MAXIMIZING THE SIMULATION OUTPUT: A COMPETITION

Jack P.C. Kleijnen & Őzge Pala[1]

Abstract

The Business Section of the VVS (Netherlands Society for Statistics and Operations Research) organized the following competition. Maximize the output of a given simulation model by selecting the best combination of six inputs; only 32 runs are permitted. Participants in this competition came from industry and academia; actually twelve teams competed. This paper is written by the winning team, explaining its design and analysis. That design proceeds in stages. First, Rechtschaffner's saturated design for estimating all main effects and two-factor interactions is used. Then factors are changed one at a time to estimate quadratic effects. Finally, the estimated second-order polynomial is used to estimate the optimal input combination.

1. Introduction: the competition explained

Two recent issues of the *VVS Bulletin* (November 1997, pages 150-151 and December 1997, pages 162-163) defined the following problem (the translation from the original Dutch text into English is our's).

'Optimize your own output! You have developed an advanced computer model that computes the output of the synthesis of zeolite on gauze pads, for given values of the following six factors:

Factor	Current Setting
(A) Aluminum	150 mM
(B) Silicon	400 mM
(C) Temperature	250 °C
(D) Template	10 mM
(E) Rotations	300 rpm
(F) Copper	100 μM

For the current setting the computer model calculates an output of 90.9 ppb. You have the impression that this setting is not optimal at all. Therefore you decide to start experimenting with

[1] Department of Information Systems and Auditing (BIKA)/Center for Economic Research (CentER), School of Management and Economics (FEW), Tilburg University (KUB) Postbox 90153, 5000 LE Tilburg, The Netherlands.

Kleijnen: phone: +3113-4662029; fax: +3113-4663377; e-mail: kleijnen@kub.nl; web: center.kub.nl/staff/kleijnen

Pala: phone: +3113-4662262; fax: +3113-4663377; e-mail: O.Pala@kub.nl

the settings of these six factors. ... you can compute no more than 32 runs. By how many ppb can you increase the output?
Rules of the game:

1. [Given is the following table:]

run	A	B	C	D	E	F
1	120	380	200	10	300	100
2	180	380	200	10	300	100
3	120	420	200	10	300	100
4	180	420	200	10	300	100
5	120	380	300	10	300	100
6	180	380	300	10	300	100
7	120	420	300	10	300	100
8	180	420	300	10	300	100

2. We [the organizers of the competition] will e-mail you a similar list, including the corresponding output.
Note:
Of course, the table above is only an example, in which only the factors A, B, and C were varied. You are permitted to vary more factors or fewer factors as long as you indicate for each of the six factors how you wish to set its value. In the example 8 runs were offered. So 24 runs remain for new experiments.

You yourself determine how you will spread the 32 runs over the experiments, e.g. 1 experiment with 32 runs, 2 experiments with 16 runs. 1 experiment of 16 runs and 2 of 8 runs, etc. ...You can register no later than 5 January 1998 ...'

At the start of our search, this was all we knew about the problem. In other words, we had no information on the process itself, the ranges of its inputs (or factors), etc.; we did know one input combination and its resulting output. We shall call this latter run the free base run.

We organize this report on our search, as follows.
§2. Solution strategy selected
§3. Rechtschaffner (1967) 's saturated R-5 design
§4. Quadratic effects: one-at-a-time design
§5. Estimating the optimal combination from the second-order polynomial
§6. Conclusions
§7. Epilogue
Appendix: All 33 runs and their inputs and outputs

2. Solution strategy selected

Any simulation model implies an input/output (abbreviated to I/O) function or response surface. Since the simulation model of this competition represents a chemical system, we assume that interactions among the six factors (or inputs) are important. Moreover, the competition concerns a maximization problem, so we assume that quadratic effects are important (as these effects can model a 'hill top'). Therefore we assume that the I/O function can be adequately approximated by a *second-degree polynomial* over the area of experimentation. This polynomial has twenty-eight parameters, namely the overall mean or intercept (say) β_0, the six main or first-order effects β_1, ..., β_6, the fifteen two-factor interactions $\beta_{1;2}$, ..., $\beta_{5;6}$, and the six quadratic effects $\beta_{1;1}$, ..., $\beta_{6;6}$. Which experimental design should we select to estimate these parameters?

We have a tight 'computer budget', since §1 stated that we can make only 32 runs; we have one run free, namely the base run. So to estimate all effects, we need 27 more runs. But we do not wish to spend 84% of our computer budget in one shot. Instead we decide to proceed stagewise: experiments with computer models are usually executed one-by-one (whereas experiments in, for example, agriculture need to be executed in one shot, as the growing season allows no sequentialization). We further decide to focus on interactions, before quadratic effects.

Note: On hindsight, the interactions are not so important as we assumed; see "Epilogue' (§7). So a resolution-4 - abbreviated to R-4 - design would have been better. Such a design, however, is not a subset of the design that we shall actually use, namely a Rechtschaffner design; see next section. An R-4 design may be a subset of a Resolution-5 (or R-5) design; to the class of the R-5 designs belong 2^{k-p} designs with $k = 6$ and $p = 1$ so 32 runs are needed; see Kleijnen (1987, p. 309). However, we cannot afford so many runs since there is a limit of 32 runs and we also want to estimate the quadratic effects. This limit also implies that we cannot apply Response Surface Methodology (or RSM), which combines a series of local designs with steepest ascent. See Kleijnen (1998) for details including nearly one hundred references; we limit the references to those publications that we really used.

Once we have also estimated the quadratic effects, we take the six partial derivatives $\partial y/\partial z_j$ with $j = 1$, ..., 6, equate to zero, and estimate the optimum factor combination.

Note: The appendix gives the inputs and outputs of all our 32 runs, plus the free base run.

3. Rechtschaffner (1967) 's saturated R-5 design

Our strategy implies that we first estimate the overall mean β_0, the six main effects β_1, ..., β_6, and the fifteen two-factor interactions $\beta_{1;2}$, ..., $\beta_{5;6}$ (in total, 22 effects). Because of the tight computer budget, we prefer a *saturated design*, that is, a design with a number of runs (say) n equal to the number of effects, q (in our case $n = q = 22$). There are many such designs, satisfying different criteria. By definition, R-5 designs give unbiased estimators of the overall mean, all main effects, and all two-factor interactions. We select a design that is readily available, namely a *Rechtschaffner* design. This design was derived in Rechtschaffner (1967) and replicated in Kleijnen (1987, pp. 310-311); see Table 1 (unlike 2^{k-p} designs, this design is nonorthogonal so the design does not give minimum variance estimators).

Table 1 gives the *standardized* values of the factors; that is, - stands for -1, and + for 1; - (respectively +) means that the factor has its lowest (respectively highest) value in the experiment.

Rather arbitrarily we decide to let + correspond with a 10% increase of the factor relative to the base value given in the problem description (§1); for example, factor A has a base value of 150, so + means that A has value $z_1 = 165$. Analogously, z_2 denotes the value of factor B, etc. In the analysis of Rechtschaffner's design we use the standardized values (say) x rather than the original values (or measurement scales) z, because the effects can then be compared with each other without thinking about their different units (A is in mM, C in C°, etc.). The effects of the standardized factors can be used to detect the most important factors. In the next stage we shall use the original scales. Also see Kleijnen (1998).

3.1 Main effects only: first eight runs

Above (§2) we stated that we prefer to experiment *stagewise*. Actually we hope that one or more factors are unimportant. Therefore we first try to estimate main effects. At least seven runs are needed to estimate six main effects and one overall mean. Run #1 in Table 1 is the free run. Misled by the fact that a 2^{k-p} design of resolution-3 requires eight runs, we decide to execute runs # 2 through # 8 (instead of #2 through #7) in Table 1. When we use eight (or seven) runs, the estimators of the main effects may be biased by higher-order effects (such as two-factor interactions and quadratic effects). Hence it is dangerous to declare a variable unimportant when its estimated main effect is not significant. But how do we analyze the experimental results?

Table 1: Rechtschaffner (1967) 's saturated R-5 design in standardized values (- is -1; + is 1)

Run	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
1	-	-	-	-	-	-
2	-	+	+	+	+	+
3	+	-	+	+	+	+
4	+	+	-	+	+	+
5	+	+	+	-	+	+
6	+	+	+	+	-	+
7	+	+	+	+	+	-
8	+	+	-	-	-	-
9	+	-	+	-	-	-
10	+	-	-	+	-	-
11	+	-	-	-	+	-
12	+	-	-	-	-	+
13	-	+	+	-	-	-
14	-	+	-	+	-	-
15	-	+	-	-	+	-
16	-	+	-	-	-	+
17	-	-	+	+	-	-
18	-	-	+	-	+	-
19	-	-	+	-	-	+
20	-	-	-	+	+	-
21	-	-	-	+	-	+
22	-	-	-	-	+	+

We use *ordinary least squares* (OLS) to estimate the effects β. We emphasize that OLS is a mathematical - not a statistical - fitting criterion. Denote these OLS estimates by $\hat{\beta}$.

Note: We do not know whether the simulation outputs have a particular distribution. So we might assume that they are deterministic, and that the fitting errors are normally identically and independently distributed (NIID), so we might apply Student's statistic to test the significance of the estimated effects. Actually, we do not test and eliminate factors or effects. We do the OLS analysis in SPSS, which routinely gives '95% confidence intervals'.

Before we submit runs #2 through #8, we check whether these runs together with the free run (run #1) permit OLS estimation: we compute the inverse of $(X'X)^{-1}$ where X denotes the eight-by-seven matrix of standardized inputs including the dummy input that corresponds with the overall mean (we use MatLab to compute this inverse).

The fitted polynomial gives a coefficient of determination, denoted as R-square, of 0.99999, and a coefficient adjusted for the number of parameters of 0.99996. Table 2 gives the OLS estimates $\hat{\beta}$. SPSS (assuming NIID fitting errors; see Note above) automatically computes the standard error of the fitting errors of the model (namely 0.01516) and the standard errors of the estimated main effects and their 95% confidence intervals. This table shows that under this NIID assumption all main effects are significant. Actually all effects have roughly the same standard error, namely 0.007, so the magnitudes of the OLS point estimates can be used to sort the factors. Then factor B is the most important factor; factor D the least important; factor F the only 'negative' factor. However, these are only tentative conclusions, because - as we said before - the main effects may be biased by higher-order effects and the statistical significance test assumes NIID. Our conclusion after the first stage is that there is not enough information also enclose either to eliminate a factor or to make any changes in the factor levels.

Table 2: Estimated main effects $\hat{\beta}$ with their standard errors and 95% confidence intervals

Factor	$\hat{\beta}$	Standard error	Low	Upper
A	0.785726	0.006628	0.701512	0.869941
B	1.793276	0.006628	1.709062	1.877491
C	0.541622	0.006816	0.455012	0.628232
D	0.419272	0.006816	0.332662	0.505882
E	0.612872	0.006816	0.526262	0.699482
F	-0.668528	0.006816	-0.755138	-0.581918
(Constant)	94.378726	0.006628	94.294512	94.462941

3.2 Two-factor interactions: remaining runs

Next we execute the remaining runs of Rechtschaffner's design: runs #9 through #22 in Table 1. The outputs vary between a minimum of 90.369 and a maximum of 99.204 (base output was 90.900); also see the appendix. Since the design is saturated, R-square is 1.0, and the coefficient adjusted for the number of parameters is also 1.0; the resulting standard error of the fitting errors is 0.0. Going from eight runs to twenty-two runs changes the factor estimates:

(i) the overall mean changes to 94.3616
(ii) the six main effects change to 0.79105, 1.79775, 0.5415, 0.41918, 0.61275, and -0.66778
(iii) all two-factor interaction can now be estimated; they equal 0.00225, except for the A-F interaction, which is 0.0014, and the D-F interaction, which is 0.002225. These interaction estimates suggest that all two-factor interactions are unimportant. We shall return to this issue.

4. Quadratic effects: one-at-a-time design

Our next step is the estimation of the six quadratic effects $\beta_{j;j}$. Therefore we change one factor at a time. Each factor should get a value that differs from its previous value so each factor has at least three values: expressing values in standardized units, we change factor j to (say) c_j with $c_j \neq -1$ and $c_j \neq 1$. Moreover we select a sequential design: we execute runs, one by one (changing the level of only one factor). In this way, we can re-estimate main effects, interactions, and quadratic effects of that one input. If the estimated optimum value of that input lies within the limits used so far, it means that we are experimenting within the optimal range of that factor. However, if that estimated optimal value is far outside the current range, we seem to be searching in the wrong area!

The first factor we change is the seemingly most important factor, B; see the preceding subsection §3.2. Furthermore, we decide to change this factor in the combination that yielded the highest output so far (namely run #7, which had output 99.204; see Appendix). Since the estimated main effect of B is positive, we increase B's value; we do so by another 10%, which gives the value 484 in the z domain (which implies the value 3.2 in the x domain; we point out that the x's and z's are related through linear transformations with non-zero constants, so a 10% change in z is not a 10% change in x). The output for this combination is 102.79, which is a further increase of 3.59%.

After adding this run to the previous 22 runs, we re-estimate the regression model. Next we take the partial derivatives $\partial y/\partial z_j$, equate them to zero, and solve for the estimated optimum factor values. Of course, the values for the other five factors do not make sense because their quadratic effects are not yet estimated. The 'optimal' B value turns out to be far away from the values we have been working with so far: the B value becomes 15.4843 in the x domain or 729.68599 in the z domain. Of course, this does not necessarily apply to the other factors. Yet, we decide to take larger steps for the other factors, in the next runs: we increase all other inputs by 20% in the original scales; except for factor F, which we decrease by the same percentage (hence the new standardized values become 6.5 for A, C, D, and E; -6 for F).

Next we execute five more runs, namely runs #24 through #28. Run #24 gives a higher output, namely 103.1224. We re-estimate the second-order polynomial. The overall mean and main effects are close to those in §3.2; the interactions remain unchanged; the six quadratic effects are -0.011488, -0.041071, -0.016225, -0.004175, -0.023099, and -0.019517.

5. Estimating the optimal combination from the second-order polynomial

Next we again compute the six partial derivatives, equate them to zero, and solve for the estimated optimum factor values. This gives the z values 530.94375, 955.02000, 623.20625, 51.96925, 647.07900, and 74.43700.

This z combination is the next run, run #29. It gives an output of 145.4481, which is a drastic increase, namely of 41.04%, compared with the highest output so far. Re-estimating the polynomial gives an overall mean, main and quadratic effects that hardly change, and interactions that change quite a bit. The re-estimated optimal z values are 409.5375, 886.8600, 516.0275, 36.1193, 529.1760, and 8.6685. This combination is the input for the next run, run #30. This yields 159.5943, a further increase.

The next re-estimated optimal inputs are 427.7513, 925.72, 533.6238, 38.71125, 551.823, and -7.0495 So the value for F is *negative*, which is impossible since F denotes the factor copper. Thus, we decide to keep its level at zero in the next run. This yields an output of 157.5518, which is a decrease compared with the immediately preceding run. The re-estimated effects hardly change. The re-estimated optimal inputs are now 396.08925, 1109.382, 632.425, 39.63895, 512.8965, and 58.2545. So some inputs increase, some decrease, factor F becomes positive again, which is more meaningful. This combination becomes the input for run #32. This yields an output of 151.3, a decrease.

In the next stage we investigate whether it makes sense to eliminate the run that had zero input for factor F, from the analysis (treat that run as an 'outlier'). This, however, again gives a negative value for F. Moreover, the output predicted by the regression model is lower than the output predicted when keeping that run (with negative F-value) in the analysis. The latter approach gives roughly the same effects as the preceding runs did. The optimal z values become 531.652, 1034.68, 488.22, 32.9985, 642.859, and 39.3584. This z combination is the input for our last run, run #33. This yields an output of 152.6. This output is not the maximum over all 33 runs; the maximum output is that of run #30 (also see the appendix).

6. Conclusions

We have a total of 33 runs, including the free base run provided in the problem definition. The first 22 runs were used for the estimation of the 6 main effects and the 15 two-factor interactions, besides the overall mean. These runs were specified by Rechschaffner's saturated design (Table 1), and by deciding to change the factors by 10% (Appendix). These runs gave outputs that increased by no more than 9% (90.9 became 99.2).

The next six runs were meant to estimate the six quadratic effects. We changed the factors one at a time, increasing them by 20% (appendix, runs #23 through #28). This increased the output to a maximum of 103.1, a modest increase.

The remaining five runs (#29 through #33) used the five combinations that were estimated to be optimal, using the second-order polynomial that was re-estimated after each run. These last five runs gave substantially improved outputs compared with the preceding set of runs.

The maximum output is the result of run #30; this maximum is 159.5943. This is a 76% increase compared with the base output, 90.9. Obviously, our estimated maximum is not necessarily the global maximum: we might have gotten stuck at a local maximum Actually, the true maximum output turns out to be 160 (see Epilogue), so we have succeeded in approximating the true maximum very closely.

7. Epilogue

After we finished our search for the maximum simulation output, the true maximum was revealed by the organizers. The simulation model that was a black box to us, turned out to be the following model:

$$y = 160 +$$
$$- (z_1 - 420)^{2/5000} - (z_2 - 870)^{2/10000} - (z_3 - 480)^{2/10000} - (z_4 - 40)^{2/70} - (z_5 - 520)^{2/10000} - (z_6 - 40)^{2/1000} +$$
$$+ 30/\{[(z_1 - 420)(z_6 - 40)/1000]^2 + 5\} - 30/5.$$

So the true maximum output is 160. There are neither main effects nor interactions except for the interaction between A and F. There is no noise. Notice that the last term (30/5) is subtracted, because the interaction term for optimal input values is $30/(0^2 + 5)$.

We have the following comments on this competition. We were disappointed to learn that the simulation model was only a mathematical function, not a real-life problem that we were helping to solve. The fact that the simulation model was only this function explains why the participants did not get any information on the process itself and the ranges of its inputs. Hence, in our view the competition is unrealistic: in real life the analysts accumulate much knowledge while developing their (simulation) model. This knowledge concerns both the model and the real system that is modeled. In real life the analysts and problem 'owners' should cooperate!

Notwithstanding this criticism, not only *we* found this an interesting and challenging problem: twelve teams competed, employed by operations research and statistics departments of well-known international companies (Philips, Unilever), research institutes (TNO, DLO), and universities (Amsterdam, Tilburg). We won the competition, but it was a 'photo finish': our maximum output was 159.6, whereas the second-place output was 159.4.

Note: The problem to be solved in this competition has no known optimal solution strategy (the organizers - not the competitors - did know the optimal solution; see above). Consequently, each competitor had to resort to a *heuristic* solution strategy. By definition, there are infinitely many heuristic strategies. Certainly it would be interesting to compare some of the better heuristics, but this requires a joint paper by the better teams. At the meeting at which the competitors presented their solutions, it turned out that typically our strategy gave relatively low (compared with our competitors) simulated outputs during the first part of our experiments; in the last part our strategy accelerated and overtook the competitors' outputs. The current paper demonstrates a strategy that seems a good heuristic for application in real life. Obviously no heuristic is always 'best' (it would not be a heuristic). Determining when a particular heuristic is applicable, is rather difficult. One practical solution might be: apply the heuristic that is most familiar ('a carpenter can solve any problem with a hammer').

Moreover, in some other respects this competition was realistic. The number of runs was limited to 32, and there was a deadline (5 January 1998).

References

Kleijnen, J.P.C. (1998), Experimental design for sensitivity analysis, optimization, and validation of simulation models, *Handbook of simulation*, edited by Jerry Banks, Wiley, New York

Kleijnen, J.P.C. (1987), *Statistical tools for simulation practitioners*, Marcel Dekker, New York

Rechtschaffner, R.L. (1967), Saturated fractions of 2n and 3n factorial designs, *Technometrics*, 9, pp. 569-575

Acknowledgment

We thank the organizers of this competition, especially P. Marres (TNO) and J.C. Steigstra (CQM), and also W. Van Groenendaal (KUB) and an anonymous referee for comments on the first draft of our report.

Appendix: All 33 runs and their inputs and outputs

	Input						Output
Run	A	B	C	D	E	F	Output
1	150	400	250	10	300	100	90.9000
2	150	440	275	11	330	110	96.2860
3	165	400	275	11	330	110	94.2709
4	165	440	250	11	330	110	96.7834
5	165	440	275	10	330	110	97.0281
6	165	440	275	11	300	110	96.6409
7	165	440	275	11	330	100	99.2037
8	165	440	250	10	300	100	96.0433
9	165	400	275	10	300	100	93.5308
10	165	400	250	11	300	100	93.2862
11	165	400	250	10	330	100	93.6733
12	165	400	250	10	300	110	91.1106
13	150	440	275	10	300	100	95.5425
14	150	440	250	11	300	100	95.2979
15	150	440	250	10	330	100	95.6850
16	150	440	250	10	300	110	93.1257
17	150	400	275	11	300	100	92.7854
18	150	400	275	10	330	100	93.1725
19	150	400	275	10	300	110	90.6132
20	150	400	250	11	330	100	92.9279
21	150	400	250	11	300	110	90.3685
22	150	400	250	10	330	110	90.7557
23	165	485	275	11	330	100	102.7941
24	206.25	440	275	11	330	100	103.1224
25	165	440	343.75	11	330	100	101.5498
26	165	440	275	13.75	330	100	101.3742
27	165	440	275	11	412.5	100	101.6581
28	165	440	275	11	330	75	101.8076
29	530.9438	955.02	623.2063	51.96925	647.079	74.437	145.4481
30	409.5375	886.86	516.0275	36.1193	529.176	38.6685	159.5943
31	427.7513	925.72	533.6238	38.71125	551.823	0	157.5518
32	396.09	1109.382	632.425	39.63895	512.8965	58.2545	151.3000
33	531.652	1034.68	488.22	32.9985	642.859	39.3584	152.6000

Ontvangen: 04-02-1998
Geaccepteerd: 18-06-1998