A prototype of an interactive program for the design of experiments
van Berkum, E.E.M.

Published: 01/01/1990

Document Version
Publisher’s PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:
• A submitted manuscript is the author's version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 11. Dec. 2018
A PROTOTYPE OF AN INTERACTIVE PROGRAM
FOR THE DESIGN OF EXPERIMENTS

E.E.M. van Berkum
June 1990
A PROTOTYPE OF AN INTERACTIVE PROGRAM
FOR THE DESIGN OF EXPERIMENTS

by E.E.M. van Berkum

IWDE REPORT 90-07
June 1990
1. Introduction

Design of experiments is an important tool for a research laboratory. In [1] a study was made concerning the possibility of the development of an interactive computer program, that can be useful in the design of experiments. The following conclusion has been drawn.

Two commercial computer programs can be used in this interactive system. CADE can be used for 2^n-experiments (all factors have 2 levels), and Echip can be used for response surface designs (including the cases where there are constraints or where one is dealing with a mixture). The interactive part of the system should be a shell around these two commercial programs and should assist in choosing the appropriate design. The interactive program must generate a batch file that can be used to start one of the commercial programs.

2. Some useful programs

The results of a literature study of, amongst other, [2] and [3] were important for the decision that Echip and CADE can be useful.

However, one really gets to know a program if one is able to use that program.

2.1. CADE

CADE proved to be a little disappointment.

The basic designs that can be chosen in CADE are the following.

\textit{Table 1. Designs of CADE}

<table>
<thead>
<tr>
<th>Number of factors</th>
<th>Experiment</th>
<th>Number of runs</th>
<th>Defining contrast</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>full</td>
<td>4</td>
<td>--</td>
</tr>
<tr>
<td>3</td>
<td>full</td>
<td>8</td>
<td>--</td>
</tr>
<tr>
<td>4</td>
<td>full</td>
<td>16</td>
<td>--</td>
</tr>
<tr>
<td>5</td>
<td>\frac{1}{2}</td>
<td>8</td>
<td>( I = ABCD )</td>
</tr>
<tr>
<td>6</td>
<td>\frac{1}{4}</td>
<td>16</td>
<td>( I = ABCDE )</td>
</tr>
<tr>
<td>7</td>
<td>\frac{1}{8}</td>
<td>16</td>
<td>( I = BCDF = ADEF )</td>
</tr>
<tr>
<td>8</td>
<td>\frac{1}{16}</td>
<td>16</td>
<td>( I = BCDF = BDEG = ACDG )</td>
</tr>
<tr>
<td>9</td>
<td>\frac{1}{16}</td>
<td>32</td>
<td>( I = ABEF = ABCG = ABDF = ACDF )</td>
</tr>
<tr>
<td>10</td>
<td>\frac{1}{32}</td>
<td>32</td>
<td>( I = ABFG = ADFJ = ACFH = BDGJ )</td>
</tr>
</tbody>
</table>
So there is only one design for each number of factors (apart from \( n = 4 \)) and the confounding structure is fixed. Therefore, after all, it seems not useful to use CADE. However, the basic \( 2^n \)-designs can not be handled by Echip. Therefore we have chosen to use OPZET in stead of CADE.

2.2. OPZET

OPZET is a program developed at URL. This program can generate

i) full experiments if one of the factors has more than two levels. The number of observations is equal to the product of all the numbers of levels;

ii) \( 2^n \)-designs with (or without) blocks.

OPZET can not be used directly to generate fractions of \( 2^n \)-designs. However, if one regards the effects confounded with blocks as defining contrasts, then the first block can be seen as the principal fraction. We have used this in the interactive program.

Therefore it is also possible to have fractions (and fractions in blocks) of \( 2^n \)-designs.

2.3. Echip

Echip seems to be satisfying for many purposes, since it incorporates many possibilities.

The basic idea is that it generates a design with a good \( G \)-efficiency for a response surface.

The \( G \)-criterion minimizes the maximum of the variance of the estimated response. In situations, where there are no qualitative variables and no blocks, Echip gives good designs. This includes the case of a mixture design and the case where one has constraints. Apart from that Echip gives the opportunity to choose standard central composite designs and Plackett-Burman designs.

However, if one has a qualitative variable there are some problems, because the number of levels is fixed. In Echip one cannot choose the number of levels oneself. The number of levels is determined by the program and depends on the model that has been chosen. For that reason it is also difficult to incorporate blocks in a proper way. These problems have been dealt with in the following way.

2.3.1. Qualitative variables

If the factor \( A \) is qualitative with \( k \) levels, then we want that the design constructed has also \( k \) levels for factor \( A \). However Echip acts as if each variable is quantitative and determines the number of levels itself. The number of levels depends on the kind of relation between the response and the factor \( A \).

If, for example, the relation between the response and factor \( A \) is quadratic, then factor \( A \) will have three levels in a \( G \)-optimal design. So if a qualitative factor \( A \) has three levels, we act as if the factor is quantitative and as if the relation between the response and factor \( A \) is quadratic.

If a qualitative variable \( A \) has \( l \) levels, then we act as if the relation between the response and factor \( A \) has degree \( l - 1 \) (or the variation sources of factor \( A \) are the linear term \( A \), quadratic term \( A^2 \), \( A^3 \), \( A^{l-3} \)). In many situations this leads indeed to a design where factor \( A \) has the appropriate number of levels. In some cases, especially where the other variables are mixture
variables, or when the model is rather complicated, the resulting design will have more levels than it can actually have. Future use of the program must give idea if this problem occurs very often.

2.3.2. Blocks

If one has to use Echip (i.e. in the case of mixture design, constraints or complicated models), then blocks cannot be handled in a proper way.

Dummy variables (the block variables) are introduced to deal with this situation. One can choose 2, 3, 4, 6, 8 or 9 blocks.

Example. In the case of 4 blocks, two block variables are introduced $B_1$ and $B_2$, and the terms $x_{B_1}, x_{B_2}, x_{B_1} \cdot x_{B_2}$ are added to the model. The idea is that the optimization with the G-criterion leads to a design where $x_{B_1}$ and $x_{B_2}$ occur equally often. All the observations that have the same combination of values of $x_{B_1}$ and $x_{B_2}$ should be done in the same block.

The complete list of the block variables for each case is the following.

<table>
<thead>
<tr>
<th>Number of blocks</th>
<th>Number of blockvariables</th>
<th>Number(s) of levels</th>
<th>Extra terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>$b_1$</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>3</td>
<td>$b_1, b_1^2$</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>2, 2</td>
<td>$b_1, b_2, b_1b_2$</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>2, 3</td>
<td>$b_1, b_2, b_2^2, b_1b_2, b_1b_2^2$</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>2, 2, 2</td>
<td>$b_1, b_2, b_3, b_1b_2, b_1b_3, b_2b_3, b_1b_2b_3$</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>3, 3</td>
<td>$b_1, b_1^2, b_2, b_2^2, b_1b_2, b_1b_2^2, b_1b_2b_3, b_1b_2^3, b_1b_2b_3$</td>
</tr>
</tbody>
</table>

2.4. Crystal

The actual part of the program, i.e. the shell around Echip and OPZET, has been made by use of Crystal. Crystal is a structured rule-based tool that can be used to build expert systems. URL has experience with Crystal and it has been used to build some expert systems. Crystal is easy to handle and has many possibilities.
3. The structure of the decision tree

In appendix A the decision tree is given. The following notation is used.

i) A rectangle means that a question is posed; from such a rectangle one or more arrows point to the next decision point(s).

Example:

```
confirmation experiment?
```

This means that the question "Is this a confirmation experiment?" is posed. If the answer is Yes, one should choose the arrow with Yes to decision point 1, if the answer is No, one should choose the arrow with No to the decision point 2.

ii) A circle is a decision point where the program makes the decision.

Example: if the number of observations that can be done is known, then the following decision point might occur.

```
number is small?
```

In some parts of the tree an assignment statement is given. It means that if one follows the path, where the assignment statement is given, then the assignment is executed.

We give some comments on parts of the tree.

Page A1

- The program distinguishes three kinds of experiments: confirmation, first and follow-up. A first experiment is the first experiment in this area and might be a screening design. Follow-up is to get more exact knowledge. Confirmation is to confirm earlier results.
- The number of observations is called small if the number is less than NumFac, the number of factors, and is called moderate if only main-effects and all first-order interactions can be estimated.
This part consists of two loops. In the first loop the names of the qualitative factors are chosen and for each factor the number of levels. In the second loop the names of the quantitative factors are chosen. For each factor the program asks whether one wants to choose the number of levels. If the experiment is a confirmation experiment, then the program assumes that one wants to choose the number for each factor. However, a question is posed to ask if this is really the case. If the experiment is a screening experiment, then the program assumes that the number of levels for each factor may be chosen two. However, a question is posed to ask if this is really the case.

If the number of observations is moderate the one can not choose more than 3 levels for these factors.

If the number of observations is large, then a question is posed concerning the purpose of the experiment.

If it is a first experiment, then the purpose is to get knowledge about the relation between response and factors.

In this part we are dealing with a $2^n$-experiment. The program asks which main effects, first-order and higher order interactions must be estimated.

If the variable OK$ has the value "No", then the design must be generated by Echip; therefore an arrow points to VI Select Blocks, this is the introduction to Echip.

If the variable OK$ has the value "Yes", then the design will be generated by OPZET (see V).

Only in the case, that one wants a full experiment, the design will be generated by OPZET; in all other cases the program asks which interactions are of interest.

The question "Do you want a special model" is posed for the case that one wants a model with for example a quadratic term, but no linear term.

In this part the file for OPZET is generated in the case of $2^n$-experiments. First the defining contrasts are chosen and after that the block contrasts.

One can choose whether one wants to give the contrasts oneself, or one wants the standard contrasts given by the program.
If one has a $k$-term model, then at least $k$ observations are needed to estimate the parameters. One can choose extra observations to estimate the variance (this estimation is model-dependent). One can choose replications of points of the design to get as model-independent estimation of the variance.

4. Practical information

The program can be run by using the file opzet.bat. This file contains the command ascr OPZET /d/a100/f. Opzet is the name of the knowledge base. The name of the file containing the knowledge base is opzet.kb.

By giving the command in opzet.bat you enter Crystal. There are several options given at the bottom of the screen in a command line: files, run, clear, build, utilities, quit.

If you enter run the program will start. The program needs the following files to be in the same directory: cr.exe, dsna.exe, mdl.exe, ascr.exe, crystal.ovl, dsn.exe, sho.exe, opzet.bat, opzet.kb and all the files with extension 80.

Now the questions will be posed. After the program has been terminated the command line is shown again. Now you have to enter quit. Three options will be given. You have to choose quit without saving. If you choose quit with saving, then you might change to knowledge base if you used the option build in the command line before.

After you have quit, there are two possibilities.

i) A design will be generated by use of Echip. In that case the file des.bat has been created and you have to enter des and press return to generate the design. You can view the file with sho design and print the design with sho design -w.

ii) A design will be generated by use of OPZET. You must enter opzetp and press return. A file fort30 will be generated, containing the design.

A rule listing can be generated by using the option files in the command line of Crystal.

5. Conclusion

This prototype will construct good designs in many cases. However, in the case of qualitative variables Echip does not always generate proper designs. This prototype must be used in the near future only by experts. They can check whether the proposed designs are satisfactory or not. So one gets an idea whether the program satisfies or not and what should be improved. Things that can be improved are the following.

- The terminology must be adapted to the situation of URL.
• More helpfiles can be made. At the moment there are 15 helpfiles.
• The outputfile can be standardized.
• The program needs about 750 kB. It might be split up into pieces that need less than 640 kB.
• The program must have more checks on validity of answers.
• The way qualitative variables and blocks are dealt with must be improved.
• Fractions and blocks in $3^k$ and $2^n \cdot 3^k$ experiments must be incorporated.
• Other commercial packages may be used.
• Echip uses a linear polynomial model. So transformations of variables and may-be non linear optimization must be included.
• The knowledge base can be easily changed. Too easily. If a user chooses the option "build" in the command line, the knowledge base can be changed. It will be useful to have a command with which the program can be executed but can not be changed.

Literature


II Choice of names and some numbers of levels

What is the number of qualitative factors?

\[ \text{numqualfac} \]

\[ i := 0 \]

\[ i < \text{numqualfac} \]

What is name of qualitative factor \( i \)?

What is the number of levels for this factor?

numquantfac := numfac - numqualfac

\[ j := 0 \]

confirm

Do you want to fix the number of levels for all factors?

if \text{realconf} = \text{Yes}

if Yes, then realconfint = \text{Yes}

Do you want to fix some of the numbers of levels?

if \text{yes}, then nocontinue = \text{Yes}

other cases

\[ j = \text{numquantfac} \]

numrealquant := number of factors of which the number of levels has not been fixed yet

Are there constraints?, see I

realconfint = \text{Yes}

Do you want to fix the number of levels of factor \( j \)?

Yes

What is number of levels of factor \( j \)?

what is lowest and highest possible level of factor?
III Choice of number of levels for factors for which this number has not been fixed yet (the "real quantitative" factors).
IV Choice of interactions.

IV a
- All numbers of levels are 2
- Other cases

IV b
- Numbers of levels are 3

IV c
- Moderate or new exp

VI Select blocks
- OK?
- Do you need blocks?
- Do you want a linear or Plackett-Burman?
- Give the number of replicates
- Construct batch file for standard ECHIP-design

V Make file for OPBET
- Yes

IV a
- Moderate or new exp
- Do you want interactions?
- OK?
- Do you need blocks?
- Randomise?

IV b
- Are you satisfied with a quadratic model?
- Do you need blocks?
- OK?
- Construct batch file for standard ECHIP-design

VI Select blocks
- OK?
- Number of factors
- Do you want higher order interactions?
- Which first-order interactions?
- New exp
- OK?
IVc

not newexp and not moderate

OK?

Do you want a full experiment?

Yes

Do you want a special model?

Yes

VIII Make file 2 for OP2ET

No

Select blocks

VI

Are you satisfied with this model?

No

Choose interactions of two factors in a loop

Yes

Kind of factor i and j

Both qualitative

Both quant.

One qual., one quant.

Give the degree of the qualitative factor

Give highest degree of product

other than (i)

Do you want interactions of more than 2 factors?

No

Select them in a loop

Yes

V

Make file for OP2ET \( 2^n \) experiment

Full?

No

Which fraction do you want? \( \frac{1}{8}, \frac{1}{4}, \text{etc} \)

numdefcon :=

Do you want the standard defining contrasts, given by the program?

No

Choose the defining contrasts

numdefcon > 0

Yes

numdefcon = 0

Do you want blocks?

Yes

Give the names of levels of the qualitative factors

Choose the block contrasts

No

Do you want the standard block contrasts

Yes

Give the number of repetitions

\( \geq 1 \)

\( > 1 \)

The input file is made
VI Select blocks

Do you need blocks?

Yes ➔ Give the number of blocks

No ➔ Give the number of extra observations ➔ Adjust the model with block variables

Give the number of replicates ➔ Randomise?

Yes ➔ Construct batch file for ECHIP

No ➔ Give the constraints ➔ The batch file for ECHIP is ready

Give the constraints ➔ mixture?

Yes ➔ Give the mixture variables

No ➔ mixture?

VII Special model

Do you want all main effects?

Yes ➔ Good model?

No ➔ Select the main effects ➔ VI SELECT BLOCKS

Yes ➔ Good model?

No ➔ Select the first-order interactions ➔ Good model?

Yes ➔ Good model?

No ➔ Select the quadratic terms ➔ Good model?

Yes ➔ Good model?

No ➔ Select higher-order terms.
VIII Make file-2 for OPZET

Do you want replicates?

Do you want standard names for the qualitative factors?

No

Choose them for each factor in a loop

j := 0

Yes

The file is ready

j := numquantfac

Do you want to choose the levels yourself?

Yes

Choose them in a loop (1)

No

The values are given by the program (2)

(1) Default is equidistant
(2) G-optimal values if the variable is the only one.
Appendix B
List of variables, used in the program

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value or meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allfoi$</td>
<td>Yes, if all first-order interactions have to be estimated</td>
</tr>
<tr>
<td>Allfoi2$</td>
<td>Yes, if all first-order interactions have to be estimated</td>
</tr>
<tr>
<td>Allmaef$</td>
<td>Yes, if all main effects have to be estimated</td>
</tr>
<tr>
<td>Allquad$</td>
<td>Yes, if all quadratic terms have to be estimated</td>
</tr>
<tr>
<td>$a[i]$</td>
<td>dummy variable, needed to denote a higher order interaction</td>
</tr>
<tr>
<td>Blockdefc$ [i ]</td>
<td>array of the defining block contrasts</td>
</tr>
<tr>
<td>blocksize</td>
<td>number, related to the blocksize in a $2^n$-experiment</td>
</tr>
<tr>
<td>Blocksnotneed$</td>
<td>Yes, if you do not need blocks</td>
</tr>
<tr>
<td>$b[i]$</td>
<td>dummy variable, needed to denote the coefficients of a constraint</td>
</tr>
<tr>
<td>confinequa</td>
<td>coefficient on the right hand side of the inequality constraint</td>
</tr>
<tr>
<td>confirm$</td>
<td>Yes, if it is a conformation experiment</td>
</tr>
<tr>
<td>Constr$</td>
<td>Yes, if there are constraints</td>
</tr>
<tr>
<td>Constterm$</td>
<td>Yes, if you want to include the constant term in the model</td>
</tr>
<tr>
<td>dofint</td>
<td>degrees of freedom of an interaction</td>
</tr>
<tr>
<td>err$</td>
<td>dummy variable</td>
</tr>
<tr>
<td>exp$</td>
<td>&quot;First&quot; or &quot;Follow-up&quot; depending on kind of experiment</td>
</tr>
<tr>
<td>ext$</td>
<td>dummy variable</td>
</tr>
<tr>
<td>Extra</td>
<td>number of extra observations</td>
</tr>
<tr>
<td>Facinv$</td>
<td>Yes, if you know which factors are involved</td>
</tr>
<tr>
<td>facmodgood$</td>
<td>Yes, if a factorial model is good</td>
</tr>
<tr>
<td>file$</td>
<td>dummy variable</td>
</tr>
<tr>
<td>Firstorik$</td>
<td>Yes, if one wants to include the first order interaction of factors $i$ and $k$</td>
</tr>
<tr>
<td>frac0</td>
<td>$\text{fraci}$ is the number of observations if a $\frac{1}{2^i}$ fraction is chosen</td>
</tr>
<tr>
<td>frac1</td>
<td></td>
</tr>
<tr>
<td>frac2</td>
<td></td>
</tr>
<tr>
<td>frac3</td>
<td></td>
</tr>
<tr>
<td>frac4</td>
<td></td>
</tr>
<tr>
<td>full$</td>
<td>Yes, if you want a full experiment</td>
</tr>
<tr>
<td>Goodmodel$</td>
<td>Yes, if the model proposed is good</td>
</tr>
<tr>
<td>grac0</td>
<td>$\text{fraci} = \text{the number of}$</td>
</tr>
<tr>
<td>grac1</td>
<td>degrees of freedom for estimating</td>
</tr>
<tr>
<td>grac2</td>
<td>variance and blockeffects if a $\frac{1}{2^i}$ fraction is chosen</td>
</tr>
<tr>
<td>grac3</td>
<td></td>
</tr>
</tbody>
</table>
Yes, if one wants higher order interactions
- dummy variable, denoting temporarily the high level
- name of the highest level of factor $i$
- highest possible value of factor $i$

$hraci = i$ if a $\frac{1}{2^i}$ fraction can be chosen and is None if a $\frac{1}{2^i}$ fraction cannot be chosen

- dummy variable

Yes, if the number of observations is large
- Yes, if the inequality in a constraint is "<"

- dummy variable
- Yes, if one wants to fix the number of levels of the factor under consideration
- Yes, if you want to give the levelvalues yourself
- dummy variable, denoting the low level of a factor
- name of the lowest level of factor $i$
- value of lowest possible value of factor $i$
- dummy variable
- Yes, if you want to include the main effect of factor $i$
- the maximum of all the numbers of levels
- the minimum of all the numbers of levels
Yes, if the experiment is a mixture
name of a mixture variable
Yes, if the number of observations is moderate
Yes, if you can do more observations
the name of factor $i$
1 if all the interactions to be estimated do not contain $i$
Yes, if one wants a second new experiment
dummy variable, needed in computation of level values
number of observations
the number of observations needed
Yes, if the number of observations is restricted
$1 + (\text{NumFac} \times (\text{NumFac} + 1))/2$
NumFac + 1
Yes, if the interactions do not have to be estimated
Yes, if it is not really a screening design
number of blocks in a $2^n$-experiment
number of block effects
number of block variables
number of defining contrasts
degree of an interaction
degree of quantitative factor in an interaction
number of factors
NumFac + numblvar
number of fractions
highest possible degree of an interaction
the number of levels of factor $i$
dummy variable
number of quantitative variables of which you have fixed the number of levels
number of qualitative factors
number of quantitative factors
number of really quantitative factors, i.e. factor for which the number of levels is not fixed
number of full replicates
number of parameters (variables) in model
total number of parameters, including those introduced by blocks
Yes, if no mixture and no constraints
Yes, if one wants to include the quadratic effect of a factor
Yes, if one wants to test significance of a quadratic effect of a factor
Yes, if you want an other constraint
otherinterac$ - Yes, if you want an other interaction
Otherterm$ - Yes, if you want an other term in your model
plackbur$ - Yes, if you want a Plackett-Burman design
prodnumlev - product of all numbers of levels
Purpos$ - Relation, Optimal, Target, Significance depending on the purpose of your experiment
Quadeff$i$ - Yes, if you want to include the quadratic effect of factor $i$
quadeffii$ -$
quadmo$ - Yes, if a factor has a quadratic relation with the response
Quadmodel$ - Yes, if you want a quadratic model
quadratic$ - Yes, if you want a quadratic model
qualint$ - Yes, if you want to include the interaction of the two qualitative factors in consideration
random$ - Yes, if one wants to randomize
range - dummy variable, needed in computation of level values
realconfirm$ - Yes, if it really is a confirmation exp. (all levels fixed)
Realconstr$ - Yes, if one want to use constraints
Realmix$ - Yes, if it is really a mixture
Relfacmod$ - Yes, if you want a factorial model to describe the relation
Repli - number of replicates
save$ - dummy variable
small$ - Yes, if the number of observations is small
Somequadeff$ - Some, None or All if one wants some, none or all quadratic effects; More if one wants more than quadratic effects
special$ - Yes, if one wants a special model
standblockdef$ - Yes, if one wants the standard block contrasts
standdef$ - Yes, if one wants the standard defining contrasts
stanlevname$ - Yes, if one wants the standard level names
stdefc$[i] - the standard defining contrasts
sum - dummy variable, needed in computation of level values
t$[i] - dummy variable
totnumdef - numdefcon + numblockeffect
typequad$ - Opt, Csc, Css, type of quadratic model
user_date$ - dummy variable
zeros$ - dummy variable