.

2.1   Initial design and search algorithm

From an initial design with 84 lines (21 per block) we reserved one line in each block for the central point (14 zeroes) and the remaining 80 lines (20 per block) for a combination of DSD_{10} and full factorial (3^4) non central points. The initial allocation was at random and those are already combination of locally optimized designs with the full factorial. So there is a population of such designs for each the criterion can be evaluated. The distribution of those values can be seen as a reference for combined designs and was used to infer if computer search can improve on original proposed designs.

Optimality criterion adopted considered the average variance of all estimable effects, corrected by local control (blocks). Thus, we evaluated the trace of corresponding submatrix from covariance matrix without rows and columns associated to blocks (A_r optimality, GOOS and GILMOUR, 2017; TRINCA and GILMOUR, 2015).

For the reference distribution approximation we used a random sample from the 80! ($\sim 7,16 \times 10^{18}$) possible designs (already good) obtained by the concatenation of DSD with levels of the full factorial. Sample size was $N = 10 \times 10^6$ and the

best design from this sample was submitted to an interchange algorithm. Resulting design is then considered the best permutation to estimate desired model.

In the actual experiment there was a mistake in the identification (and assemblage of) one treatment from the full factorial. Nevertheless, criterion value and the efficiency for this design and the variance of estimates for the full factorial and corresponding terms in DSD were also evaluated.

2.2 Combined design structure

Table 1 brings the basic structure of the design. From the full factorial of the 3^4 series, fixing on central point for each block, we proceed to interchange remaining 80, keeping, for each block, a fixed structure for a DSD_{10} for the other factors. Resulting design has $t = 81$ treatments, but just one (the central point for all factors) is repeated (4 times). Total numberof experimental units is $n = 84$ arranged in $b = 4$ blocks of size $k = 21$.

All 14 factors represent suplements for mineral nutrition of banana trees, previously selected by the research group, being seven biostimulators: "Multiturbo" (MTU), "Acadian" (ACA), "PDQ Perenes" (PDQ), "Biof-Nature1" (BKP), "Biof-Nature2" (BF2), "Litho-Natus" (LTL) and "Vitakelp" (VTK), seven soil conditioners (B): "Shell Limestone" (CC), "Agro Silicon" (AG), "Eggshell" (CO), "Dolomite Limestone" (CD), "Celtonite" (CT), "Seaweed" (HA) and "Lithotane" (LT). To make for a simples presentation of the design we represent all 14 levels in the Table 1 by alphabetic sequence "A", "B", $\cdots$, "N".

Table 1 - Combined Design Structure: for each block a fraction of the full factorial of the 3^4 series is combined to a DSD_{10}

Block	E.U.	Factors from 3^4 series A B C D	Factors from DSD_{10} E F G H I J K L M N
1	1 to 20	20 out of 80 treatments from 3^4 series	$DSD_{10.1}$
	21	0 0 0 0	0 0 0 0 0 0 0 0 0 0
2	22 to 41	next 20 from 3^4 series	$DSD_{10.2}$
	42	0 0 0 0	0 0 0 0 0 0 0 0 0 0
3	43 to 62	next 20 from 3^4 series	$DSD_{10.3}$
	63	0 0 0 0	0 0 0 0 0 0 0 0 0 0
4	64 to 83	next 20 from 3^4 series	$DSD_{10.4}$
	84	0 0 0 0	0 0 0 0 0 0 0 0 0 0

3 Results

3.1 Starting design and distribution for optimality criterion values

In Figure 1 is depicted the reference distribution for the criterion evaluated for each design in the sample of size 10×10^6. Note that this is a random sample of designs in which all factors are estimable. In some sense they are all good designs, with better properties than usual full factorial choices. The average of distribution is 0.4707 and the best design has variance 0.4549 (the smallest value). we depicted this value with a thick vertical line in the graph. Best design that can be found by chance is 3.47% more efficient than the average of good designs. A typical design assembled by hand with the researchers (Basic Plan) will also have good properties, with criterion value slightly lower than average (0.4623), i.e. 1.81% more efficient than the expected random choice.

After the search algorithm, the best design found has criterion value 0.4525. This value is marked with a thin vertical line and represents a design 4.02% more efficient than expected by a random choice of a good design.

This implies that a sensible initial suggestion for researchers, following good experimental considerations, i.e., trying to combine contrasting levels in the same block, could be found without search algorithms, only using statistical considerations such as orthogonality and balance. However, a better design could be found by a random search and an even better design can be (almost) always found using interchange algorithms

It is commonplace that discrete interchange algorithms may not result in global optima, however, the best design found repeat consistently the criteriom values from different starting points. This value is also very unlikely in the random sample of good designs (in this case, of course, $p < 10-7$).

This result is a clear indication that we could enhance the efficiency of fractional factorials even when odd restrictions from researchers are imposed, such as the need to have a full factorial design for some factors (for comparison purposes), or some unexpectedly inconvenient block size, for instance. In this case we used blocks with $k = 21$ and $n = 84$, numbers that were very interesting to have a good control of experimental conditions for sucha great number of factors.

This is also an indication that a definitive screening design with 14 columns (DSD_{14}) would be a suitable solution for the search of causal factors, and that other optimal designs would also be useful and could use less experimental effort (and money). This means that on optimized fractional factorial could be built for smaller number of units with possibly better local control, if not for initial restrictions.
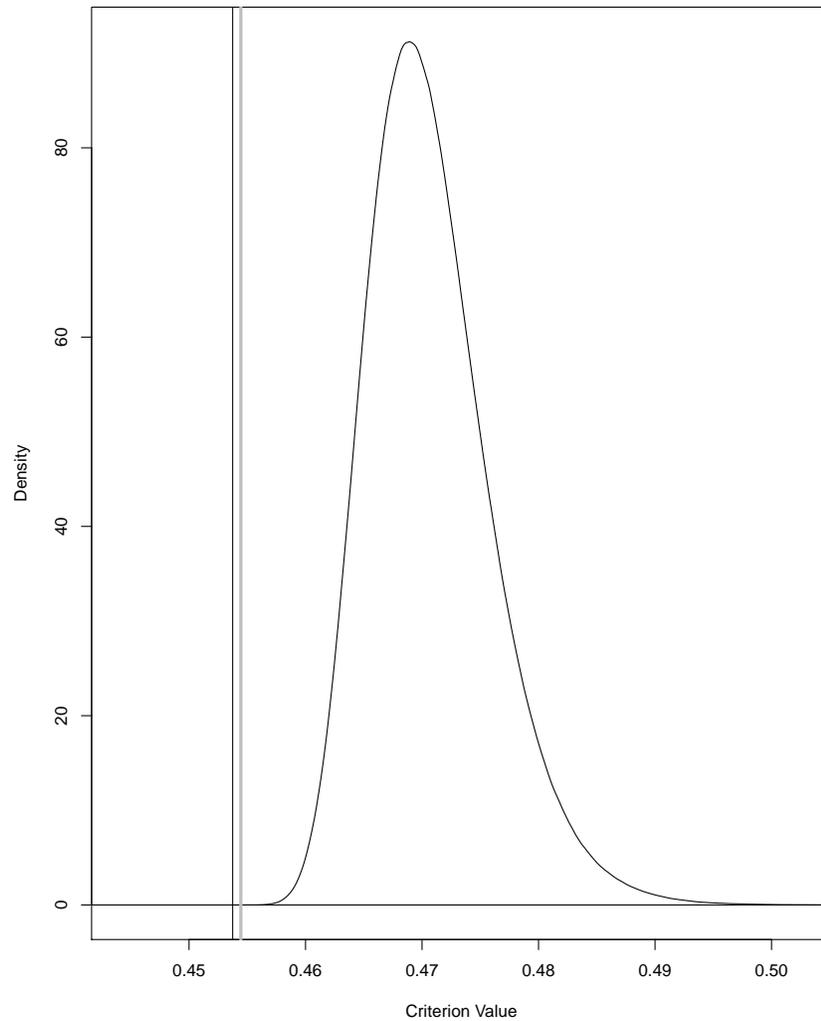
Figure 1 - Frequency density of optimality criteriom values evaluated in a sample of combined designs constructed as a combination of full factorial of 3^4 series and four DSD_{10}. The two vertical lines represent the criterion evaluated for the best design found at random (gray line) and the best design found by search algorithm (black line to the left).

3.2 Efficiency for partial models in resulting design

In table 2 is described an interesting aspect of the results, that is the efficiency of fractions from full factorial within each block as compared to factors from DSD_{10}.

This might help to enlighten the potential of using Fractional factorials at smaller costs, keeping low variances for the desirable estimates. This makes for a better hypothesis testing in the search for causal factors.

Table 2 - Average variance of estimates for estimable factors in each block. Designs considered are of a design organized by hand (Basic Plan) and the best found by search algorithm (Optimized Design). Efficiency of the latter is evaluated within each block

| | Experimental Design | | |
Bloco	Basic Plan	Otimized Design	Efficiency
I	2.50	2.26	1.1062
II	2.44	2.20	1.1091
III	2.58	2.24	1.1518
IV	2.65	2.33	1.1373

As for Block I, optimized design is 10.62% more efficient than the Basic Plan. In all four blocks efficiency is even higher. This means that average variance of estimates from optimized design is smaller for every block. In particular, this is due to a better concatenation of rows from full factorial and DSD_{10}. The initial reason for combining those levels was to show researchers that screening causal factors could be done without repeated full factorial designs (point that is already extensively made in the literature, e.g. Box, Hunter and Hunter, 2005). On the other hand, design strategy has a potential to make screening designs widely used in agronomic trials. A good combined design could also be assembled by hand, without relying on computer search. As a general indication, however, a DSD_{14} would be the optimal design for screening with small variance. This design would use just $n = 29$ experimental units (or at least $n = 30$ to allow for replication of the central point, or how much possible for more than 2 replications of it).

3.3 Efficiency of actual design (sub-optimal)

Due to treatment identification problems (a change in a composition of a treatment), the actual field design was changed, this brought larger impact on the full factorial than in the DSD factors, as expected. The lack of efficiency could be easily evaluated. For 4 factors in 3^4 series average variance rise to 2.6953 (from 2.3811 in the original plan). This means actual design is 88.4% efficient as compared to planned design. For the DSD, of course, efficiency did not change. The estimability for quadratic model was unnafected.

The covariance of estimates for quadratic effects in the actual experiment was also evaluated and is presented in the following matrices:

1. Full factorial

$$\begin{bmatrix} 5 & 0 & 1 & -2 \\ 0 & 5 & -2 & 2 \\ 1 & -2 & 24 & -16 \\ -2 & 2 & -16 & 24 \end{bmatrix},$$

2. Fractional factorial

$$\begin{bmatrix} 5 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 \\ 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 5 \end{bmatrix}.$$

This means that simpler models were better estimated (uncorrelated estimates) by the fractional factorial and this strategy is clearly recommended.

In Table 3 is presented the optimized design resulting from interchange algorithm. In the last row, the average bias for each factor is evaluated. Of course, we yield an unbiased design, but what happen in the actual experiment was biased, with changes most factors. As an illustration, the analysis for one of the variables is presented in the next section.

4 Example for plant heigh

The experiment was evaluated for many agronomic traits, such as plant height, diameter of pseudocaule, number of leaves by plant, number of tillers per plant, etc. The evaluation was done in two seasons, we decide to show an example of a continuous trait (plant height) evaluated in last season. The main purpose of this example is to illustrate the simplicity of ANOVA and associated tests for hypotheses on linear and quadratic effects.

Data and model passed ordinary normality tests such as Box and Cox (1964), performed with R package MASS (VENABLES e RIPLEY, 2002). There was no need to transform original data.

In ANOVA summary (Table 4) we verify the relevance of main effects for CT and CD as well for AG:LT interaction .

Table 3 - Optimized design from interchange algorithm. In the last row it is evaluated the average bias

Block	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	1	0	-1	1	0	1	1	-1	1	1	1	1	-1	1
1	1	-1	0	-1	0	-1	-1	1	-1	-1	-1	-1	1	-1
1	1	1	0	0	1	0	-1	1	1	-1	1	1	-1	-1
1	0	-1	-1	1	-1	0	1	-1	1	-1	-1	-1	-1	1
1	-1	-1	0	1	-1	1	0	-1	-1	-1	1	-1	-1	-1
1	1	-1	1	-1	1	-1	0	1	1	-1	1	1	1	1
1	-1	1	1	0	-1	1	1	0	1	-1	-1	1	1	-1
1	0	-1	1	-1	1	-1	-1	0	1	1	-1	-1	-1	1
1	1	0	-1	0	-1	-1	-1	-1	0	1	1	1	-1	-1
1	-1	1	-1	-1	1	1	1	1	0	-1	-1	-1	-1	1
1	1	1	1	1	-1	-1	1	1	1	0	1	1	-1	-1
1	-1	0	1	1	1	-1	1	-1	-1	0	-1	1	-1	-1
1	0	1	-1	-1	1	1	1	-1	-1	-1	0	-1	-1	-1
1	0	0	1	1	-1	-1	1	1	1	1	0	-1	-1	-1
1	0	1	0	0	1	1	1	1	-1	1	1	0	1	-1
1	-1	0	-1	1	-1	-1	-1	-1	1	-1	-1	0	-1	1
1	0	0	0	-1	1	1	1	1	1	1	-1	-1	0	-1
1	-1	0	-1	0	-1	-1	1	-1	1	-1	1	1	0	1
1	1	-1	-1	0	-1	1	-1	1	-1	1	-1	-1	-1	0
1	-1	-1	0	-1	-1	1	-1	1	-1	1	-1	1	-1	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	1	1	0	1	0	1	1	-1	1	1	1	1	-1	1
2	0	-1	0	-1	0	-1	-1	1	-1	-1	1	1	-1	-1
2	0	1	1	1	1	0	1	-1	-1	-1	1	-1	1	1
2	-1	-1	-1	0	-1	0	1	-1	-1	-1	1	-1	1	1
2	-1	0	0	-1	1	0	0	-1	1	-1	1	-1	1	-1
2	-1	1	0	1	1	-1	0	1	1	-1	-1	1	1	1
2	0	0	1	-1	-1	1	1	0	1	-1	-1	1	1	-1
2	0	1	-1	-1	-1	-1	1	0	-1	1	-1	-1	-1	-1
2	1	1	1	0	-1	-1	-1	-1	0	1	1	-1	-1	1
2	-1	0	0	0	-1	1	-1	1	0	-1	1	-1	1	1
2	-1	0	0	1	-1	1	-1	1	1	0	1	-1	1	1
2	0	0	-1	-1	-1	1	-1	-1	-1	0	-1	1	1	-1
2	1	-1	-1	1	1	-1	-1	-1	-1	-1	0	1	1	1
2	-1	-1	1	1	-1	-1	1	1	1	1	0	-1	1	-1
2	-1	1	-1	0	1	1	1	1	1	-1	1	0	1	-1
2	-1	-1	1	0	-1	-1	1	-1	-1	-1	-1	0	-1	1
2	0	1	1	-1	1	-1	1	-1	-1	1	-1	-1	0	-1
2	0	0	-1	1	-1	-1	1	-1	1	-1	1	1	0	1
2	1	-1	0	0	1	-1	1	-1	1	-1	1	-1	1	0
2	1	1	-1	-1	-1	1	-1	1	-1	1	-1	1	-1	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	1	-1	1	0	0	1	1	-1	1	1	1	1	-1	1
3	1	0	0	0	0	-1	1	-1	-1	-1	-1	-1	1	-1
3	-1	-1	0	0	1	0	-1	1	-1	-1	-1	-1	-1	-1
3	1	0	1	1	-1	0	1	-1	-1	-1	1	-1	-1	-1
3	0	1	-1	1	-1	1	0	-1	-1	-1	1	-1	-1	-1
3	0	0	1	0	1	-1	0	1	1	1	1	1	1	-1
3	0	-1	1	0	-1	1	1	0	1	-1	-1	1	1	-1
3	-1	0	-1	-1	-1	-1	-1	0	-1	1	-1	-1	-1	-1
3	0	1	0	-1	-1	-1	-1	-1	0	1	1	1	-1	-1
3	-1	1	1	1	-1	1	-1	1	0	-1	1	-1	1	1
3	-1	-1	-1	1	-1	1	-1	1	1	0	1	-1	1	1
3	1	-1	-1	-1	1	1	-1	-1	-1	-1	0	1	1	1
3	1	1	1	-1	1	1	-1	-1	-1	-1	0	1	1	-1
3	0	0	-1	0	-1	1	1	1	-1	1	1	0	-1	-1
3	1	0	-1	-1	1	1	1	1	-1	1	1	0	-1	-1
3	0	-1	0	0	1	-1	-1	1	-1	-1	-1	0	-1	1
3	1	0	0	1	1	-1	-1	1	-1	-1	-1	-1	0	-1
3	-1	-1	1	-1	1	-1	1	1	-1	1	1	1	1	0
3	-1	1	0	-1	1	-1	1	1	-1	1	1	-1	1	0
3	0	1	0	1	-1	1	-1	1	-1	1	-1	1	-1	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	1	1	0	-1	0	1	1	-1	1	1	1	1	-1	1
4	1	0	1	-1	0	-1	-1	1	-1	-1	-1	-1	1	-1
4	0	-1	1	1	1	0	-1	1	-1	1	-1	-1	1	-1
4	-1	-1	-1	-1	-1	0	1	-1	-1	1	-1	-1	1	1
4	-1	1	0	0	-1	1	0	-1	1	-1	-1	-1	-1	-1
4	-1	-1	1	1	1	-1	0	1	1	-1	1	1	1	1
4	-1	0	1	0	-1	1	0	1	-1	-1	-1	1	1	-1
4	-1	0	0	0	1	-1	-1	0	-1	1	1	-1	-1	1
4	0	1	1	0	-1	-1	-1	-1	0	1	1	-1	-1	-1
4	1	0	0	-1	-1	1	1	1	0	-1	-1	-1	-1	1
4	0	-1	0	-1	-1	1	1	1	1	0	1	-1	-1	1
4	1	1	-1	1	-1	1	-1	-1	-1	0	-1	1	-1	-1
4	-1	1	1	-1	1	1	-1	-1	-1	-1	0	1	-1	-1
4	0	0	0	-1	1	-1	1	1	-1	1	0	-1	-1	-1
4	0	-1	0	0	1	-1	-1	1	-1	1	1	0	-1	-1
4	-1	0	1	-1	-1	-1	-1	-1	1	-1	-1	0	-1	1
4	1	1	-1	0	-1	-1	-1	1	-1	1	-1	-1	0	-1
4	-1	1	-1	1	-1	-1	1	1	-1	1	-1	-1	1	0
4	0	-1	-1	-1	-1	1	1	1	-1	1	-1	1	1	0
4	1	-1	0	1	-1	1	-1	1	-1	1	-1	1	-1	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Bias	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 4 - Summary of the ANOVA for plant height of banana trees, Gurutuba cultivar, measured at 3.5 month age

S.V.	D.F.	S.S.	M.S.	F_c	$P(F > F_c)$
Block	3	0.3430	0.1143	0.5762	0.6329
CC	1	0.0136	0.0135	0.0684	0.7946
AG	1	0.0065	0.0065	0.0327	0.8570
CT	1	0.8274	0.8273	4.1697	0.0455 *
LT	1	0.2248	0.2248	1.1329	0.2914
CO	1	0.0498	0.0498	0.2510	0.6182
CD	1	0.9675	0.9674	4.8758	0.0310 *
HA	1	0.4442	0.4442	2.2387	0.1398
MTU	1	0.3065	0.3065	1.5447	0.2187
ACA	1	0.4006	0.4006	2.0190	0.1605
PDQ	1	0.0915	0.0915	0.4612	0.4996
BKP	1	0.2275	0.2275	1.1465	0.2885
BF2	1	0.0837	0.0837	0.4218	0.5185
LTL	1	0.3708	0.3707	1.8686	0.1767
VTK	1	0.0156	0.0155	0.0785	0.7803
CC:AG	1	0.0590	0.0589	0.2971	0.5877
CC:CT	1	0.0136	0.0136	0.0685	0.7943
CC:LT	1	0.0253	0.0252	0.1273	0.7225
AG:CT	1	0.1885	0.1885	0.9501	0.3336
AG:LT	1	1.6424	1.6423	8.2770	0.0055 **
CT:LT	1	0.1211	0.1210	0.6101	0.4378
Res	60	11.9055	0.1984		

Final remarks

In this paper we introduce a screening design alternative that combines levels of a full factorial for some factors with *Definitive Screening Designs* for the others. It was possible to assembly combined designs with efficiency that are smaller and more efficient than ordinary full factorials, restricting our estimation to lower order models.

A design assembled by hand is not bad, but had average variance that was in the middle of a distribution for good designs that could be found by chance. An interchange algorithm, however, makes possible to find a much better design.

Optimized design made the whole experiment possible to be carried out (due to smaller size and better control). It was tested in practice with some implementation problems, but still could be used to make inference on causal factors.

The use of fractional factorials and optimized screening designs is strongly recommended for initial studies in Agronomy.

Acknowledgments

RIBEIRO, P. C. M.; CAMPOS, M. P.; PIO, L. A. S.; BUENO FILHO, J. S. S. Fatoriais fracionários em um estudo de caso de experimento nutricional com bananeiras. *Rev. Bras. Biom.,* Lavras, v.37, n.3, p.335-349, 2019.

■ *RESUMO: Neste artigo, estudamos a combinação de planejamentos, concatenando níveis de um fatorial completo para alguns fatores com alternativas de triagem para os demais. Isso foi feito para lidar com uma situação prática em experimentos de nutrição de plantas. O problema original era o delineamento de um estudo para 14 fatores potenciais da nutrição da bananeira, e os pesquisadores imaginaram que quatro fatores fatoriais completos eram necessários para testar suas hipóteses, sendo dois da série 3^3 e dois da série 3^4. Como isso demandaria pelo menos 216 unidades experimentais e, enfrentando recursos limitados, buscamos uma estratégia de planejamento diferente. A idéia foi combinar no mesmo experimento quatro instâncias de DSD (Definitive Screening Designs) para 10 fatores de três níveis, cada um em um bloco diferente, com uma fração do fatorial completo da série 3^4. Um ponto central de tratamento, com nível médio para todos os fatores, esteve presente em todos os blocos. Algoritmos de troca foram usados para concatenar os níveis dos fatores. O projeto otimizado resultante foi comparado aos projetos amostrados seguindo o mesmo princípio. O critério de comparação de delineamento foi a variância média esperada das estimativas para fatores. A otimização reduziu $4,02\%$ dos valores médios do critério em uma população de referência de delineamentos amostrados. Foi possível mostrar que a variância para efeitos lineares e quadráticos no fatorial completo foi maior que no plano otimizado. Como exemplo, a análise de um teste de campo real é apresentada. Os autores recomendam o uso de estratégias do tipo fatorial fracionária, incluindo desenhos de DSD em ensaios agronômicos, especialmente na fase de triagem.*

■ *PALAVRAS-CHAVE: Algoritmo de intercâmbio; delineamentos combinandos; delineamentos de triagem.*

References

BOX G. E. P.; COX D. R. An analysis of transformations. *Journal of the Royal Society*, v.26, n.2, p.211-252, 1964.

BOX G. E. P.; HUNTER, J.S.; HUNTER, W.G. *Statistics for experimenters: Design, innovation, and discovery.* 2.ed. 2005. 633p.

CHENG, S. W.; WU, C. F. J. Factor screening and response surface exploration (with discussion). *Statistica Sinica*, v.31, n.11, p.553-604, 2001.

GOOS, P.; GILMOUR, S.G. Testing for lack of fit in blocked, split-plot, and other multi-stratum designs. *Journal of Quality Technology*, v.49, n.4, p.320-336, 2017.

JONES, B.; NACHTSHEIM, C. J. Efficient designs with minimal aliasing. *Technometrics*, v.53, n.1, p.62-71, 2011.

JONES, B.; NACHTSHEIM, C. J. A class of three-Level desigs for definitive screening in the presence of second-order effects. *Journal of Quality Technology*, v.43, n.1, p.1-15, 2011.

JONES, B.; NACHTSHEIM, C. J. Definitive screening designs with added two-level categorical factors. *Journal of Quality Technology,* v.45, n. 2, 2013.

LIN, C-Y. Construction and selection of the optimized balanced blocked definitive screening design. *Metrika*, v.78, p.373-383

MEAD, R.; GILMOUR, S. G.; MEAD, A. *Statistical principles for the design of experiments: applications to real experiments.* Cambridge: Cambridge University Press, 2012. 584p.

PLACKETT, R. L.; BURMAN, J. P. The design of optimum multifactorial experiments. *Biometrika*, v.33, n.4, p.305-325, 1946.

R CORE TEAM. *R: a language and environment for statistical computing.* Vienna: R Foundation for Statistical Computing, 2018. Disponível em: <http://www.R-project.org>. Acesso: 14 mar. 2018.

TRINCA, L.; GILMOUR, S. G. Split-plot and multi-stratum designs for statistical inference. *Technometrics*, v.59, n.4, p.446-457, 2017.

TSAI, P.W.; GILMOUR, S. G.; MEAD, R. Projective three-level main effects designs robust to model uncertainty. *Biometrika*, v.87, n.2, p.467-475, 2000.

VENABLES, W. N.; RIPLEY, B. D. *Modern applied statistics with S.* New york: Springer, 2002.

WANG, Y.; AI, M.; LI, K. Optimality of pairwise blocked definitive screening designs. *Annals of the Institute of Statistical Mathematics*, v.68, I.3, p.659-671, 2016.

YANG, J.; LIN, D. K. J.; LIU, M-Q. Construction of minimal-point mixed-level screening designs using conference matrices *Journal of Quality Technology* v.46, I.3, p.251-264, 2017.

Appendix

Call a 84x15 table from the file Base.csv

- 1st column: Block fator

- 4 columns for the 3^4 series full factorial

- 10 columns for the DSD

```
> Base    <- as.matrix(read.table("Base.csv"))
> Blocos <- factor(Base[,1])
> Xbase  <- Base[,2:15]
> Xb      <- model.matrix(~ -1+Blocos)
```

Starting designs and criterion evaluation

```
> inicia  <- function(Xbase){
+   indice  <- sample(c(1:20,22:41,43:62,64:83))
+   sorteio <- c(indice[1:20],21,indice[21:40],42,indice[41:60],63,indice[61:80],84)
+   Xt      <- Xbase[sorteio,1:4]
+   X       <- cbind(Xb,Xt,(Xt^2),
+                 (Xt[,1]*Xt[,2]),(Xt[,1]*Xt[,3]),(Xt[,1]*Xt[,4]),
+                 (Xt[,2]*Xt[,3]),(Xt[,2]*Xt[,4]),(Xt[,3]*Xt[,4]))
+   XX   <- t(X)%*%X
+   XXi  <- solve(XX)
+   Crit <- sum(diag(XXi[5:18,5:18]))
+   return(list(Xt=Xt,Crit=Crit))
+ }
> inicial  <- inicia(Xbase)
> Xt        <- inicial$Xt
> Crit      <- inicial$Crit
> Crit.fin <- Crit
```

Sampling good combined designs (using small N)

```
> N <- 100
> critvec <- 1:N
> for(i in 1:N){
+   inicial <- inicia(Xbase)
+   critvec[i] <- inicial$Crit
+   if(inicial$Crit < Crit){
+     Xprov <- inicial$Xt
+     Crit  <- inicial$Crit
+   }
+   if(i%%10000==0){
+     print(round(100*i/N,2))
```

```
+    }
+ }
```

Vector for criterion values in the sampled designs

```
> summary(critvec)
> CritMedio <- mean(critvec); CritMin   <- min(critvec); CritMedio/CritMin
```

Other definitions to the search function

```
> b  <- 4; k  <- 21; l1 <- NULL; l2 <- NULL; trocado  <- 1
> Xt       <- inicial$Xt
> Crit.old <- inicial$Crit
```

Interchange function

```
> trocas <- function(Xt){
+   for(i in 1:(b-1)){
+     for(j in 1:(k-1)){
+       for(ii in (i+1):b){
+         for(jj in 1:(k-1)){
+           l1    <-  (i-1)*21+ j
+           l2    <- (ii-1)*21+jj
+           Xt.n <- Xt
+           Xt.n[c(l1,l2),] <- Xt.n[c(l2,l1),]
+           Xn    <- cbind(Xb,Xt.n,(Xt.n^2),(Xt.n[,1]*Xt.n[,2]),
+                        (Xt.n[,1]*Xt.n[,3]),(Xt.n[,1]*Xt.n[,4]),
+                        (Xt.n[,2]*Xt.n[,3]),(Xt.n[,2]*Xt.n[,4]),(Xt.n[,3]*Xt.n[,4]))
+           XXn  <- t(Xn)%*%Xn
+           if(det(XXn)>0){
+             XXni <- solve(XXn)
+             Crit <- sum(diag(XXni[5:18,5:18]))
+             if(Crit < Crit.old){
+             Crit.old <- Crit
+             Xt        <- Xt.n
+             trocado  <- trocado+1
+             }
+           }
+         }
+       }
+     }
+   }
+   return(list(Crit=Crit, Xt=Xt, trocado=trocado))
+ }
```

Calling the interchange:

```
> tt <- trocas(Xt)
```

Controling the search process:

```
> Crit      <- tt$Crit
> Crit.old <- Crit
> Xt        <- tt$Xt
> XFin      <- tt$Xt
> trocado  <- tt$trocado
> if(Crit<CritMin){
+   Xprov   <- Xt
+   CritMin <- Crit
+ }
```

Criterion value for best design found

```
> CritMin
```