AUTOMATED RESPONSE SURFACE METHODOLOGY FOR STOCHASTIC OPTIMIZATION MODELS WITH UNKNOWN VARIANCE

Robin P. Nicolai
Rommert Dekker
Nanda Piersma

Department of Econometrics and Operations Research
Erasmus University Rotterdam
POB 1738, 3000DR Rotterdam, THE NETHERLANDS

Gerrit J. van Oortmarssen

Department of Public Health
Erasmus University Rotterdam
POB 1738, 3000DR Rotterdam, THE NETHERLANDS

ABSTRACT

Response Surface Methodology (RSM) is an optimization tool that was introduced in the early 50´s by Box and Wilson (1951). In this paper we are interested in finding the best settings for an automated RSM procedure when there is very little information about the objective function. We will present a framework of the RSM procedures that is founded in recognizing local optima in the presence of noise. We emphasize both stopping rules and restart procedures. The results show that considerable improvement is possible over the proposed settings in the existing literature.

1 INTRODUCTION

Response Surface Methodology (RSM) is an optimization tool that was introduced in the early 50´s by Box and Wilson (1951). It is a collection of mathematical and statistical techniques that is useful for the approximation and optimization of stochastic functions. These techniques are employed in order to estimate the optimization function and to find search directions to sub-regions of the domain with improved and hopefully optimal solutions.

RSM is based on approximations of the objective function by a low order polynomial on a small sub-region of the domain. Using regression analysis based on a number of observations of the stochastic objective function, the best local solution is determined together with a search direction for possible improvement. To this end, the stochastic function is evaluated in an arrangement of points referred to as an experimental design. Many applications for the RSM procedure are performed in a manual setting, for example in physical, engineering, biological, clinical and food sciences (Myers, Khuri, and Carter 1989).

In a manual setting the user can interfere in the optimization process according to his/her personal intuition and likings. In an automated RSM optimization exercise the settings of the algorithm have to be fixed in a systematic manner. We want to design a RSM algorithm that will not stop to ask for input from the user during an optimization run, instead the algorithm will read the input, performs a systematic search for an (local) optimum and reports the optimum back to the user. The OPTQUEST (Glover, Kelly, and Laguna 1999) simulation optimization procedure operates in similar way, yet it is primarily oriented at optimization of discrete decision variables and it uses other techniques.

In this paper we are interested in finding the best settings for such an automated RSM when there is very little information about the objective function. We consider stochastic objective functions with unknown variance and objective functions that are very time consuming to evaluate for each solution. When optimizing a simulation model, one estimates the model parameters that optimize specific stochastic output statistics of the simulation model. In this optimization exercise the simulation model is considered to be a black box. The advantage of such a procedure is that the original simulation model can be left intact, while procedures like infinitesimal perturbation require changes that are not desirable for very complex models.

For an automated RSM to be called successful it should be precise and fast. The procedure should recognize when no further progress is being made and the differences in the subsequent iterations can only be attributed to the noise in the objective function. The procedure should also be able to recognize sub-regions of the domain with optima and it should be able to find directions for improvement.

We will present a framework of the RSM procedures that is founded in recognizing local optima in the presence of noise. It includes feedback iterations, precision checks and restart procedures. We iterate between first-order and second-order approximations in order to continue the search for optima beyond the first-order approximation. We therefore also extensively discuss the use of stopping criteria.

In the literature we found that there are rather confusing and non-systematic recommendations for the settings of automated RSM procedures. To standardize the algorithm we fix some of the possible choices in the process based on existing literature. Other choices are defined in a

number of test algorithms that are compared using test functions and a simulation model well known in the medical literature.

The setup of this paper is as follows. In Section 2 we extensively discuss the setting of the RSM that we found in the literature and we fix a number of the settings based on pre-tests and literature. Other choices for the settings are subject to experiments and the design of the tests is discussed in Section 3. The test functions are described in Section 4 and the test results are given in Section 5. Finally, in Section 6 we discuss the test results and give our recommendations.

2   RESPONSE SURFACE METHODOLOGY

In this section we will describe all steps and procedures of the RSM algorithm with emphasize on the choices that need to be defined for an automated RSM. Without loss of generality, we discuss the RSM method for a minimization problem. The second assumption that we will make is that the objective function is of a stochastic nature, and that we want to optimize the expected value of the stochastic output. Mathematically, this problem can be described by

$$\min f : D \rightarrow \Re, D \subseteq \Re^k \qquad (1)$$

where $f(\xi_1,...,\xi_k)$ is equal to $E(F((\xi_1,...,\xi_k)))$. Here, $F((\xi_1,...,\xi_k)$ denotes stochastic output for given input $\{\xi_1,...,\xi_k\}$, and $E(F((\xi_1,...,\xi_k))$ denotes its expected value. We further assume that the variance in the function values is not known. This situation especially arises in simulation studies, where the objective function can be seen as a black box that returns an output value for a given input. Simulation models do not assume a function form and are subject to an unknown stochastic error. A minimization exercise aims to find the input parameters that result into the minimum output value, such as finding the best set of service employees that will lead to the minimum waiting time for a queuing model.

In general the RSM procedure comprises two phases. In the first phase the objective function is locally approximated by first-order polynomials, in the second phase the objective function is approximated by a second-order polynomial. In both phases we define a region of interest, which is a sub-region of the domain. For the approximations the stochastic objective function is evaluated a number of times in the points of an experimental design, which is a specific arrangement of the points that usually lie on the borders of the region of interest. When the first-order model is found to be adequate a steepest descent procedure is applied to find a new region of interest. Otherwise the RSM moves to the second phase. When a second-order model is approximated and found to be adequate a stationary point needs to be found and classified and an appropriate action should be taken. Usually the algorithm is termi-

nated and the stationary point is returned. However, we will discuss why it is profitable to continue the algorithm beyond the second order approximation especially for a stochastic function with unknown error variance. In particular we will define an extension where we return to a first-order approximation in some cases and we include stopping rules based on the quality of the current center point rather than a certain phase in the optimization process. Figure 1 shows the optimization process on a very global scale and displays when a stopping rule is checked. The dotted arrow shows the proposed extension.
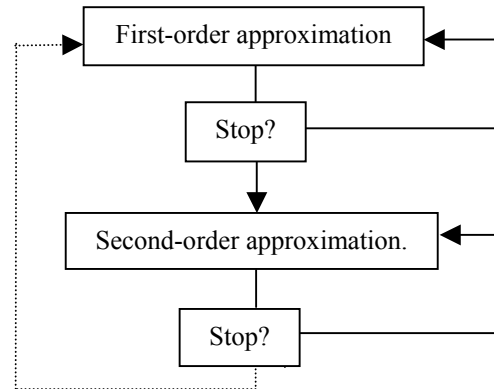

Figure 1: Schematic Overview of the Algorithm

The stopping rules applied are the same after first- and second-order approximation. We define an iteration of the RSM algorithm as the run between two checks of the stopping criteria of the algorithm. There are a number of choices that need to be implemented in an automated RSM procedure using a consistent decision rule. These choices can be divided into "building blocks", "strategic moves" and "stopping rules" of the algorithm. Notice that we follow the steps of the framework proposed by Neddermeijer et al. (2000a). Each step in this framework is either a building block or a strategic move.

The building blocks of the algorithm consist of well-defined procedures that can be used to determine the next move of the algorithm. Strategic moves of the algorithm determine the action taken when a building block returns a result. In the next sections we will discuss the literature and our extensions of the RSM by describing the building blocks and strategic moves. Then we discuss the stopping rules and we discuss the option to restart the algorithm with its initial settings in the current center point. Finally there are a number of parameter settings such as the significance levels of statistical tests.

2.1 Building Blocks

2.1.1 Design for the First-Order Approximation

There are many designs to choose from, like fractional or full factorial, and two-level or three-level designs (Myers

and Montgomery 1995). All these designs can be augmented by the center point of the region of interest. In non-automated optimization the user tries to fit a first-order approximation with different designs, apply coding of the parameters to find better regression estimates or recalculate the objective value in the design points. For instance, replicating the evaluation of the objective function in the center point provides protection against curvature (Myers and Montgomery 1995). For an automated RSM procedure we follow the literature and evaluate the objective function once in the 2^k points of a two-level full factorial design and 5 times in the center point of the current region of interest (Myers and Montgomery 1995; Joshi, Sherali, and Tew 1998). This design is orthogonal and does not require as many points as a three-level full factorial design. In our opinion two-level fractional factorial designs consist of too few points to approximate objective functions with 2 or 3 parameters well enough. Furthermore, full factorial designs can quite easily be augmented to derive a second-order design (see e.g. Neddermeijer et al. 2000a).

2.1.2 Test the First-Order Model for Adequacy

Usually, a test for lack of fit (Weisberg 1985) and a test for significance of regression are performed (Myers and Montgomery 1995). Box and Draper (1987) showed that the test for lack of fit is a joint test for interaction between factors as well as for curvature. In non-automated optimization one can decide to use other tests and one can vary the significance levels based on the results of these tests.

If there is no lack of fit and not all the regression coefficients are equal to zero we will perform the line search. If one of the tests fails then we conclude that this model is not adequate and we fit a second order model.

2.1.3 Perform a Line Search in the Steepest Descent Direction

If the first-order model is found to be adequate a line search is performed from the center point of the current region of interest in the steepest descent direction to find a point of improved response. Numerous implementations of the line search have been proposed (Box and Draper 1987; Myers and Montgomery 1995; Khuri and Cornell 1996; Joshi, Sherali, and Tew 1998; Neddermeijer et al. 1999; Kleijnen, den Hertog and Angün 2003). We will use increments $\Delta_1,...,\Delta_k$ along the path of steepest descent equal to the distance from the center point to the point of intersection of the direction of steepest descent and the sphere given by $\sum_{i=1}^{k} \Delta_i^2 = 1$ (Neddermeijer et al. 1999). In a manual RSM algorithm one can observe the results of a line search and use personal likings to stop the search. However, in automated optimization, the algorithm needs a

stopping rule that recognizes the lack of improvement in response during the line search.

The most straightforward rule ends the line search when an observed value of the objective function is higher than the preceding observation (Del Castillo 1997). We will not use this stopping rule, known as the 1-in-a-row rule, because it is sensitive to the noise from the response surface function. In a similar way, the n-in-a-row stopping rule ends the line search when n observed values of the objective function are higher than the preceding observation. In the Myers and Khuri stopping rule (1979), the line search is ended when the mean response in a line search point, is significantly higher than the mean response in a preceding line search point. This rule requires evaluating the response surface function in a line search point more than once, because the variance of the response is not known at the start of the algorithm. In our algorithms we use the small sample t test in order to compare the mean responses in different points. This statistical test is robust with respect to both non-normality and unequal variances (Wackerley, Mendenhall, and Schaeffer 1996). We will test our version of the Myers and Khuri rule against the 3-in-a-row rule for our setting, where the variance in the stochastic objective function is not known a priori (contrary to the setup by Del Castillo and Myers and Khuri).

2.1.4 Design for the Second-Order Model on the Region of Interest

The regression coefficients of this model are again determined by regression analysis, applied to observations performed in an experimental design. A popular second-order design is the central composite design (CCD; Myers and Montgomery 1995). The CCD arises when the full (or fractional) factorial design is augmented by the first-order design with $2k$ axial points (Box and Wilson 1951). We make this design spherical by choosing the new points such that all points are equidistant to the center point of the current region of interest (Myers and Montgomery 1995). This design is chosen for two reasons. First of all this design can almost be rotated and the loss in rotation is trivial (Myers and Montgomery 1995). Furthermore, if we would use a design that is rotatable, the distance of the new points to the center point would be large as compared to the distance of the existing points to the center point.

2.1.5 Test the Second-Order Model for Adequacy

This module checks if a second-order model describes the behavior of the objective function in the current region of interest. Similar to the first-order model a lack of fit test can be used. Now, the null hypothesis of this test is that the true regression model is quadratic. In non-automated optimization one can use different significance levels or decide to overrule the outcomes of the lack of fit test.

2.2 Strategic Moves

2.2.1 When the First-Order Model is Adequate

If the first-order approximation is found to be adequate, a steepest descent procedure will be applied from the current center to find a new center point as used in most literature (Box and Wilson 1951; Box and Draper 1987; Fu 1994; Myers and Montgomery 1995; Khuri and Cornell 1996; Joshi, Sherali, and Tew 1998). This new point is then used as the center point of the next region of interest. On this new region, the objective function will be approximated again by a first-order model (Myers and Montgomery 1995).

2.2.2 When the First-Order Model is Inadequate

If the first-order model is not accepted, there is some evidence of curvature or interaction between the factors in the current region of interest, or the regression coefficients are all equal to zero. Most references suggest approximating the response surface function by a second-order model (e.g. Fu 1994; Myers and Montgomery 1995; Neddermeijer et al. 2000a). An alternative is to increase the precision of the function evaluation in the design points. However this alternative is very time consuming and does not guarantee that the inefficiency is solved.

2.2.3 When the Second-Order Approximation is Adequate

If the second order approximation is found to be adequate then the appropriate action depends on the location and the nature of the stationary point. It is shown (Greenwood, Rees, and Siochi 1998) that for many functions a first-order model is inappropriate over a large percentage of the domain, so the algorithm can turn to the second phase quite early. The first stationary point found by a second-order approximation is therefore not likely to be the best point in the domain. If we stop the algorithm at this point (Fu 1994; Kleijnen 1998) the optimum could still be located far away from the current region of interest. If a minimum is found inside the region of interest, we can use this point as the center of a new design and a new second order approximation can be performed. We suggest reducing the size of this region of interest because we assume that we are close to the minimum of the objective function. If the stationary point of the second-order polynomial is a maximum, a saddle point or a minimum outside the current region of interest we perform ridge analysis to find a new stationary point that lies inside the current region of interest, because it is not correct to extrapolate the second-order model outside the current region of interest (Myers and Montgomery 1995). We conclude that we are not close to the optimum and propose to return to a first-order approximation in order to find a direction of improvement. By making this choice we now need stopping rules to decide when we are satisfied with the current point. In section 3.2 we will discuss a number of rules and our tests will show the best stopping rules for specific functional forms.

2.2.4 When the Second-Order Model is Inadequate

If the second-order model is found to be inadequate, we assume that either the region of interest is too large or that the stochastic nature of the function disturbs the approximation process. Stopping the algorithm at this point is only a good idea if there is some good indication that the current center point is close to the optimum. If we do not have this indication we propose to continue the algorithm. If we increase the precision used in evaluating a design point, the variance of the response will be reduced and therefore the second-order polynomial will better approximate the objective function. We will thus either reduce the size of the current region of interest (Joshi, Sherali, and Tew 1998) or we will increase the precision used in evaluating the design points.

2.3 Stopping Rules

In automated optimization the RSM algorithm needs to be ended by consistent stopping rules that do not end the algorithm before a good solution is found and also do not unnecessarily prolong the algorithm. In section 2.2. we referred to the RSM literature where the optimization is ended after estimating only one second-order model (Fu 1994; Kleijnen 1998). We recommend ending the optimization if either the estimated response value does not improve sufficiently anymore, or, in case there are budget constraints, if a fixed maximum number of (function) evaluations have been performed. In this section we explain why these criteria seem to be consistent and how we apply them on the automatic algorithm. We also propose one more stopping criterion: stop the algorithm if the input values do not change anymore, i.e. if consecutive center points are close to each other.

2.3.1 The Estimated Response Does Not Improve Sufficiently Anymore (IMPROVE)

Algorithms to find the optimum of a deterministic function can simply be ended if the function value does not improve sufficiently in consecutive iterations. When optimizing stochastic objective functions though, one has to take noise into account. Because we assume that the variance of the response is not known at the start of the algorithm we have to estimate it by evaluating the response in the new center point of the region of interest more than once. We then need some statistical test to determine if there is sufficient progress or if the different mean response in the center points is completely due to the noise. Notice that if the mean response does not decrease significantly in consecutive center points we could still make progress. For in-

stance the mean response can decrease from value 10.2 to 8.1 in 5 iterations, while in each separate iteration the change is not significant. We therefore implement the following criterion. Stop the algorithm if the mean response in the latest center point is not significantly different from the mean response in the penultimate center point in *n* consecutive iterations. It is important to note that the penultimate center point is only changed in case the mean response differs significantly from the mean response in the latest center point. The number of iterations *n* is subject to tests. Notice that this stopping rule is based on the known stopping rules for the steepest descent line search. It makes use of elements of both the Myers and Khuri rule as well as the n-in-a-row rule.

2.3.2 Convergence of Input Values (CONVERG)

Algorithms that are used to find the optimum of a deterministic function are sometimes ended if the input parameters of the function do not change anymore. We also propose to end the algorithm if the Euclidian distance between two consecutive center points is small, i.e. less than $\varepsilon \cdot \sqrt{k}$, where ε is a small number and k is the number of parameters of the objective function. In this way, the precision of each estimated parameter will be approximately ε.

2.3.3 Fixed Maximum Number of Function Evaluations (EVAL)

Our interest in the RSM is especially intended for stochastic models where the calculation of the corresponding stochastic objective function is very expensive or time-consuming. Therefore, ending the algorithm after a maximum number of function evaluations is appropriate when there are budget constraints. Notice that this stopping criterion does not consider the noise.

2.4 A Restarting Mechanism

Because RSM is a local search method there is no guarantee for finding a global optimum. The first center point used in the RSM procedure is chosen by the user or randomly selected and can influence the outcome of the procedure. Neddermeijer et al. (2000a) consider multiple starting points and/or multiple searches from the same starting point, when optimizing a stochastic objective function. In this study, we will use an adjusted mechanism that is based on these suggestions.

To escape from a non-optimal region we propose to restart the algorithm as soon as the algorithm is stopped by one of the stopping criteria using the current best center point found as the new starting point. Because the algorithm cannot escape from a non-optimal region when the size of the region of interest is too small, we propose to reset the size of the region to its initial value. In all optimization runs the best solution, i.e. the parameter values for

which the mean response of the stochastic objective function is best, will be remembered. So this restarting procedure will not deteriorate the solution found in the 'normal' optimization run. Of course the response function is stochastic, so it is possible that the mean response measured in a non-optimal point is better than the mean response measured in the real optimal point.

2.5 Parameter Settings

The parameter settings will all be subject to tests. For instance the significance of the lack of fit test can be performed at a significance level of 2.5%, 5% or 10%. We will also test by how much we have to increase the precision of the function evaluation in the design points and by how much we have to decrease the size of the region of interest to solve the inadequacy of the second order model.

3 TEST DESIGN

In this paper we want to find automated RSM procedures that consistently perform well for stochastic objective functions with unknown variance. For such procedures there are two important issues: how can we find an (area that contains an) optimum and when do we decide that the current center point is a satisfactory solution, i.e. when do we stop the algorithm? There is a trade-off between the running time of the algorithm and the quality of the algorithm. An efficient algorithm produces a solution of high quality within a reasonable time. Our emphasize on stochastic functions further complicates the task for the RSM procedure, we now need to recognize when the difference in response in the sequence of recent center points is only due to noise, and that no further improvement can be expected.

We therefore emphasize stopping rules and restart procedures. In a pre-testing phase we have determined the parameter values that give the best performance with respect to running time and quality of the solutions returned. We also discuss only strategic moves that we feel have the most impact on the efficiency of the RSM procedure. We benchmark our set-up against the set-up by Fu (1994) who stops after the second-order model is found to be inadequate for the first time.

In particular we consider the following alternatives: We test the 3-in-a-row stopping rule against our version of the Myers and Khuri stopping rule for the steepest descent line search in the first phase. The steepest descent procedure should efficiency recognize a new region of interest. If it fails to do so then the second-order approximation, that performs locally, will be of no use. Del Castillo (1997) concludes that the Myers and Khuri stopping rule dominates the 3-in-a-row rule for stochastic functions with known variance. When the variance is unknown and it should be determined by the algorithm it is not clear whether the Myers and Khuri rule is still as successful.

When the second-order model is found to be adequate and a minimum is found within the region of interest we assume that we are close to a (local) optimum. We then shrink the region of interest with either 50% or 10%. The exact values are set by pre-testing, but the difference between these shrink-percentages is important. When the region of interest is shrunk by 10% it will take more time to focus on a certain region that is suspected to contain the global optimum and the algorithm could be unnecessarily prolonged. However if we shrink the region of interest too soon, we are at higher risk of returning a local optimum. The success of the alternatives will closely interact with the stopping rules and restart procedures. For instance, we have good hope that the 50% shrink procedure (ensures fast convergence) in connection to a strict stopping rule with respect to noise (ensures fast convergence) and a restart procedure that returns to the original size of the region of interest (recognize quality of solutions) will perform well.

When the second-order approximation is rejected either the noise is dominant or the region of interest is too large. Since we do not believe that there is any justification that the current center point is close to the optimum we want to continue either with a second-order model in the neighborhood of the current center point or with a first-order approximation. We have no intuition which correction (reduce noise or shrink region of interest) will give better results and test both options.

As a result we will test eight algorithms, using all proposed alternatives in combination with the others (see Table 1). For all these eight algorithms we will study the effect of the stopping rules and restart procedures explained in Section 2.3 and 2.4.

Table 1: The Test Design

Algo-rithm Nr	Stopping rule steepest descent	Shrink region of interest	Solve second-order inade-quacy
1	3-in-a-row	50%	Noise reduction
2	3-in-a-row	50%	Shrink design
3	3-in-a-row	10%	Noise reduction
4	3-in-a-row	10%	Shrink design
5	Myers and Khuri	50%	Noise reduction
6	Myers and Khuri	50%	Shrink design
7	Myers and Khuri	10%	Noise reduction
8	Myers and Khuri	10%	Shrink design

The set-up of the tests is as follows. First we apply the algorithms for a large number of iterations and for each algorithm record the relevant values for the application of the stopping criteria. Especially, at each iteration we record (the mean response in) the current (best) center point, the number of function evaluations, and the Euclidean distance between the current and the last center point. We then decide on the exact implementation of each stopping rule such that the average performance of each algorithm for a

test set defined in the next section is "best". Each algorithm can have a different implementation of the stopping rules for which it performs best. The criteria for the preferred performance of the stopping rules are based on the quality of the solutions compared to the optimum values of the deterministic test functions. Notice that we have full knowledge of the optima of the test functions and corresponding solution values. We can therefore observe the parameters of each stopping rule of the RSM procedure for every test function and determine when the procedure is no longer effectively improving the solutions. The parameter values are taken such that the average performance over the test functions is best because we assume that we have no previous knowledge of the objective function. Note that we apply each algorithm on each test function 100 times to find the average behavior of the algorithm.

It should be said that the optimality of a solution comprises two things: the response of the objective function and the parameter values. In chemical processes it may be more important to find a minimum response, whereas in the estimation of parameters of a simulation model the values of the parameters are more important. Therefore, we will not only compare the minimized stochastic output with the minimum response, but also the estimated parameter values with the actual location of the minimum.

In a second phase we run the algorithms again for the new implementations of the stopping rules and we apply a restart procedure when each algorithm is stopped. The restart procedure is applied once and the result of the algorithm after the restart is returned as the output of the algorithm.

The performance of the algorithms is compared to the performance of four algorithms using the stopping criterion of Fu applied to the same set of test functions.

4   THE TEST PROBLEMS

In this study we will test the optimization algorithms on a set of 7 test functions with known optima, which consist of a deterministic term and a stochastic term. These functions, e.g. the Rosenbrock, Powell, Beale and Wood function, are also used in comparing different versions of Nelder and Mead simplex method (NMSM) (Neddermeijer et al. 2000b) and in comparing RSM algorithms with NMSM algorithms (Neddermeijer et al. 1999). We will also consider a (micro)simulation version of an existing cancer-screening model (Day and Walter 1984). This model has three parameters that need to be estimated from an observed data set using constrained optimization of a goodness-of-fit statistic (van Oortmarssen et al. 1990; Neddermeijer et al. 1999). For this particular model the optimal parameters can also be determined analytically, but we are interested in the performance of the RSM procedure for this simulation exercise. For a more detailed description of above test problems we refer to Nicolai and Dekker (2004).

5 RESULTS

From a number of preliminary tests we obtained the parameter settings as described in section 2.5. Table 2 shows the settings we use in all algorithms.

Table 2: Parameter Settings

Parameter	Setting
Significance level statistical tests	5%
Shrink design	50%
Increase precision	0.95 / 25%

Table 2 shows that all statistical tests are performed at a significance level of 5% and the size of the design will be decreased by 50% to solve the inadequacy of the second order model. For the first seven test functions the precision of the function evaluation will be increased by multiplying the standard deviation of the stochastic part by 0.95. For the micro-simulation model the number of simulated life histories will be increased by 25%.

In the first phase of the numerical experiments we want to determine the exact implementations of the stopping criteria. Because we cannot compare the mean of the solutions of an algorithm over iterations we have recorded the results of the eight algorithms for a number of implementations of the stopping criteria. These implementations are shown in Table 3. Note that these implementations may depend on the number of parameters (nPar) of a test problem.

Table 3: Implementations of Stopping Criteria

Criterion	Implementation
IMPROVE	$5 \cdot i$, $i = 1,2,3,4,5$
CONVERG	$5E\text{-}04 j \cdot \sqrt{nPar}$, $j = 1,2,4,10,20$
EVAL	$50 \cdot k \cdot 2^{nPar}$, $k = 1,2,3,4,5$

*nPar: the number of parameters of a test function

In Table 4 the results of the application of algorithm 1 on test problem 6 are shown. The first column contains the description of the stopping criterions and in the second column the setting of each criterion is given. The third column shows the mean (standard deviation) of the absolute difference between the actual optimum and the actual value (i.e. without noise) of the objective function in the found minimum. In the fourth column the mean (standard deviation) of the Euclidean distance between the minimum and the found minimum is given. The last row actually shows the results of stopping the algorithm after the maximum number of iterations, in this case 150. Note that this particular test problem has two parameters.

The results in Table 4, which are representative for the majority of the results of the first phase of numerical experiments, show that the more iterations the algorithm runs the lower the actual mean response becomes. However, it appears that the Euclidean distance between the es-

Table 4: Results of Applying Algorithm 1 on Function 6

Criterion	SETTING	Difference	DISTANCE
IMPROVE	5	0.329(0.307)	1.877(0.029)
	10	0.283(0.190)	1.874(0.035)
	15	0.266(0.177)	1.876(0.036)
	20	0.255(0.168)	1.878(0.045)
	25	0.254(0.171)	1.878(0.044)
CONVERG	1.41E-02	0.319(0.302)	1.877(0.038)
	7.07E-03	0.264(0.189)	1.878(0.046)
	2.83E-03	0.249(0.165)	1.879(0.052)
	1.41E-03	0.241(0.156)	1.880(0.054)
	7.07E-04	0.243(0.158)	1.880(0.054)
EVAL	200	0.347(0.317)	1.875(0.024)
	400	0.275(0.198)	1.873(0.035)
	600	0.263(0.176)	1.874(0.038)
	800	0.262(0.176)	1.877(0.042)
	1000	0.241(0.155)	1.883(0.056)
Max. iterations	150	0.228(0.141)	1.882(0.074)

timated parameters and the actual location of the minimum does not always become smaller as the algorithm runs longer. In general we may conclude that the algorithms iteratively find solutions with smaller and smaller mean responses although the difference with the actual best parameter values does not change so much anymore. In general it appears that the CONVERG stopping criterion outperforms the other two criterions with respect to the absolute difference and Euclidean distance (Nicolai and Dekker 2004).

Note that in Table 4 the standard deviation of the difference between the actual minimum value and the actual objective value in the found minimum is relatively large. It actually appears that the algorithm also finds solutions for which the deviation is very close to zero. Therefore, applying the algorithm more than once gives a high probability of finding a good solution. Since in practice the actual values are not known, the estimated response in the found solutions must be compared to determine which solution is better.

As it was stated before, the performance of an algorithm does not only depend on the precision of the solutions found, but also on the computing time. When optimizing stochastic objective functions the number of function evaluations needed to obtain a certain solution is an important indicator for the computing time. In Table 5 the average (standard deviation) of the number of evaluations needed to fulfill the stopping criteria IMPROVE and CONVERG when algorithm 1 is applied on function 6 are shown.

The values in Table 5 show that the difference in the number of function evaluations between the least restrictive setting and the most restrictive setting of each stopping criterion is a factor 5. Moreover, the number of evaluations in running the algorithm 150 iterations is three times higher than the number of evaluations done for the most restrictive setting of the CONVERG stopping rule. For this specific algorithm applied on function 6 the gain in precision, see Table 4, does not balance the extra computing time.

Table 5: Number of Function Evaluations Required to Fulfill the Stopping Criteria

Criterion	SETTING	EVAL
IMPROVE	5	279(89)
	10	575(376)
	15	886(572)
	20	1226(763)
	25	1424(762)
CONVERG	1.41E-02	274(110)
	7.07E-03	442(204)
	2.83E-03	702(292)
	1.41E-03	878(336)
	7.07E-04	1024(361)
Max. iterations	150	3331(106)

If we combine the results in Tables 4 and 5, we see that algorithm 1 minimizes test function 6 more precisely using less function evaluations. However, in general it appears that the number of evaluations required to fulfill stopping criterion CONVERG is much higher than for IMPROVE (Nicolai and Dekker 2004). This especially holds for the test functions with more than three variables. In the second phase of the numerical experiments we compare the above-defined eight algorithms with four algorithms, which use the stopping criterion of Fu. Moreover, we test whether the restart mechanism has an effect on the solutions.

In Table 6 the results of applying algorithm 1 using the stopping criterion of Fu are shown. Note that we use the numbering of the algorithms as given in Table 1 and that we have applied algorithms 1, 2, 5 and 6 using the stopping criterion of Fu. In these algorithms the option of decreasing the size of the design is not used, since the Fu algorithms stop after one second-order model. In the first column of Table 6 the number of the test function is given. Test function number 8 is the micro-simulation model. In the second and third column the average (standard deviation) of the absolute deviation between the actual objective value in the optimum and the actual value of the objective function in the found minimum are given, where the third column contains the results for applying the restarting mechanism. The fourth column shows the mean number of function evaluations that the algorithm carried out to find the minimum, the restarting mechanism included.

Table 6: Results of Applying Fu Algorithm 1

Test Function	No Restart	Restart	EVAL
1	9.92(86)	1.28(0.40)	127
2	19334(0)	18287(0.01)	108
3	0.96(1.16)	0.242(0.25)	61
4	1102(3524)	120(1200)	455
5	7.83(19.8)	0.33(0.36)	2264
6	1207(701)	679(739)	80
7	4203(3174)	2664(3069)	387
8	0.109(0.11)	0.094(0.099)	54

From extensive experimentation (Nicolai and Dekker 2004) it appears that algorithm 1 outperforms the other algorithms, so we continue with this algorithm. The results in Table 6 show that the restarting mechanism increases the precision of the found solutions very much. However, if we compare the solutions for test function 6 in Table 6 with the results in Table 4, we see that even applying the least restrictive version of stopping criterion CONVERG results in a far more precise solution. Note that the average number of evaluations for the Fu algorithm is 80, whereas it is 274 for test algorithm 1. In general it appears that the Fu stopping criterion ends the algorithm too soon (Nicolai and Dekker 2004).

6 CONCLUSIONS

In this paper we have analyzed settings for an automated RSM procedure for simulation optimization. Numerical results show that considerable improvement is possible over the proposed settings in the existing literature. Stochastic objective functions with unknown variance need different settings than objective functions with known variance in order to find good solutions. The trade-off between running time of the algorithm and the quality of the solutions returned can be tuned by setting the appropriate stopping criteria and applying restart procedures.

REFERENCES

Box, G.E.P., and N.R. Draper (1987). *Empirical Model-Building and Response Surfaces*. New York: John Wiley & Sons.

Box, G.E.P., and K.B. Wilson (1951). On the Experimental Attainment of Optimum Conditions. *Journal of the Royal Statistical Society* B13**,** 1-38, discussion: 38-45.

Day, N.E., and S.D. Walter (1984). Simplified Models of Screening for Chronic Disease: Estimation Procedures from Mass Screening Programmes. *Biometrics* 40, 1-14.

Del Castillo, E. (1997). Stopping Rules for Steepest Ascent in Experimental Optimization. *Communications in Statistics. Simulation and Computation* 26(4), 1599-1615.

Fu, M.C. (1994). Optimization via Simulation: A Review. *Annals of Operations Research* 53, 199-247.

Glover, F., J. Kelly, and M. Laguna (1999). New advances for wedding optimization and simulation, *Proceedings of the 1999 Winter Simulation Conference*, eds. P. Farrington, H. Nembhard, D. Surrock and G. Evans, 255-259. Piscataway, NJ: IEEE.

Greenwood. A.G., L.P. Rees, and F.C. Siochi (1998). An Investigation of the Behavior of Simulation Response

Surfaces. *European Journal of Operational Research* 110, 282-313.

Joshi, S., H.D. Sherali, and J.D. Tew (1998). An Enhanced Response Surface Methodology (RSM) Algorithm Using Gradient Deflection and Second-Order Search Strategies. *Computers and Operations Research* 25(7/8), 531-541.

Kleijnen, J. P. C., D. den Hertog, and E. Angün (2003). Response surface methodology's steepest ascent and step size revisited, *European Journal of Operational Research* (in press).

Kleijnen, J.P.C. (1998). Experimental Design for Sensitivity Analysis, Optimization, and Validation of Simulation models. In *Handbook of Simulation: Principles, Methodology, Advances, Applications and Practice*, ed. J. Banks, 173-223. New York: John Wiley & Sons.

Khuri, A.I., and J.A. Cornell (1996). *Response Surfaces: Designs and Analyses*. New York: Marcel Dekker, Inc.

Myers, R.H., and A.I. Khuri (1979). A New Procedure for Steepest Ascent. *Communications in Statistics. Theoretical Methods* A8(14), 1359-1376.

Myers, R.H., A.I. Khuri, and W.H. Carter, Jr. (1989). Response Surface Methodology: 1966-1988. Technometrics 31(2), 137-157.

Myers, R.H., and D.C. Montgomery (1995). *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*. New York: John Wiley & Sons.

Neddermeijer, H.G., G.J. van Oortmarssen, N. Piersma, and R. Dekker (2000a). A Framework for Response Surface Methodogy for Simulation Optimization. In *Proceedings of the 2000 Winter Simulation Conference*, eds. J.A. Joines, R.R. Barton, K.Kang, and P.A. Fishwick, 129-136. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Neddermeijer, H.G., N. Piersma, G.J. van Oortmarssen, J.D.F. Habbema, and R. Dekker (1999). *Comparison of Response Surface Methodology and the Nelder and Mead Simplex Method for Optimization in Microsimulation Models*. Econometric Institute Report EI-9924/A, Erasmus Universiteit Rotterdam, The Netherlands.

Neddermeijer, H.G., G.J. van Oortmarssen, N. Piersma, R. Dekker, and J.D.F. Habbema (2000b). *Adaptive Extensions of the Nelder and Mead Simplex Method for Optimization of Stochastic Simulation Models*. Econometric Institute Report EI2000-22/A, Erasmus Universiteit Rotterdam, The Netherlands.

Nicolai, R.P., and R. Dekker (2004). *Automated Response Surface Methodology for Stochastic Optimization Models with Unknown Variance*. Econometric Institute Report EI2004-37, Erasmus Universiteit Rotterdam, The Netherlands.

Van Oortmarssen, G.J., J.D.F. Habbema, J.Th.N. Lubbe, and P.J. van der Maas (1990). A Model-based Analysis of the HIP Project for Breast Cancer Screening. *International Journal of Cancer* 46, 207-213.

Wackerley, D.D., W. Mendenhall, and R.L. Scheaffer (1996). *Mathematical Statistics with Applications*. 5th Edition, Duxbury.

Weisberg, S. (1985). *Applied Linear Regression.* New York: John Wiley & Sons.

AUTHOR BIOGRAPHIES

ROBIN P. NICOLAI is a Ph.D. Candidate at the Department of Econometrics and Operations Research of the Erasmus University Rotterdam, The Netherlands. He received a M.Sc. degree in econometrics at the same university. His research interests include optimization of stochastic objective functions and maintenance optimization. His email address is <rnicolai@few.eur.nl>.

ROMMERT DEKKER is a full professor in operations research of the Department of Econometrics and Operations Research at the Erasmus University Rotterdam. He obtained his Ph.D. in operations research at the State University of Leiden and his M.Sc. degree in industrial engineering form the Twente University of Technology. His current research interests are: maintenance and logistics (inventory control, spare parts, containers and reverse logistics). He has many times applied simulation models in various logistical activities. His email address is <rdekker@few.eur.nl> and his web address is <www.few.eur.nl/few/people/rdekker>.

NANDA PIERSMA was assistant professor in operations research at the Econometric Institute of the Erasmus University Rotterdam. She obtained her Ph.D. in operations research and her M.Sc. degree in mathematics at the University of Amsterdam.

GERRIT J. VAN OORTMARSSEN is assistant professor of Public Health at the Department of Public Health. His research interest is (simulation) modeling of public health interventions for diseases, in particular applied to screening for early detection of cancer, and control of tropical infectious diseases. His email address is <vanoortmarssen@mgz.fgg.eur.nl>.