Comparison of On-line Design of Experiments Methods on Physical Models

Rutten, Koen
KULeuven, Department of Biosystems
Kasteelpark Arenberg 30, bus 2456
3001 Heverlee, Belgium
E-mail: koen.rutten@biw.kuleuven.be

De Ketelaere, Bart
KULeuven, Department of Biosystems
Kasteelpark Arenberg 30, bus 2456
3001 Heverlee, Belgium
E-mail: bart.deketelaere@biw.kuleuven.be

De Baerdemaeker, Josse
KULeuven, Department of Biosystems
Kasteelpark Arenberg 30, bus 2456
3001 Heverlee, Belgium
E-mail: josse.debaerdemaeker@biw.kuleuven.be

The general goal of the research was to develop Design of Experiments (DOE) methods that can be used during the production process. They differ from classical DOE because new experiments are incrementally defined based on the outcome of previous experiments. A first step consisted of comparing a heuristic online optimization strategy, the simplex, with the most basic form of sequential response surface modeling (sRSM), a factorial augmented with steepest ascent, called factorial EVOP in this text. The comparison was made using the mathematical model of a real physical process. The optimization methods were programmed in Matlab®. To compare these methods, a set of hundred randomly chosen starting points was obtained as an input for all methods. Several comparisons were made: number of steps needed to reach the optimum, quality of the optimum… The first results are presented and discussed.

Statistical Design of Experiments has been used for decades to optimize a variety of industrial processes. One of its greatest failings has always been that the production process usually has to be stopped for the experimentation to be carried out. Such one-shot optimization presents a problem in processes that drift in time. Not only process drift is of concern, modern industrial processes are often highly complex and are difficult to model. The higher the complexity of the model, the more experimental runs are necessary, resulting in high costs. Therefore, it is interesting to look at methods that can continuously optimize a process without stopping the production line. In this paper an adaptation of Evolutionary Operation (EVOP) by George Box (Box, 1957), and the simplex algorithm, as described by Nelder and Mead (Nelder & Mead, 1965), are compared. More information about the simplex method can be found in (Walters *et al.*, 1991).

The Unilever research model, as used in the VIRTEX factory applet (Darius *et al.*, 2003), was used. In this simulation it was used as a three-factor model. The three factors are: temperature, concentration and time. The response of the system is the yield, expressed in monetary profit [EUR]. Limits for these factors can be found in table 1.

Table 1 - Factor settings

Factor	Min	Max
Temperature	60	140
Concentration	0	0.08
Time	80	160

For a two-factor experiment the situation in which temperature could not change due to production setup limitations was considered. The temperature was held constant at 140 °C.

Table 2 - Settings for simulations

Simulation #	**EVOP** *Factor point measurements*	**EVOP** *Number of centerpoints*	**Simplex** *Point repeats*
1.1	1	1	1
1.2	2	2	2
1.3	3	3	3
2.1	1	1	1
2.2	1	2	2
2.3	1	3	3

For two factors, 25 equidistant points were used as starting coordinates for the optimization methods, and each optimization is repeated 30 times. For three factors, 125 equidistant points were used as starting coordinates, repetition of an optimization stayed at 30 times.

Since this initial research resembles an offline optimization and no intelligent choices have been imposed to stop optimizing a maximum number of optimization runs is defined. For the 2-factor optimization the maximum number of runs is defined as 150 EVOP runs. For Simplex the maximum number of points is calculated from the EVOP runs as shown in equation [1], with k the number of factors and c the number of centerpoints.

$$Runs_{simplex} = 150 \cdot \left(2^k + c\right) \qquad [1]$$

Increasing variation through SNR

To check robustness by increasing variation the signal-to-noise ratio was defined as the square root of the variation in response when no noise is present, divided by the standard deviation of the currently tested mathematical model, as in equation [2].

$$SNR = \frac{\sqrt{Var\left(Y_{\sigma=0}\right)}}{\sigma} \qquad [2]$$

The SNR, as a relative indicator for variation, allows us to test all models within the same parameters. All models were tested with these settings for the SNR: ∞ (no noise), 100, 10, 5, 1 (as much signal as noise).

Results

Keep in mind that we know the theoretical range in the experimental domain. In a normal situation we only know the yield at the current operating conditions and can only see if the yield improves. In all following graphs the results are scaled to the theoretical range within the experimental domain. Thus: the yield in the experimental domain ranges from 0 % to 100%.

The two-factor model has two additional difficulties because the theoretical optimum lies outside the experimental domain and the model has two slopes inside the experimental that the optimization methods can see as planes toward an optimal response, as can be seen in figure 1.

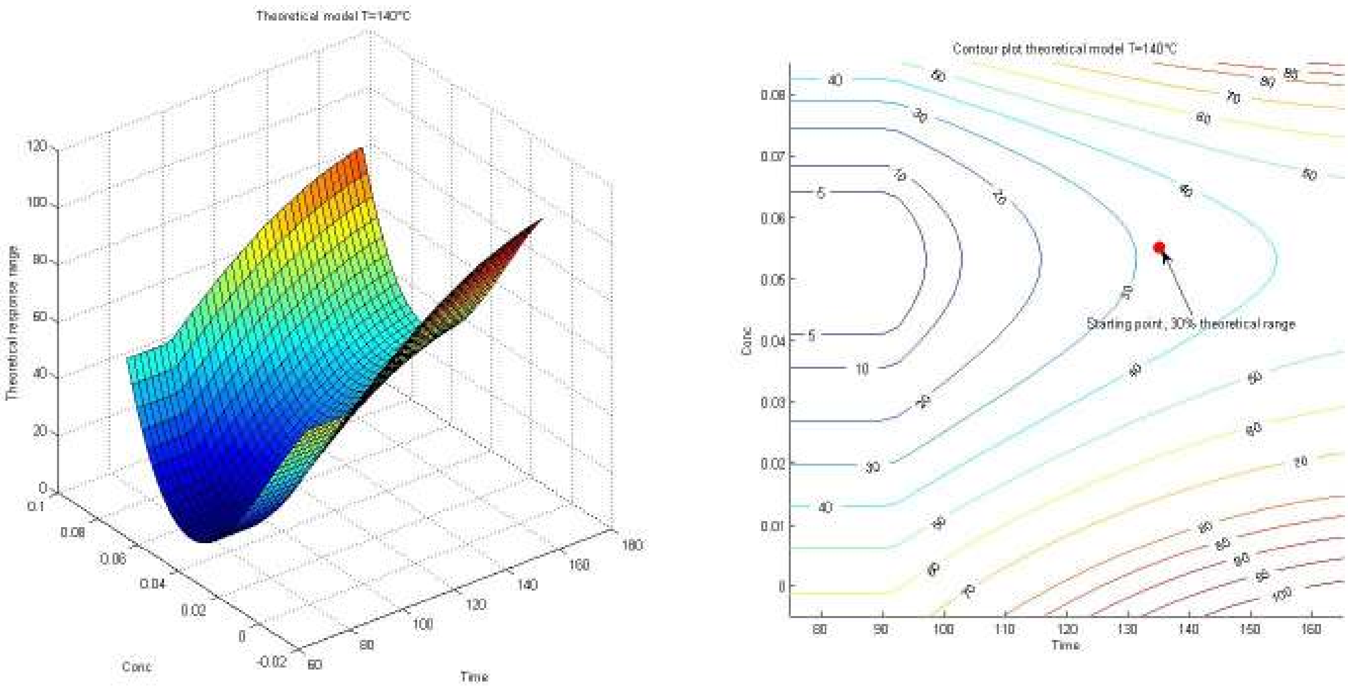

Figure 1 - Two-factor model

In figure 2 the boxplots for EVOP (blue) and Simplex (red) are given per inverse SNR. Optimizations are done with noise but the response on the optimal point is calculated on the model without noise, to be able to compare all optimizations.

It can be seen that the spread increases with the addition of more noise. The simplex spread for 1/SNR=1 seems to be an exception to this. This small spread is due to the stop criteria of the simplex method. If a certain point is retained for more then 12 * *k* times, where *k* is the number of factors, the optimization stops. Due to the high noise levels the simplex stops quite rapidly. It is also interesting to note that every optimization increases yield or keeps it around 30%.

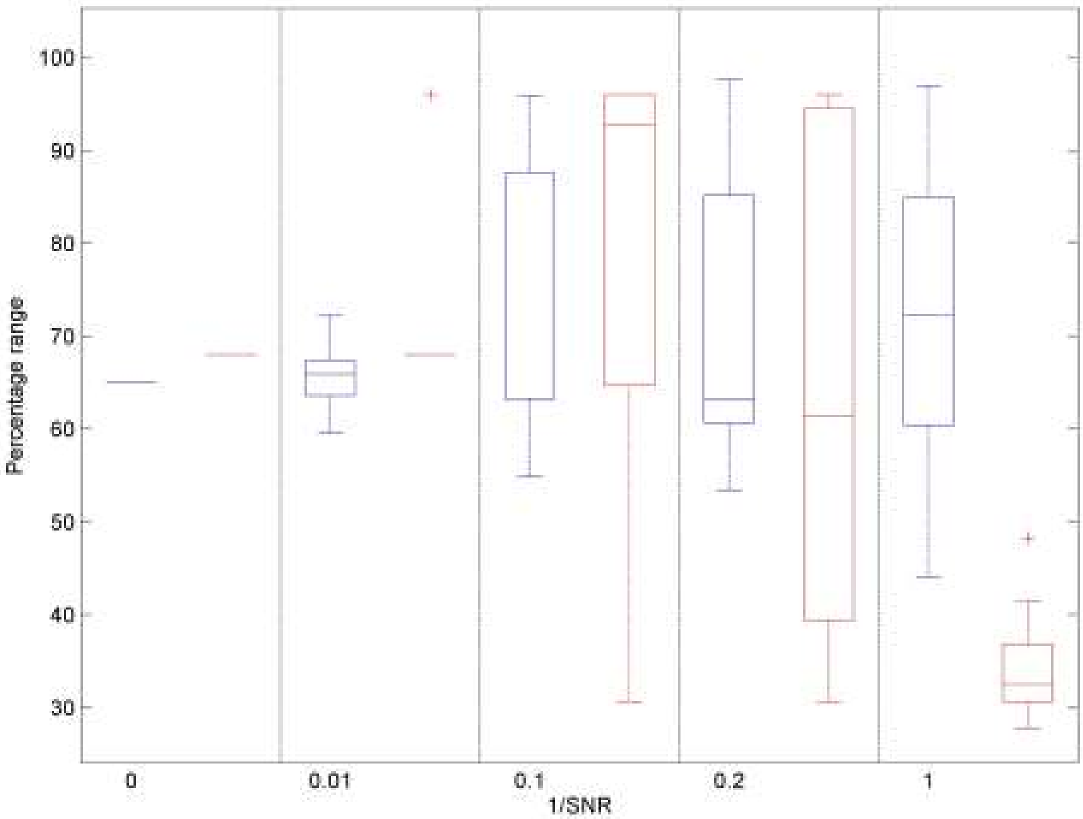

Figure 2 – Two-factor ptimizations over all noiselevels, starting point closest to 30%, simulation 1.1, EVOP (blue)/Simplex (red)

In low-noise situations simplex often performs better than simplex. The higher the noise-level, the beter EVOP does compared to EVOP. Additional experiments at factor points and centerpoints improve the overall median of EVOP but often increase the spread.

In the case of three factors, the situation becomes less clear. 30% of the theoretical range is a special point that seems to correspond to a local maximum thus for three factors thus two other starting points were chosen. In figure 3 all optimizations for the starting point closest to 15% theoretical range are shown, figure 4 shows the same for the starting point closest to 40% theoretical range.

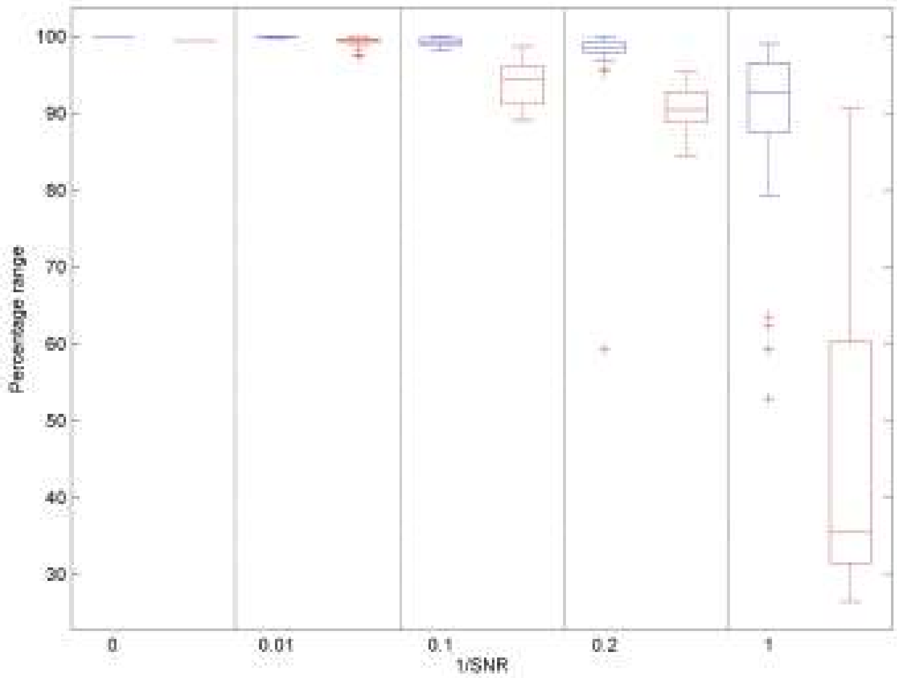

Figure 3 - Three-factor ptimizations over all noiselevels, starting point closest to 15%, simulation 2.1, EVOP (blue)/Simplex (red)

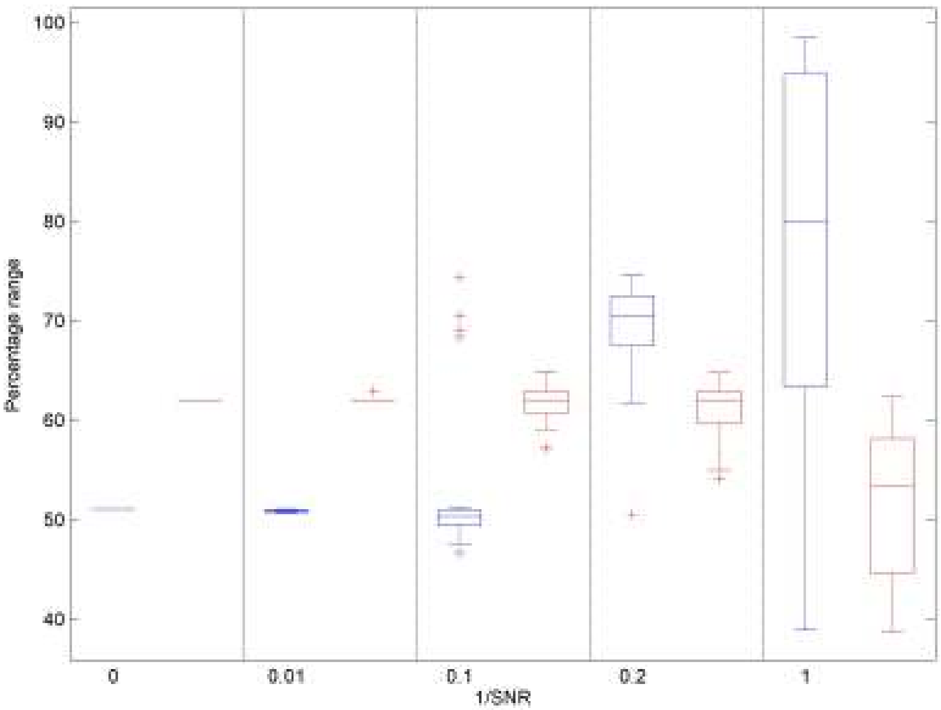

Figure 4 - Three-factor ptimizations over all noiselevels, starting point closest to 40%, simulation 2.1, EVOP (blue)/Simplex (red)

This shows clearly that the coordinates of the starting point are extremely important for these optimizations. An intelligent way to determine initial stepsizes could overcome some of the problems that we see in the two figures above.

REFERENCES

Box, G. E. P. (1957). Evolutionary Operation: A Method for Increasing Industrial Productivity. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *6*, pp. 81-101.

Darius, P., Schrevens, E., van der Knaap, H., Portier, K., Massonnet, G., Lievens, L., Duchateau, L. & Thas, O. (2003). Using web-based tools for teaching statistical concepts and experimentation skills. *Proceedings of Statistics and the Internet Conference*, , .

Nelder, J. A. & Mead, R. (1965). A Simplex Method for Function Minimization. *The Computer Journal*, *7*, 308-313.

Walters, F. H., Lloyd, R. J., Morgan, S. L. & Deming, S.N. (1991). Sequential Simplex Optimization: A Technique for Improving Quality and Productivity in Research, Development, and Manufacturing (Chemometrics series). CRC.