Combinatorial Chemistry & High Throughput Screening, **2007,** ***10,*** **149-159**

Optimisation Methodologies and Algorithms for Research on Catalysis Employing High-Throughput Methods: Comparison Using the Selox Benchmark

Sílvia Raquel Morais Pereira[*,a], Frédéric Clerc[b], David Farrusseng[b], Jan Cornelis van der Waal[a] and Thomas Maschmeyer[*,c]

[a]*Laboratory of Applied Organic and Catalytic Chemistry, Delft University of Technology, Julianalaan 136, 2628 BL Delft, The Netherlands*

[b]*Institut de Recherches sur la Catalyse, 2 Av. A. Einstein, 69626 Villeurbanne Cedex, France*

[c]*Laboratory of Advanced Catalysis for Sustainability, School of Chemistry – F11, The University of Sydney, NSW 2006, Australia*

Abstract: The Selox is a catalytic benchmark for the selective CO oxidation reaction in the presence of H_2, in the form of mathematical equations obtained *via* modelling of experimental results. The optimisation efficiencies of several Global Optimisation algorithms were studied using the Selox benchmark. Genetic Algorithms, Evolutionary Strategies, Simulated Annealing, Taboo Search and Genetic Algorithms hybridised with Knowledge Discovery procedures were the methods compared. A Design of Experiments search strategy was also exemplified using this benchmark. The main differences regarding the applicability of DoE and Global optimisation techniques are highlighted. Evolutionary strategies, Genetic algorithms, using the sharing procedure, and the Hybrid Genetic algorithms proved to be the most successful in the benchmark optimisation.

Keywords: Catalysis, benchmark, DoE, optimisation algorithms, GA, SA, Taboo.

INTRODUCTION

The wide acceptance of High Throughput Experimentation (HTE) and combinatorial methods in recent years has opened a broad range of possibilities to the catalyst researcher [1, 2]. However, the automation and parallelization of the experimentation poses new challenges to the chemist in the planning of experimental work so as to take full advantage of HTE capabilities. Chemometric research methods [3, 4], initially dedicated to tackle analytical chemistry issues, are nowadays receiving special interest from the catalyst HTE community and a considerable number of studies has been performed in order to improve the understanding of these methods and to adjust them as well as other computer-science methods to the field of catalyst optimisation [5]. A method used already for many years within the combinatorial catalysis field is Design of Experiments (DoE). Applications include catalyst formulation and preparation [6-9], catalytic kinetic modelling [10, 11], reactor engineering [12, 13], and the optimisation of catalytic reaction conditions [14-18]. For the latter, DoE is usually well-suited, since for this application a well defined parameter space is commonly employed, the general understanding of a given chemical system is targeted and, by using descriptors, successful modelling of the response surface can be achieved in the presence of discrete variables [19]. When the goal is the optimisation of a heterogeneous catalyst composition, however, global stochastic algorithms appear to be promising alternatives. These are attractive due to a) the large parameter space, b) the non-linear shape of the response surface, where synergistic effects are commonly encountered, and c) the difficulty of defining DoE descriptors that characterise heterogeneous catalysts [20, 21]. Under these conditions the use of DoE methodology might not be the most appropriate since the optimum can be easily missed.

Usually in Catalysis, catalytic properties cannot be predicted from physical models. This implies that the catalysts (candidate solutions) must be synthesized and tested. Catalyst screening and optimization is a tedious task even with the help of automated parallel equipments. As a consequence, the optimisation procedure should be reliable and should minimize the risks of failures (e.g. optima not found). This represents a substantial increase in the effort required to asses the reliability of algorithms and methodologies. In addition, unlike the solution of a mathematical equation, laboratory experimentation implies experimental error and outliers. Algorithms robust with respect to noisy data are therefore required for catalyst optimisation [50]. Furthermore, adjusting and testing the applicability of the different algorithms to different types of chemical problems is essential. Much effort has been recently invested in validating optimisation algorithms using catalytic benchmarks. For instance, Genetic Algorithms search procedures have been studied in virtual catalytic benchmarks in the form of a Neural Network [22, 28, 32, 39], or in some form of virtual mathematical benchmarks. While several of these methods have been tested separately, few studies compare their performances directly on the same benchmark.

---

*Address correspondence to either author at the Laboratory of Applied Organic and Catalytic Chemistry, Delft University of Technology, Julianalaan 136, 2628 BL Delft, The Netherlands;
E-mail: s.r.m.pereira@tnw.tudelft.nl or Laboratory of Advanced Catalysis for Sustainability, School of Chemistry – F11, The University of Sydney, NSW 2006, Australia; E-mail: th.maschmeyer@chem.usyd.edu.au

1386-2073/07 $50.00+.00

In this paper a general overview of the application of common chemometric optimisation methods in the field of catalyst optimisation is presented. Several global optimisation algorithms are compared to each other using the same mathematical function, or benchmark, which derived from a dataset on the Selox reaction. Finally, Design of Experiment methodology is also studied using the same Selox benchmark.

1. THE SELOX BENCHMARK

A large data set for the selective catalytic oxidation reaction of CO in the presence of H_2 (for fuel-cell applications) was modelled to obtain the Selox benchmark [23]. The parameter space modelled includes the factors relating to catalyst composition and reaction temperature effects on the conversion and selectivity of CO oxidation. The catalyst composition parameters considered were: the type and amount of metal (NM, [NM]%), acting as the main active phase, the type and amount of transition metal (TM, [TM]%) acting as support (Table **1**). By modelling, for every combination of the remaining discrete parameters, the effects of varying [NM%] and [TM%] on the fitness function, 81 quadratic functions were obtained - one for each of the conversion and selectivity responses (Fig. **1**). The application of a desirability function that combined the two responses and a penalty for high reaction temperatures formed the Selox benchmark response surface (for more details see ref. [23]). The Selox benchmark has the advantage of being composed of a series of mathematical functions and, therefore, it can be easily optimised computationally. Moreover, it also has the advantage of being a real catalytic case study the response surface of which can be easily visualised. It comprises a series of sub-optimal surfaces and its maximum fitness value is 0.935. The optimal areas (fitness > 0.92) are situated in the Pt-Zr-Nb and Pt-Ti/Zr-V sub-regions of the response surface (Fig. **1b**).

Table 1. Parameter Space for the Selox Benchmark

Parameter	Level 1	Level 2	Level 3
NM	Au	Cu	Pt
TM	Mo	Nb	V
Support	CeO2	TiO2	ZrO2
Temp. (°C)	200	250	300
[NM] (mass %)	0.1	…	2.1
[TM] (mass %)	1	…	5

1.1. Global Optimisation Algorithms

A common methodology for the optimisation of catalytic problems is the use of global optimisation strategies. Unlike, for example, Simplex and Gradient Descent local optimisation methodologies, which are deterministic algorithms, stochastic global optimisation algorithms do not, as long as the necessary number of iterations is performed, get trapped in local optima. Among the global methodologies available, Evolutionary Algorithms (EA) represent currently the most popular method which has found a wide range of applications in chemistry [24]. After the introduction, by Wolf *et al.* of this optimization strategy in heterogeneous catalyst formulation, the interest in this approach has risen considerably [25-32]. The increasing familiarity with this methodology, and its efficiency in the discovery of new catalytic materials, make EA a methodology of choice for this kind of application. Advanced Genetic Algorithms (GA) are also being tested in order to further improve the optimisation efficiency. Most of these enhancements deal with hybridization of the GA algorithm with learning techniques such as Neural Networks, knowledge learning systems and others. Different promising algorithms like Simulated Annealing (SA) [33-35], and Taboo Search (TS) [36], while being widely used in other scientific fields, have been used to a lesser extent by chemical researchers and their advantages have not been sufficiently explored. Still other search algorithms recently applied to heterogeneous catalysis include MAP [37], Kriging [38], Holographic [39, 40] search, etc. In the following section, a brief set of explanations regarding search algorithms to be tested on the Selox benchmark is presented.

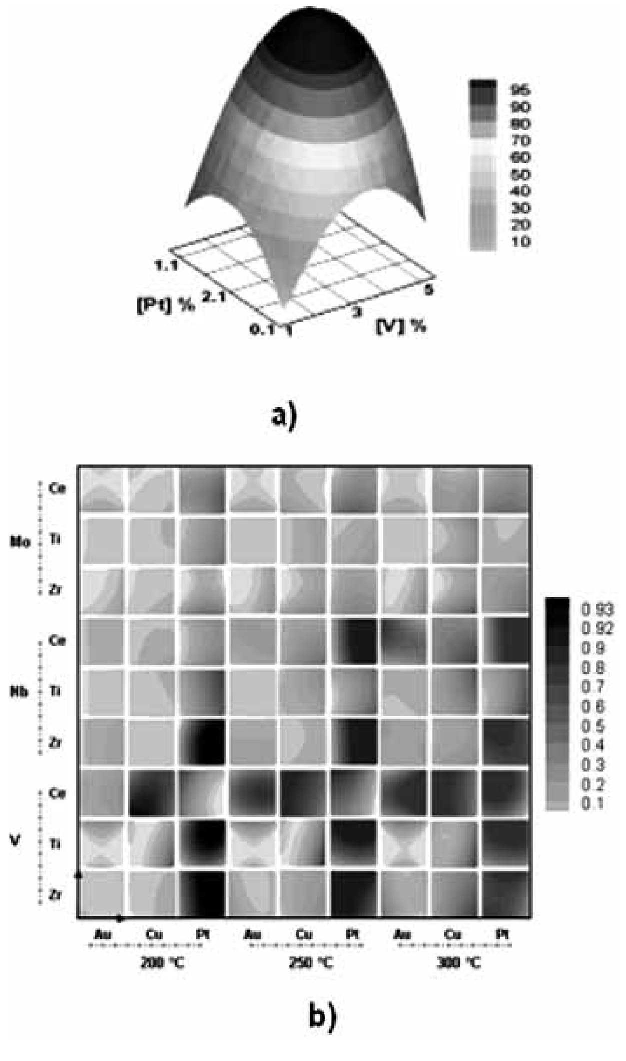

Fig. (1). Response surfaces. **a**) Calculated conversion response surface for the ternary system Pt-V-Ti at 200°C ($y = 53 - 12[TM] - 29[NM] + 8{,}67[NM]^2 - 30[TM]^2 + 20[NM][TM]$). **b**) Contour Plot representation of the total desirability response surfaces.

Evolutionary Strategy (ES) and Genetic Algorithms (GA)

Genetic Algorithms (GA), and also Evolution Strategies (ES), try to mimic the evolutionary process of living species by using similar genetic operators such as selection, crossover and mutation [41, 42]. Both share the same basic concepts, but differ in the way they encode the solutions. Genetic Algorithms use chromosomes composed of binary code, whereas the evolutionary strategies use a real-vector coding representation [43]. In both methods a random initial population is evaluated, from which the strings of the selected catalysts are recombined and mutated, creating a new population that is evaluated again. Performing the loop once is called a generation, and this is then repeated until a termination criterion (for example a maximum number of catalysts tested, or convergence to a catalyst) is met. The degree of browsing/exploitation of the search system is settled by adjusting its selection pressure, crossover rate and mutation rate parameters [23]. Results obtained for the Selox benchmark with a GA algorithm are shown in Fig. **2** and the workflow chart for a GA and an ES are presented in Fig. **3a**.

Simulated Annealing (SA)

Simulated Annealing algorithms work as an analogy to the way in which a heated metal changes towards a minimum-energy crystalline structure on cooling (the annealing process). If it is cooled quickly, it will solidify in a less organised and higher potential-energy state then when cooled slowly. The method can be generalised to a combinatorial approach in a straightforward way [44, 45]. The state of the thermodynamic system is analogous to the candidate solution, e.g. the catalyst, and the energy of the state are analogous to the value of the objective function for the specific solution. The perturbations to move to another state can be compared to moving to a neighbour candidate solution and the ground state to the global minimum or the final solution found by the algorithm. The temperature T is a parameter within the algorithm. Its initial value and the way in which it is decreased during the optimisation (CT, Cooling temperature scheme) controls the degree of system browsing and the algorithm's optimisation speed. The workflow chart for a SA procedure is presented in Fig. **3b**. In this study the simulated annealing tested is a classic single-candidate algorithm. However, parallel SA optimisations that consider, for the same temperature, several neighbourhood candidate solutions have already been implemented [46].

Taboo Search (TS)

The word Taboo comes from Tongan, a language of Polynesia, where it was used by the aboriginal inhabitants of the Tonga island to indicate things that cannot be touched because they are sacred [47]. A recent meaning of the word is a social prohibition imposed as a protective measure. It is on the original meaning that the taboo algorithm is based. The most important association with this meaning is that taboos are transmitted by means of social memory which is subject to modification over time. In this way also the Taboo

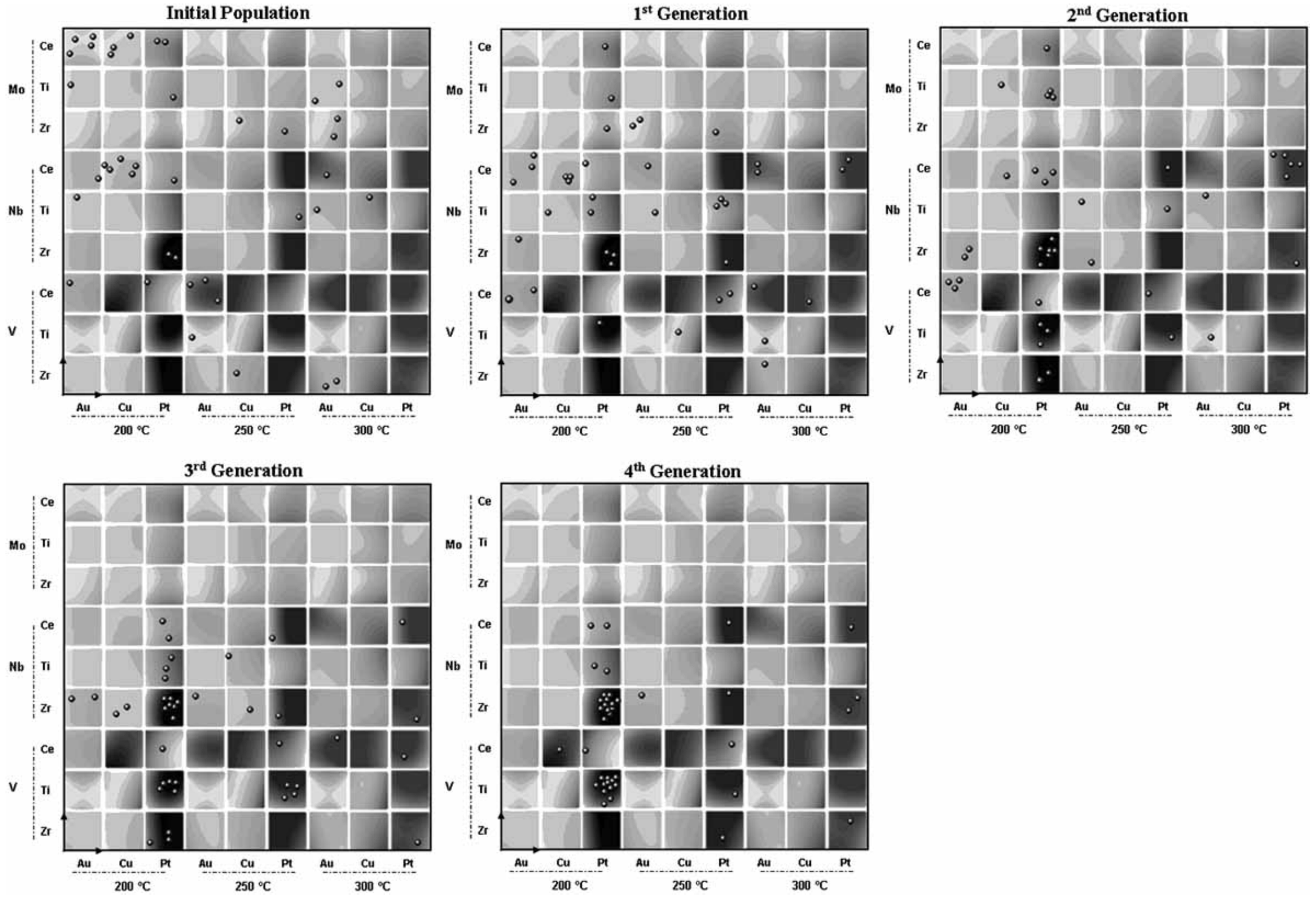

Fig. (2). Scatter plot of the initial random population and the subsequent 4 generations on the Selox benchmark for the GA_1 algorithm settings (Table **2**).

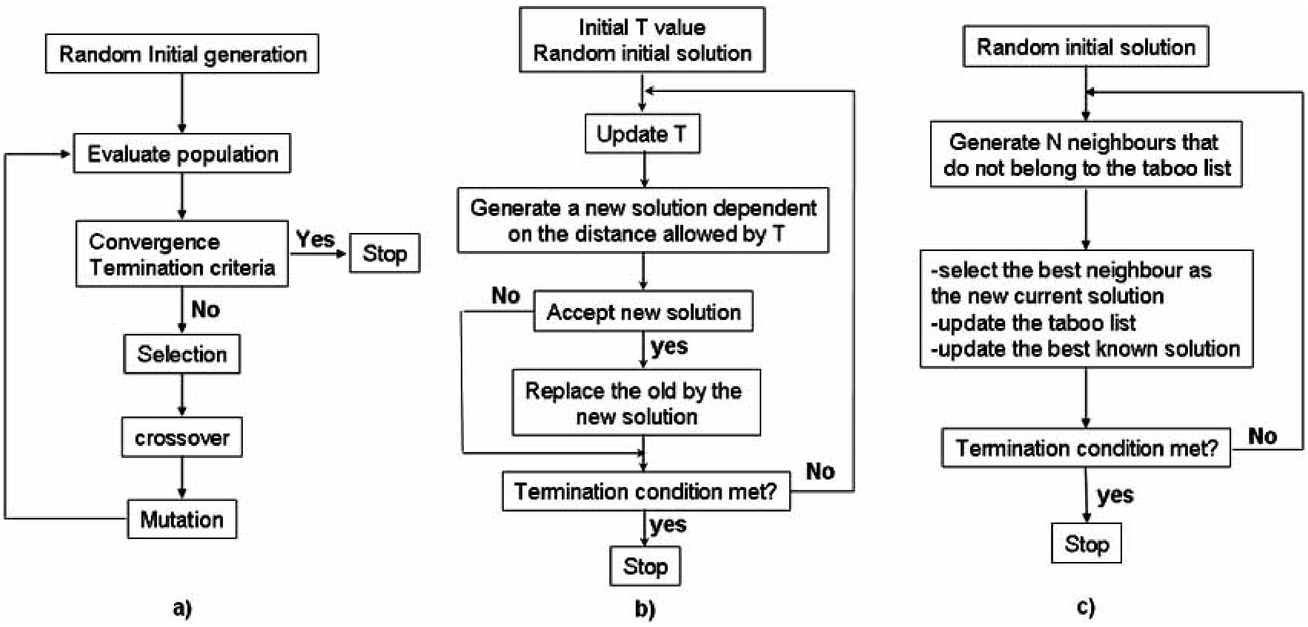

Fig. (3). Flow charts of the algorithm strategies. **a**) Genetic/Evolutionary Algorithm, **b**) Simulated Annealing, and **c**) Taboo Search algorithm.

search memorizes a list of taboo candidate solutions that are not repeated in the subsequent iteration and that is updated at each iterative step. The Taboo algorithm uses a neighbourhood search procedure to iteratively move from a candidate solution to a selected one in its neighbourhood, until some stopping criterion has been satisfied. The new candidate solution, is accepted if it has not previously been memorised as a taboo, or bad, candidate solution [48]. Taboo Search is a deterministic algorithm. The workflow chart for a TS is presented in Fig. **3c**.

Hybrid Algorithms

Several efforts have been made to improve the GA algorithms by introducing some sort of knowledge learning procedure into its search structure. Knowledge discovery algorithms [49-51] and Neural Networks [52-54], among others, have been hybridized with GA in order to extract knowledge from the past generations and use it in designing the new candidate solution generation. In this paper a multi-linear regression procedure was introduced in the GA (GA-LR), consisting of a mathematical function built according to the evaluation of the previous candidate solutions. The new individuals are estimated and proposed according to this mathematical function. The Zone Definer (GA-ZD) GA hybrid includes a special supervised learning algorithm which divides the search space into zones the boundaries of which are based on the average value of each predictive variable. Zone Definer is a modification of the *k-d-trees* algorithm [55, 56]. For the prediction, the estimation of an unknown candidate solution value will be the value of the zone it belongs to [57]. The linear-regression learning procedure models the search space from a general point of view, while the zone-definer learning approach is a non-linear and local procedure.

1.2. Comparison of Global Optimisation Algorithms Using the Selox Benchmark

Candidate Solution Representation

The catalyst candidate solutions were coded for all the algorithms, except Evolutionary Strategy (ES), as bit-strings composed of binary numbers (0,1). Each string comprised 30 digits, 4 for each discrete variable and 7 for each continuous variable. The binary code (base 2) was converted to decimal numbers and the corresponding value for the encoded variable calculated from that number. An example of coding and decoding of the bit-string representation is shown in Fig. **4a**. Vector coding representation of the variables was used for the Evolutionary Strategies as described in Fig. **4b**.

Algorithm Settings

Prior to the experimental optimisation the algorithm settings need to be defined. This will influence the speed of optimisation and browsing by the algorithm. The efficiency of the optimisation is dependent on whether the settings adopted are appropriate for the problem to optimise [23]. However, since the knowledge about the shape of the response surface is usually not known *a priori*, chemists have usually to base their choice on settings commonly adopted and/or previous experience. In our study commonly used settings were adopted. For SA, three different Cooling Temperature (CT) schemes were studied (SA_1, SA_2 and $SA_{3;}$ Table **2** and Fig. **5**), which in this case defines the neighbourhood size change allowed at each iterative step, controlling the browsing/exploration speed of the optimisation.

For the Taboo Search (TS) a neighbourhood size of 5 neighbours of the last retained best individual for generation and evaluation was considered. The best of these becomes the new reference individual and is marked as taboo. For the

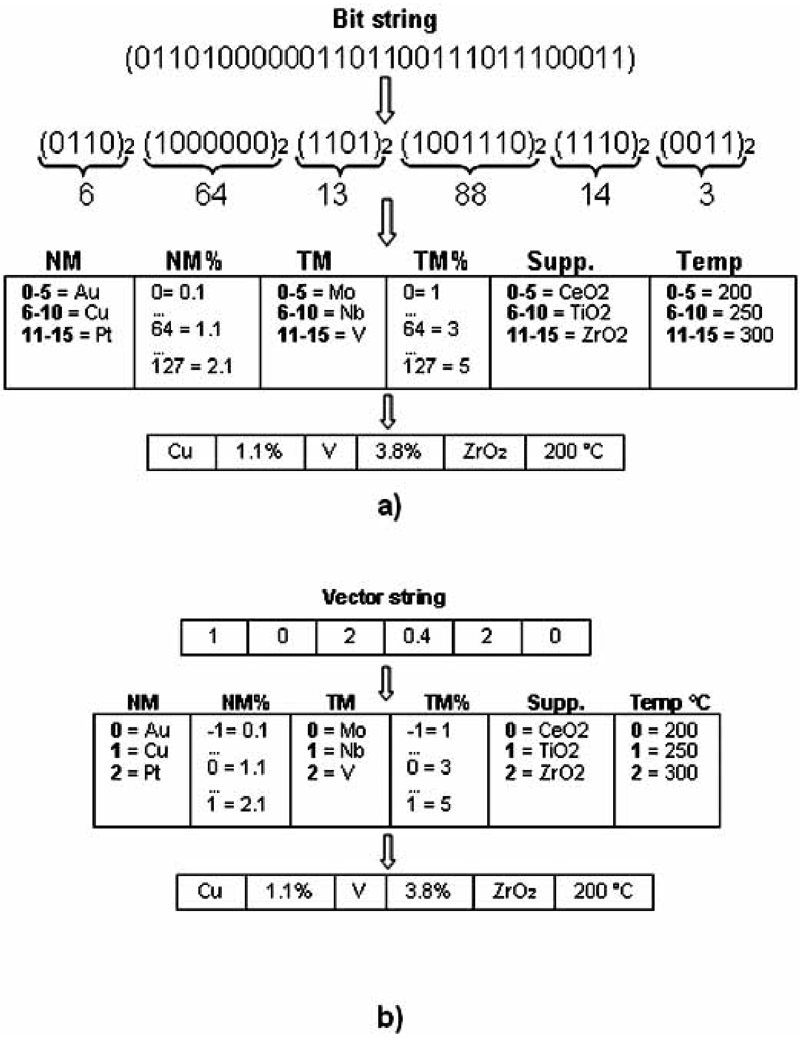

Fig. (4). Candidate solution representation. **a**) Binary string representation and decoding for the GA, SA, TS, random algorithms, and the ZD-GA and LR-GA Hybrid algorithms. **b**) Vector string representation and decoding for the ES algorithms.

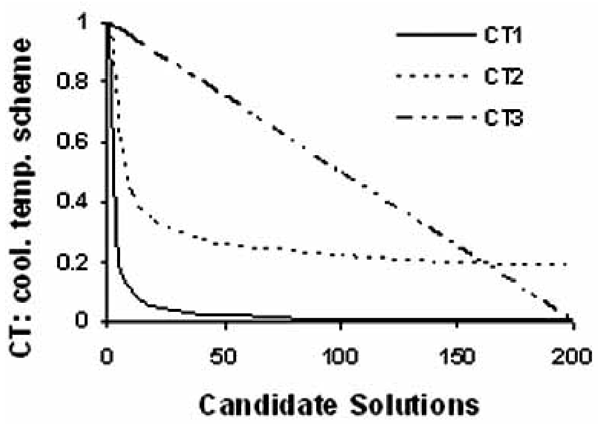

Fig. (5). Simulated annealing cooling temperature schemes (CT). (Table **2**).

GA and ES, two selections (Ranking and Tournament selections) and the use of the sharing property were studied. For GA a constant 10 % bit-flip mutation rate and 80% uniform crossover were considered, and for ES a 10% gene-mutation probability and 80% 1-point crossover. In some algorithms a sharing property was added to the generation evaluation step, such as to induce the simultaneous search of different optima [58]. The sharing property (Eq. 1) uses a similarity operator, *sim* (*a*,*b*), (Eq. 2), to measure the similarity between each of the generation candidate solutions. This similarity operator uses a constant α (in our case, α =1) and is proportional to the distance between each two individuals *d*(*a*,*b*). The individuals are penalized if they share a similar neighborhood by decreasing their performance value (*f*(a) → *f'*(a)) [58]. With this operator the regrouping of individuals is penalized and the population diversity increased, as shown below (Eq. 1 and 2).

$$f'(a) = \frac{f(a)}{\sum_{i=1}^{n} sim(a,b_i)} \qquad \text{Eq. 1}$$

$$sim(a,b) = 1 - \frac{d(a,b)}{\alpha} \qquad \text{Eq. 2}$$

where *f(a)* is the real fitness value of the candidate solution *a*, *f´(a)* the possible penalised fitness value, *d(a,b)* the distance between the candidate solutions *a* and *b* (0-1 normalised), and *n* the number of candidate solutions from the current generation.

For the SA and TS, and any methods that use the sharing property, the neighbourhood of the candidate solution needs to be defined, since it is necessary for these methods to order the distance between the candidate solutions which possess discrete variables. For this purpose the Jaccard-Tanimoto co-efficient has been used [59]. This co-efficient establishes how distant two candidate solutions are from each other by calculating, according to their bit-string representation, the ratio between the shared and non-shared bits.

As a blank reference to test the performance of the other algorithms, a random search was included in the comparison.

Algorithm Efficiency Criteria

The efficiency in achieving the Selox benchmark optimum catalytic system was tested for all the global algorithms described above. A stopping criterion of maximal 200 catalyst solutions to be tested was adopted for each algorithm search. Every algorithm was run 50 times and two efficiency criteria were computed from the average results: the percentage of the 50 runs that reached the optimal values (>0.92, >0.93 and 0.935 % Reliability), and the average maximum value reached after the performed 50 runs (Max. Fit. Average). The Opticat software was used for performing all the algorithmic optimisations [60]. The results obtained are presented in Table **2**.

Comparison of the Global Optimisation Algorithms

The algorithms used for optimising the Selox benchmark response surface and the results obtained are presented in Table **2**.

From Table **2** we can note that the maximum fitness average obtained, after 200 catalytic candidate solutions tested, was high for many of the algorithms. However, with some of the algorithms the maximum fitness average indicates that the algorithms were often trapped on sub-optimal solutions (SA, Taboo, GA_1, GA_2, Random). This is quite obvious with simulated annealing and taboo searches where, besides the low maximum fitness average, the percentage of runs reaching values above 0.92 (% Reliability) is even smaller than in the random search procedure. This demonstrates their risk of not achieving the right solution when just one search optimisation is performed. The poor performance of the SA search is likely to be related to having only a single candidate per iterative step (Table **2**). When starting from a bad candidate solution it can be more difficult to find the right track and the optimisation can be easily trapped into a local optimum solution. Another obvious practical disadvantage of a single-candidate solution methodology is its unsuitability to parallel high-throughput experimentation. Parallel Simulated Annealing, which does not suffer from the single-candidate limitation [46], is a methodology which was not within the scope of the present investigation but would be worthwhile to evaluate in future studies. For SA, an improvement can be observed when using a slower decrease of the cooling temperature schedule (CT, SA_1>SA_2>SA_3) indicating that a longer initial browsing period of the search space is required to improve the algorithm performance and to increase the probability of the algorithm finding the optimum track. Global optimisation algorithms, as the name indicates, are usually good in finding the global optimal solution area. They are, however, not the best for fine local optimisation of the continuous variables. This is indicated by the reduced percentage that reached the 0.93 and 0.935 % values - better results were achieved with SA and TS. These algorithms, which operate with the concepts of distance between the individuals and neighbourhood, are better able to exploit the optimal region, provided they are not trapped in a sub-optimal area. A common optimisation procedure to overcome this limitation is to employ local optimisation deterministic methods, like Simplex or Gradient Descent, after the global algorithm optimisation.

Evolutionary Strategies (ES) performed in general better than the Genetic Algorithms (GA), indicating that the repre-

Table 2. Optimisation Algorithms, Their Settings and Optimisation Results on the Selox Benchmark

Search Settings							% Reliability			Max. Fit. Average
Algorithm		**Distance**	$N_{Ind/Loop}$	N_{Loop}	**Specific Alg. Parameters**		**0.935**	**0.93**	**0.92**	**(0-0.935)**
Simulated Annealing	SA_1	yes	1	200	CT_1	$T_0=1$, $T_i = 1/N_{Loopi}$	4	6	30	0.837
	SA_2	yes	1	200	CT_2	$T_0=1$, $T_i = 1/\ln(N_{Loopi})$	12	24	45	0.838
	SA_3	yes	1	200	CT_3	$T_0=1$, $T_i = 0.005N_{Loopi}+1$	8	17	44	0.860
Taboo Search	TS	yes	5	40		---	5	12	33	0.821
Evolutionary Strategy	ES_1	no	40	5	Sel.	Ranking	5	24	87	0.919
	ES_2	no	40	5	Sel.	2 individuals-Tournament	6	35	90	0.920
Genetic Algorithm	GA_1	no	40	5	Sel.	Ranking	0	2	62	0.897
	GA_2	no	40	5	Sel.	2 individuals-Tournament	0	4	74	0.907
	GA_3	yes	40	5	Sel.	2 individuals-Tournament and Sharing	2	11	93	0.923
Hybrid Algorithms	GA-ZD	no	40	5	L.A.	Zone-Definer	0	11	89	0.912
	GA-LR	no	40	5	L.A.	Multi-linear Regression	0	10	93	0.922
Random	RD	no	40	5		---	0	4	58	0.871

SA: simulated annealing, TS: taboo search, GA: genetic algorithm, ES: evolutionary strategy, GA-LR: genetic algorithm hybridised with multi-linear regression learning algorithm, GA-ZD: genetic algorithm hybridised with zone-definer learning algorithm, CT: cooling temperature schedule, Sel.: Selection type, L.A.: learning algorithm.

sentation of the candidate solutions can play an important role. The higher variability usually obtained when using the bit-string representation seems not to be beneficial for the Selox benchmark optimisation. In respect to the selection types applied, the Tournament selection type produced better results than the Ranking selection for both the ES and GA cases. This is consistent with the results of previous investigations on the Selox benchmark regarding the effect of this setting for the GA optimisation [23]. The use of the Sharing property (Eq. 1, 2) with the GA appears to be highly beneficial, bringing about a considerable increase in performance and giving results even better than with the ES optimisation. The Sharing property, by monitoring the diversity of the candidate solutions from one iterative step to the other, disincentives premature convergence to a local optimum and enables several alternative optima to be pursued.

The hybridisation of learning algorithms (LR, ZD) with the GA improved the normal GA efficiency. The integration of an algorithm in which the existing knowledge from the previous experiences is used to choose the next generation of candidate solutions improves the GA search.

1.3. Design of Experiments (DoE)

Design of Experiments (DoE) aims at maximising the amount of information obtained from experimentation while minimising the number of experiments. DoE uses regression techniques to obtain the relationship between the response surface and the system factors in order to obtain a real model for the chemical system or to just understand the importance of the effects of the factors. When the factors that influence the system studied are continuous, quantitative interpolations are provided to minimize the number of experiments needed to obtain the model. However, when discrete factors are also present, which is the most common case in catalysis studies, other techniques have to be applied to avoid the exhaustive performance of all the discrete factor level combinations [61]. A common technique used is the D-Optimal criterion, where a sub-set of the total combination of experiments is selected that has an optimal distribution (the experiments are situated as far from each other as possible) [62].

DoE resembles to some extent conventional laboratory research, in the sense that, to improve efficiency, the search can be divided into different stages or designs. The parameter space is progressively reduced to the most relevant variables from one stage to the next and detailed information is obtained in the end for the most important factors [63]. The interpretation of the results, and the selection of the factors for further investigation, is the responsibility of the chemist. The parameter space is modified at each design, making it possible to eliminate or introduce new parameters or levels, according to the current understanding of the system under study. A representative DoE optimisation strategy, using the Nemrodw 2000 software, is discussed below.

1.4. DoE Strategy for the Selox Benchmark Optimisation

Screening for the Effects of the Main Factors

The cost of the total amount of information obtained is the number of experiments it is necessary to perform. To estimate simple first-order or main effects, a small number of the total experimental effort is usually necessary, while in order to account for interaction effects as well, more datapoints are demanded. For the parameter space presented in Table **1**, the minimum number of experiments recommended to obtain the trends for the main effects of the factors is 18. If we would like to study all the two-factor interaction effects, a minimum of 130 experiments would be required. Frequently the option adopted will need to take into account the research time and resources available. A common optimisation procedure comprises an initial study of the main effects of the factors, following which a subspace of the initial parameter space is chosen for further investigation of the interaction effects between the factors. In the present case, ten different D-Optimal designs were evaluated to observe the variance of the estimated effects of the factors with the experimental points chosen. The position of the experimental points in the Selox response surface of one of the D-optimal designs is represented in Fig. **6**, and, the results obtained for the main factor effects study are depicted in Fig. **7**.

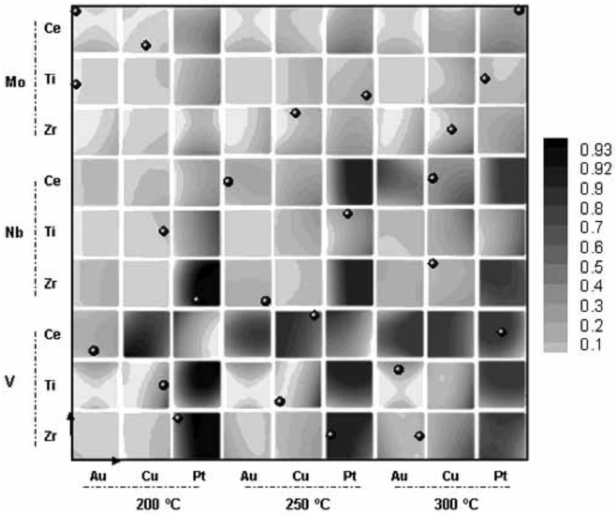

Fig. (6). Representation of one of the possible designs in the Selox search space.

The estimate of the main effect of a factor is related to the average response values of the experiments performed to its factor levels. An important factor causes a large effect because the system will perform significantly better or worse at one of its levels on average.

The results obtained for the main factor analysis in the Selox response and its variance by choosing difference sets of experimental designs is represented in Fig. **7**. From this analysis we see that the main trends in the Selox surface are reproduced. The NM type has by far the largest effect on the Selox response; if we look at the Selox response surface we see that, on average, when Pt is used the performance increases significantly, while Au is the poorest performer of the three NM, followed by Cu. The use of Mo as TM type also affects significantly the performance, but in a negative way, and this can be confirmed by noticing that on average it matches with the lightest areas from the Selox surface (Fig. **6**). The effects of the other factors are smaller and it is not possible to rely on their significance at this level. This means that those trends are not sufficiently clear to predict the effect of choosing one of their factor levels. The principal outcome of the effect study pinpoints Pt as a positive determi-

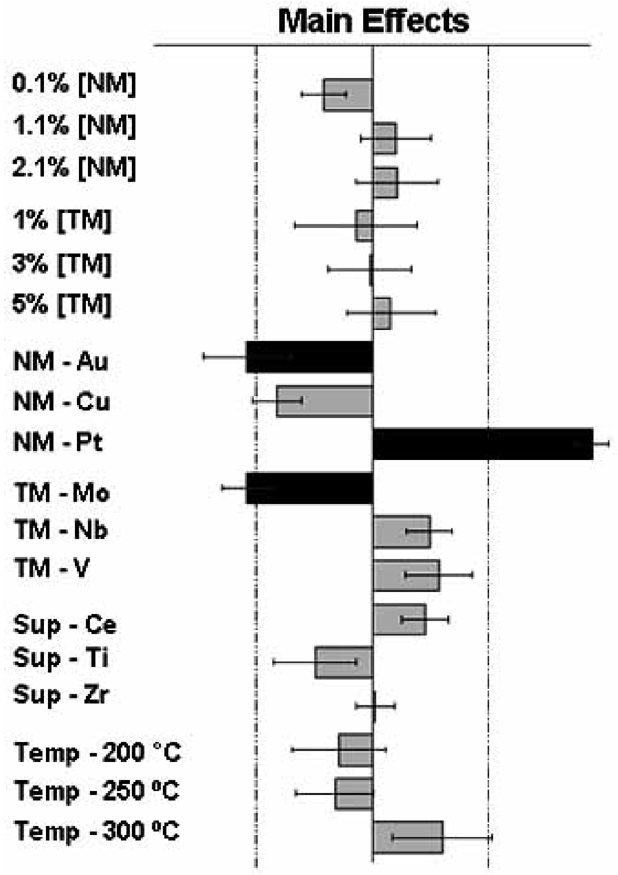

Fig. (7). Average main factor effects with 95% confidence interval and their variance. The bars directed to the right signify a positive relative effect and those to the left a negative one. The dotted lines represent the 95% confidence interval calculated from the estimated experimental variance. Effects higher than this confidence level are considered statistically significant and are represented in black.

nant factor level and Mo as a negative one. Based on these results, Pt could be selected and Mo eliminated in a subsequent more detailed parameter study.

Screening for Interaction Effects

In this second design the interaction effects of the discrete variables Support, TM and Temp were studied (Table **3**). By opting to perform fine-tuning optimisation of the [NM]% and [TM]% continuous variables at a later stage and by setting their variables at their median-values, a substantial reduction of the parameter space was achieved. A full factorial design (Fig. **8**) could then be chosen comprising 18 experiments.

Table 3. Factors to be Investigated in the Second Screening and Their Correspondent Levels

Factor Name	Level 1	Level 2	Level 3
1. TM	V	Nb	-
2. Support	CeO2	ZrO2	TiO2
3. Temp ºC	200	250	300

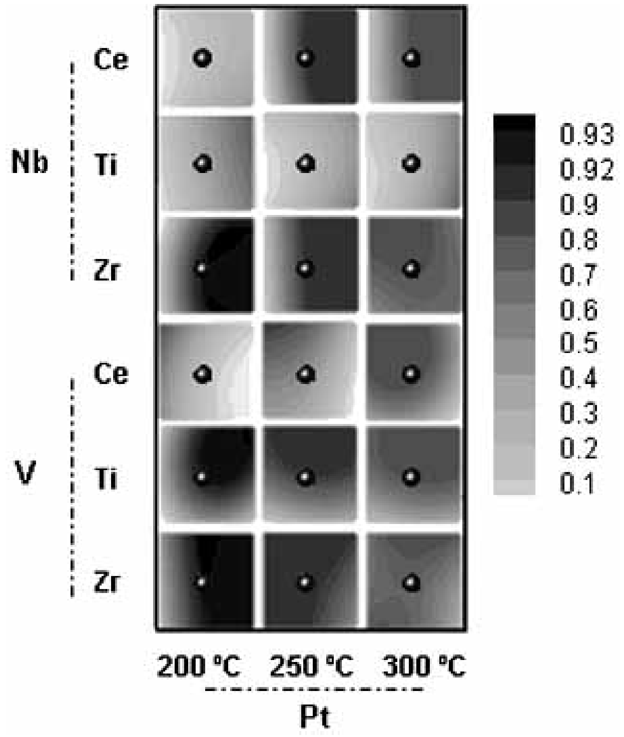

Fig. (8). Representation of the reduced search space after the main factors analysis, and the new experimental design points.

The results obtained for the two-factor main and interaction effects are depicted in Fig. **9**. A detailed analysis of the main effects for this reduced search space show us that on average the TM type V is better than Nb. Also ZrO_2 is on average a better support. In the interaction analysis we notice that even if TiO_2 is not, on average, considered the better support, there is a strong positive interaction between the TM V and the support TiO_2 (V-Ti). Since the response surface presents several optimal areas, several trends could be followed to search them. The interaction effect between the transition metals and the reaction temperature does not show significant trends. However the interaction of ZrO_2, or of TiO_2, with this parameter indicates a positive interaction at 200 ºC. The most significant interactions are however the TM-Support interactions. The interaction Ti-V is very strong, and there is a difference between the Nb-Zr and V-Zr interactions, where the Nb-Zr interaction has a positive effect. These trends can be confirmed by comparing them with the response surface in Fig. **8**. At this stage of the optimisation we have the indication that Pt is the best performing NM type and that the optimal region could lie in the combination of this metal with V supported on TiO_2 or Nb supported on ZrO_2 at 200 ºC reaction temperature.

Fine-Tuned Optimisation

The final optimisation step is exemplified with the selected Nb-Zr interaction at 200 ºC; in this final step the exact location of the optimum in terms of the concentration of TM and NM is to be found. This usually requires a mathematical model to be built which can be used to predict the response surface of the problem being investigated. More detailed information about each of the factors is then necessary; therefore these modeling experiments are much larger than

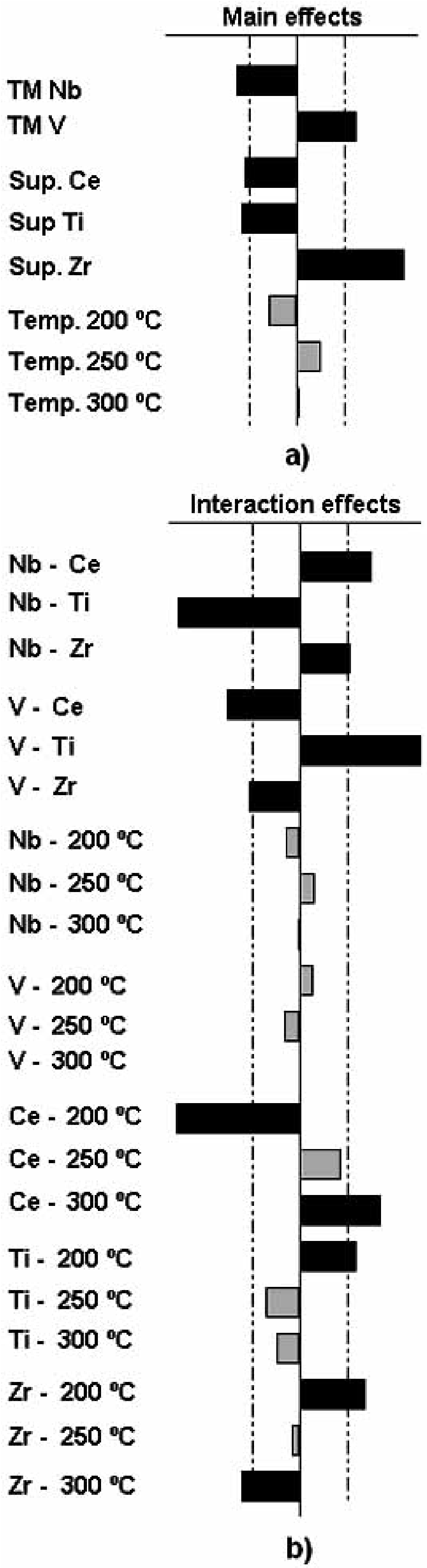

Fig. (9). Statistical effect analysis for the second screening. **a**) Main effects. **b**) Interaction effects.

screening experiments and are only performed for a few factors. To model the curvature of the response surface a second-order model is usually considered to be sufficient. A Central Composite design [3] (Fig. **10a**), was chosen to plan the experimental points to be performed. For the efficient modelling of the response surface 7 experiments plus two replicates of the central point were selected. The resulting response surface obtained and the location of the optimal composition region is shown in Fig. **10b**.

In this response surface (Fig. **10b**), the representative curvature of the benchmark surface (Fig. **10c**) is obtained. The optimal area is allocated, but with this second order model the small areas where the final optimum (>0.93) resides are not revealed. To obtain more detailed information, a higher order model design would be necessary.

CONCLUSIONS

The comparison of the different algorithms applied to the same benchmark has proven valuable for understanding the way they work and how they may be adjusted to the chemical problem under study. The efficiency of any optimisation algorithm depends largely on the difficulty of the problem to be studied. Adjustment of the browsing/exploitation ratio of the algorithm is essential to avoid becoming rapidly trapped on a local optimum or performing too lengthy an optimisation. Since chemists do not know *a priori* the shape of the response surface to be studied, they need to base their choice of the algorithm parameters in the beginning of the optimisation procedure on existent benchmark studies.

If the parameter space to be investigated is large, the Design of Experiments does not usually constitute the most efficient strategy. The research can be divided in different stages, in which the scientist can adjust the size of the parameter space and the detail of information withdrawal at each stage, making the research more flexible and efficient. This strategy runs, however, the risk that synergetic variables are discarded at an early stage because their interactions were not detected. In this way, even if better cause-effect knowledge about the effects of the variables in the chemical system is obtained (which is necessary to understand the chemical system), the achievement of the optimal solution cannot be guaranteed. With global optimisation algorithms, the final optimum is more efficiently achieved and guaranteed if the number of necessary experiments is performed, but only a poor understanding of the chemical system and the effects of the variables on the response is obtained. The main goal with Global optimisation is the achievement of the optimal solution. For small search spaces many of the global optimisation methods are not adequate. For instance, for the normal procedure of a GA, a minimum population size of 20 candidate solutions and 3 iterative generations are usually necessary. A more efficient design could be obtained with DoE where such restrictions do not exist.

Chemometric methods for catalysis composition optimisation are excellent aids for exploring and allocating the optimal parameter space regions avoiding an inefficient exhaustive experimental investigation. The appropriate method to be used, however, is dependent on the research resources available, the nature of the parameter space and the precise research aims.

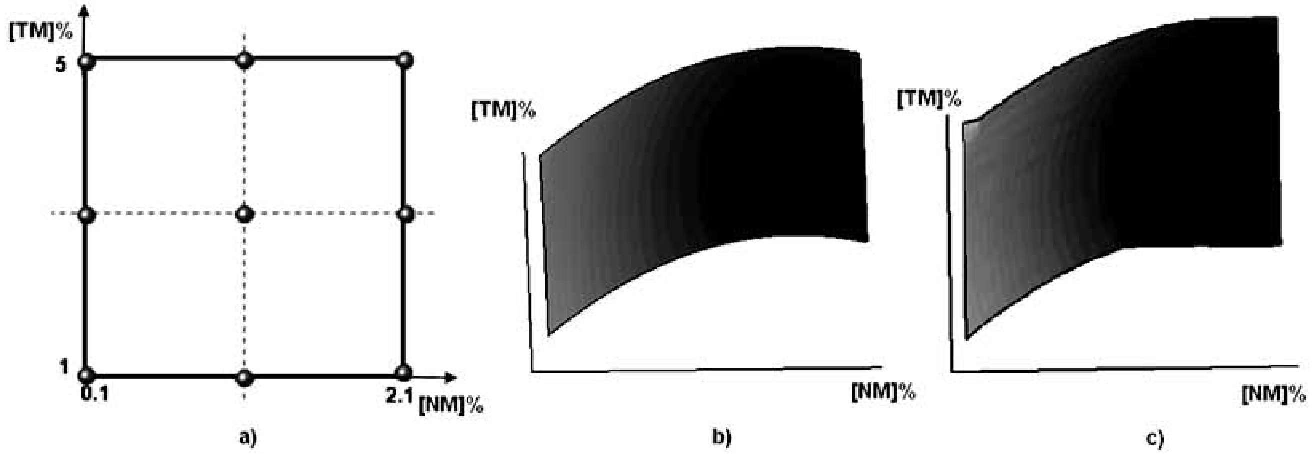

Fig. (10). Composition optimisation in terms of the [NM]% and [TM]% of the catalyst composed of Pt-Nb-Zr at 200 ºC. **a**) Central Composite representation of the experimental plan; **b**) resulting response surface with central composite; **c**) Selox benchmark response surface for the Pt-Nb-Zr catalysts composition.

ACKNOWLEDGEMENTS

The authors thank R. S. Downing for discussions and revision of the manuscript, the financial support of NWO/ ACTS and TopCombi FP6 programs. The authors also thank Prof. M. Sergent for discussions on DoE.

REFERENCES

[1] Maxwell, I.E.; van den Brick, P.; Downing, R.S.; Sijpkes, A.H.; Gomez, S.; Maschmeyer, T. *Top. Catal*., **2003**, *24*, 125.
[2] Pescarmona, P.P.; van de Waal, J.C. *Catal. Lett.*, **1999**, *63*, 1.
[3] Massart, D.L.; Vandeginste, B.G.M.; Buydens, L.M.C.; De Jong, S.; Lewi, P.J.; Smeyers-Verbeke, J. *Handbook of Chemometrics and Qualimetrics: Part A and B*, Elsevier, Amsterdam, **1997**.
[4] Brereton, R.G. *Chemometrics: Data analysis for the laboratory and chemical plants*, John Wiley & Sons, Bristol, UK, **2003**.
[5] Holzwarth, A.; Denton, P.; Zanthoff, H.; Mirodatos, C. *Catal. Today*, **2001**, *67*, 309.
[6] Dawson, E.A.; Barnes, P.A. *Appl. Catal. A*, **1992**, *90*, 217.
[7] Weckhuysen, B.M.; Verberckmoes, A.A.; Debaere, J.; Ooms, K.; Langhans, I.; Schoonheydt, R.A. *J. Mol. Catal. A*, **2000**, *151*, 115.
[8] Tagliabue, M.; Carluccio, L.C.; Ghisletti, D.; Perego, C. *Catal. Today*, **2003**, *81*, 405.
[9] Bricker, M.L.; Sachtler, J.W.A.; Gillespie, R.D.; McGonegal, C.P.; Vega, H.; Bem, D.; Holmgren, J.S. *Appl. Surf. Sci.*, **2004**, *223*, 109.
[10] Barsan, M.M.; Thyrion, F.C. *Catal. Today*, **2003**, *81*, 159.
[11] Hunter, W.G.; Mezaki, R. *Can. J. Chem. Eng*., **1967**, *45*, 247.
[12] Nordin, A.; Eriksson, L.; Ohman, M. *Fuel*, **1995**, *74*, 128.
[13] Sjöblom, J.; Papadakis, K.; Creaser, D.; Odenbrand, C.U.I. *Catal. Today*, **2005**, *100*, 243.
[14] Carlson, R. *Data Handling in Science and Technology*, Vol. *8*: Design and optimisation in organic synthesis, Elsevier, Umeå, Sweden, **1992**.
[15] McWilliams, J.C.; Sidler, D.R.; Sun Y.; Mathre; D.J. *Applying statistical design of experiments and automation to the rapid optimisation of metal-catalysed processes in process development*, Merck, JALA, **2005**, *10*, 394.
[16] Carlson, R.; *Chemom. Intel. Lab. Systems*, **2004**, *73*, 151.
[17] Stazi, F.; Palmisano, G.; Turconi, M.; Santagostino, M. *Tetrahedron. Lett.*, **2005**, *46*, 1815.
[18] Veum, L.; Pereira, S.R.M.; van der Waal, J.C.; Hanefeld, U. *Eur. J. Org. Chem*., **2006**, *7*, 1664.
[19] Leach, A.R.; Gillet, V.J. *An Introduction to Chemoinformatics*, Kluwer Academic Publishers, London, **2003**.
[20] Klanner, C.; Farrusseng, D.; Baumes, L.; Mirodatos, C.; Schüth, F. *QSAR Comb. Sci*., **2003**, *22*, 729.
[21] Sieg, S.; Stutz, B.; Schmidt, T.; Hamprecht, F.; Maier, W.F. *J. Molec. Model.*, **2006**, *12*, 611.
[22] Rodermerck, U.; Baerns, M.; Holena, M.; Wolf, D. *Appl. Surf. Sci*., **2004**, *223*, 168.
[23] Pereira, S.R.M.; Clerc, F.; Farrusseng, D.; van der Waal, J.C.; Maschmeyer, T.; Mirodatos, C. *QSAR Comb. Sci*., **2005**, *24*, 45.
[24] Kordabadi, H.; Jahanmiri, A. *Chem. Eng. J*., **2005**, *108*, 249.
[25] Rodermerk, U.; Wolf, D.; Buyeskaya, O.V.; Claus, P.; Senkan, S.; Baerns, M. *Chem. Eng. J*., **2001**, *82*, 3.
[26] Grubert, G.; Kolf, S.; Baerns, M.; Vauthey, I.; Farrusseng, D.; Veen, A.C.; Mirodatos, C.; Stobbe, E.R.;. Cobden, P.D. *Appl. Catal. A-Gen*., **2006**, *306*, 17.
[27] Paul, J.S.; Janssens, R.; Denayer, J.F.M.; Baron, G.V.; Jacobs, P. J. *Comb. Chem*., **2005**, *7*, 407.
[28] Wolf, D.; Buyevskaya, O.V.; Baerns, M. *Appl. Catal. A-Gen*, **2000**, *200*, 63.
[29] Wolf, D.; Baerns, M. *Experimental Design for Combinatorial and High Throughput Development*, John Wiley & Sons, Hoboken, New Jersey, **2003**, p. 147.
[30] Serra, J.M.; Corma, A.; Farrusseng, D.; Baumes, L.; Mirodatos, C.; Flego, C.; Perego, C. *Catal. Today*, **2003**, *81*, 425.
[31] Yamada, Y.; Ueda, A.; Nakagawa, K.; Kobayashi, T. *Res. Chem. Intermed*., 28, *5*, **2002**, 397–407.
[32] Watanabe, Y.; Umegaki, T.; Hashimoto, M.; Omata, K.; Yamada, M. *Catal. Today*, **2004**, *89*, 455.
[33] McLeod, S.; Gladden, L.F. *J. Chem. Inf. Comput. Sci*., **2000**, *40*, 981
[34] Maier, R.W.; Whiting, W.B. *Comp. Chem. Eng*., **1998**, *23*, 47
[35] Yao, R.; Yang, B.; Cheng, G.; Tao, X.; Meng, F. *Chem. Eng. J*., **2003**, *94*, 113
[36] Lin, B.; Miller, D.C. *Comp. Chem. Eng*., **2004**, *28*, 2287.
[37] Baumes, L. *J. Comb. Chem*, **2006**, *8*, 304.
[38] Scheidtmann, J.; Klaer, D.; Saalfrank, J.W.; Schmidt, T.; Maier, W. F. *QSAR Comb. Sci*., **2005**, *24*, 203.
[39] Tompos, A.; Margitfalvi, J.; Tfirst, E.; Végvári, L. *Appl. Catal.* A, **2006**, *303*, 72.
[40] Végvári, L.; Tompos, A.; Gobolos, S.; Margitfalvi, J. *Catal. Today*, **2003**, *81*, 517.
[41] Lucasius, C.B.; Katemen, G. *Chem. Intel. Lab. Systems*, **1993**, *19*, 1.
[42] Lucasius, C.B.; Kateman, G. *Chemom. Intel. Lab. Systems*, **1994**, *25*, 99.
[43] Hibbert, D.B. *Chemom. Intel. Lab. Systems*, **1993**, *19*, 277.
[44] Kirkpatrick, S.; Gelatt, C.D.; Vecchi, M.P. *Science*, **1983**, *220*, 671.
[45] Cerny, V. *J. Opt. Theory Appl*., **1985**, *45*, 51.
[46] Sanvicente-Sánchez, H.; Frausto-Solís, J., MPSA: *A methodology to parallelize simulated annealing and its application to the traveling salesman problem*, http://citeseer.ist.psu.edu.
[47] Glover, F.; Laguna, M. *Tabu Search*, Kluwer Academic Publishers, Boston, **1997**.
[48] Glover, F.; Laguna, M. *Modern Heuristic Techniques for Combinatorial Problems*, Blackwell Scientific Publishing, Oxford, England, **1993**.

[49] Baumes, L., PhD dissertation, Lyon, **2005**.
[50] Clerc, F.; Lengliz, M.; Farrusseng, D.; Mirodatos, C.; Pereira, S.R.M.; Rakotomalala, R. *Rev. Sci. Inst*., **2005**, *76(6)*, 1-13.
[51] Clerc, F.; Rakotomalala, R.; Farrusseng, D. Learning fitness function in a combinatorial optimization process, in proceedings of ASMDA conference: Applied Stochastic Models and Data Analysis, Brest, may **2005**.
[52] Corma, A.; Serra, J.M.; Serna, P.; Valero, S.; Argente, E.; Botti, V. *J. Catal.*, **2005**, *229*, 513.
[53] Huang, K.; Zhan, X.L.; Chen, F.Q.; Lü, D.W. *Chem. Eng. Sci*., **2003**, *58*, 81.
[54] Umegaki, T.; Watanabe, Y.; Nukui, N.; Omata, K.; Yamada, M. *Energ. Fuels*, **2003**, *17*, 850.
[55] Bentley, J.L. *Comm. ACM*, **1975**, *18*, 509
[56] Bentley, J.L.; Friedman, J.H. *ACM Comput. Surv*. **1979**, *11*, 397.
[57] Parzen, E. *Ann. Math. Sta.*, **1962**, *33*, 1065.
[58] Goldberg, D.E.; Richardson, J., Genetic algorithms with sharing for multimodal function optimization, in *Second International Conference on Genetic Algorithms and Their Application*, Lawrence Erlbaum Associates Inc, **1987**, 41.
[59] Tan, P.-N.; Steinbach, M.; Kumar, V. *Introduction to Data Mining*, Pearson/Addison Wesley, Boston, **2005**.
[60] Clerc, F., *OptiCat - A Combinatorial Optimisation Software*. Free of charge download from http://eric.univ-lyon2.fr/~fclerc/
[61] Du, H.; Lindsey, J. *Chemom. Intel. Lab. Systems*, **2002**, *62*, 159
[62] Johnson, M.E.; Nachtsheim, C.J. *Technometrics*, **1983**, *25*, 271.
[63] Haaland, P. *Statistical Design and Analysis of Industrial Experiments*, Chapter 3, Marcel Dekker, New York, **1990**.

Received: August 22, 2006 **Revised: December 12, 2006** **Accepted: December 15, 2006**