Chemometric Optimization Studies in Catalysis Employing High-Throughput Experimentation

Sílvia Raquel Morais Pereira

Cover: drawing by Carlos Ribeiro

Chemometric Optimization Studies in Catalysis Employing High-Throughput Experimentation

Proefschrift

ter verkrijging van de graad van doctor
Aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. dr. ir. J. T. Fokkema,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen

op maandag 3 november 2008 om 10:00 uur.

door

Sílvia Raquel MORAIS PEREIRA

Engenheira em Química, Instituto Superior Técnico, Portugal
geboren te Bragança, Portugal

Dit proefschrift is goedgekeurd door de Promotoren:

Prof. dr. R. A. Sheldon

Prof. dr. G. J. Witkamp

Samenstelling promotiecomissie:

Rector Magnificus	voorzitter
Prof. dr. R. A. Sheldon	Technische Universiteit Delft, promotor
Prof. dr. G. J. Witkamp	Technische Universiteit Delft, promotor
Prof. dr. ir. H. van Bekkum	Technische Universiteit Delft
Prof. dr. F. Kapteijn	Technische Universiteit Delft
Prof. dr. G. Rothenberg	Universiteit van Amsterdam
Dr. D. Farrusseng	CNRS, Lyon, France
Dr. ir. J. C. van der Waal	Avantium Technologies B.V.

Dr. Ir. J. C. Van der Waal heeft als begeleider in belangrijke mate aan de totstandkoming van het proefschrift bijgedragen.

To my grandparents

Para os meus avós

Contents

Scope of the thesis

Catalysis is of fundamental importance to industry and the economy. It is estimated that around 20-40% of GDP is generated through processes involving catalysis. In the petroleum and bulk-chemical industries, catalytic processes are almost universal while, in fine and pharmaceutical chemicals, catalytic reactions play an increasingly important role. In these industries the acceleration of the catalyst development and research can have a large economic and ecological effect, because more than 90 % of current industrial processes involve catalytic steps. Catalysis research and development is therefore an activity of major importance for the industry, and also in universities and research organisations. A great effort is therefore put nowadays in the research and discovery of new catalyst processes and the optimization of existent process parameters.

In catalysis the challenge of research is to find catalysts with high activity and selectivity for certain chemical transformations, and to optimize them. Most improvements in catalyst design and process conditions optimization arise from extensive empirical catalyst synthesis and mass screening. A way to accelerate this optimization is to use High-Throughput Experimentation (HTE) methodologies. In recent years, HTE has been recognised as a set of tools of great value in improving the productivity of research into new and improved catalysts and processes. The use of miniaturised and parallel reactor assemblies decreases the limitations on the number of experiments that can be performed, maximises diversity, and enables the various stages from discovery to process development to be integrated effectively.

The increase in experimental capabilities allowed with HTE brings with it a requirement for effective experimental design, if the potential of high information yield is to be realised and data explosions avoided. To use HTE in an efficient and rational way the chemist needs to a certain extent to change his practical research habits. Multiple parallel reactions need to be planned, to be performed in grouped reaction batches and the resulting large volumes of analytical data processed. By the application of appropriate chemometric techniques, appropriate designs for all stages of catalysis research, from catalyst preparation, through reactor testing to process optimization, can be constructed. Chemometric methods can also provide tools for efficient data analysis of large sets of HTE results such as analysis of variance (ANOVA), principal components analysis (PCA) and principal least squares (PLS). In this thesis alternative strategies are discussed and applied to create suitable optimization procedures for examples of both homogeneous and heterogeneous catalytic reactions.

The use of several chemometric design planning methodologies in HTE catalysis research is experimented in this thesis through different catalytic case studies. Three different catalytic systems have been considered: the investigation of homogeneous Brønsted-acid catalytic hydration of terpenes; the optimization of heterogeneously catalysed reaction conditions for the reduction of cyanohydrin esters to N-acylated-ß-amino alcohols; and finally the optimization of mixed-oxide heterogeneous catalyst compositions for the oxidation of CO, with and without hydrogen, for fuel-cells applications. To tackle these three diverse catalytic subjects different chemometric optimization strategies have been applied. These strategies can be divided into two main classes: the Design of Experiments (DoE) and the Global Optimization algorithms.

The main topic of this thesis is the investigation of the synergies between High-Throughput Experimentation (HTE) and Chemometric Optimization methodologies in Catalysis research. The uses of these strategies for catalysis research are discussed in this thesis through its various chapters:

Chapter 1 provides an introduction to High-Throughput Experimentation (HTE), Chemometric optimization methods and considerations about their combined application for catalysis research.

Chapter 2 presents an overview of the Design of Experiments techniques, since this methodology is central to most of the thesis research.

Chapter 3 deals with the exploration of a terpene hydration reaction parameter space using a Design of Experiments approach.

Chapter 4 presents the optimization of the reaction conditions for the reduction of cyanohydrin esters to N-acylated-ß-amino alcohols using a multi-step Design of Experiments approach.

Chapter 5 deals with the planning and modelling of catalytic benchmark response surfaces using Design of Experiments. A library of 189 mixed-oxide catalysts was prepared and their activity tested and modelled for the CO oxidation reaction in both the absence (COOX) and presence (SELOX) of hydrogen.

Chapter 6 is concerned with the optimization procedure of employing the Global Optimization Genetic algorithm and the influence of the algorithm settings on its optimization efficiency. Both COOX and SELOX benchmarks are used to validate the Genetic algorithm performance.

Chapter 7 compares the performance of several algorithms in their optimization based on the SELOX benchmark. Evolutionary strategies, Genetic algorithms, Simulated Annealing, Taboo Search, and Hybrid Genetic algorithms are the Global Optimization strategies tested. Considerations between the uses of Design of Experiments or Global Optimization algorithms for catalysis research are also discussed.

Appendix A describes the High Throughput Experimentation equipment used.

Appendix B presents the mathematical models (obtained in Chapter 5) that constitute the COOX and SELOX benchmarks.

Appendix C presents a brief description of the Opticat software used in Chapters 6 and Chapter 7.

1

OPTIMIZATION STUDIES IN CATALYSIS EMPLOYING HIGH-THROUGHPUT EXPERIMENTATION: DESIGN OF EXPERIMENTS AND GLOBAL OPTIMIZATION ALGORITHMS

Abstract

High-Throughput experimentation (HTE) is being increasingly used in the field of catalysis research. In order to realise the benefit of the higher throughput obtainable using these methods, simultaneous parallel reactions need to be planned and the large amount of data generated needs to be analysed. Chemometric tools, originally designed to tackle analytical chemistry issues, are proving to be a valuable aid for both experimental planning and data analysis. In combination with HTE, they enable diversity to be maximised and the productivity and efficiency of the research to be greatly improved. This combined application offers therefore new possibilities and a step forward in catalysis research.

1.1 - Introduction

Catalysis research allowed the development of more efficient and greener chemical processes through the acceleration of chemical reactions and the increase in selectivity towards the desired product. A great effort is put in the research and discovery of new catalytic processes and the optimization of existent process parameters. A way to accelerate this discovery is to use High-Throughput Experimentation (HTE) methodologies. The use of HTE also allows more diversity to be taken into consideration in the catalysis search parameter space due to the lower experimental effort, but the possibility of screening large parameter spaces can also lead to an experimental combinatorial explosion, impossible to handle even with HTE. Multiple parallel reactions need to be planned, to be performed in reaction batches and the resulting large analytical data volumes processed. Efficient experimental planning and data analysis can be partially solved by chemometric methods. These methods allow to further enhance the capabilities of HTE experimentation and their application in HTE catalysis is nowadays increasingly studied.

1.2 - High-Throughput catalysis

In contrast to traditional methods, in which product formulations or sets of process conditions are tested sequentially, High-Throughput Experimentation (HTE) is a method that allows the synthesis and screening of tens or hundreds of reactions or catalysts simultaneously. Robotics in the form of liquid and solid handling devices and, usually, miniaturised multiple reactor systems are used to speed-up the screening procedures of research and development in several chemistry fields. The multiplicity and parallelism of the reactor arrays enable productivity and diversity to be maximised, by reducing limitations in the number of experiments that can be conducted. HTE research, besides the saving in time, also allows safer, and ultimately cheaper research, and has a lower environmental impact since only small quantities of reactants are used [1]. The initial applications of HTE occurred in the field of combinatorial drug discovery in the late ‘eighties. Since then, HTE use in this field is widespread, being nowadays a standard method for drug discovery in industry [2,3]. The use of HTE methods in the catalysis field is more recent, the first reports appearing about fifteen years ago. Schültz *et al.* in 1995 disclosed the application of HTE methods for the discovery of novel materials and library screening of properties such as superconductivity and magneto resistance [4,5]. Later in the same year, libraries of compounds for use as catalysts for asymmetric synthesis [6] and phosphatase hydrolysis [7] were also reported. Nowadays a significant amount of work is still being performed, aimed at realising the full potentialities of the method in this area. Some good review papers about these efforts are presented in [8-10]. HTE can be used in most procedures involved in the development and testing of catalysts. Catalyst synthesis (both homogeneous and heterogeneous), screening for activity and selectivity, and optimization of process parameters can all be performed with higher efficiency using HTE methods.

The advantages of HTE have sometimes prompted the conclusion that less scientific insight is necessary when applying these methods. This is, however, a wrong perception since the challenge to the chemist is in fact increased rather than decreased. Due to the removal of limitations on the experimental effort that can be performed, the chemist can take into consideration more variables that can influence the system under study. This increased experimental power compels the chemist to use his/her chemical knowledge to determine the relevant variables and their settings in order to carry out a meaningful experimental plan. And, finally, a great deal of critical chemical insight is again essential for interpretation of the results, knowledge extraction and the planning of subsequent research steps. High-Throughput Experimentation is not a new field of science but a series of tools developed for helping chemists. Optimization studies are without doubt one of the areas where HTE methodology presents great advantages, since it allows larger parameter spaces to be investigated in a systematic and experimentally efficient manner. In Appendix-A the High-Throughput equipment employed in this thesis is presented and briefly described.

1.3 - Chemometrics for HTE catalysis

Chemometrics is defined by the International Chemometrics Society (ICS) as “the science of relating measurements made on a chemical system or process to the state of the system via application of mathematical or statistical methods”. These methods, initially dedicated to tackle analytical chemistry issues, are nowadays receiving special interest from the catalyst HTE community and a considerable number of studies have been performed in order to improve the understanding of these methods and to adjust them to the field of catalysis optimization [11-13].

The wider acceptance of High-Throughput Experimentation (HTE) and combinatorial methods has opened, in recent years, a broad range of new possibilities to the catalyst researcher [14]. The automation and parallelisation of the experimentation poses, however, new challenges to the chemist in the planning of the experimental work so as to take full advantage of the HTE capabilities. HTE introduces the need of batch experimental planning and the analysis of large volumes of data. The research routine is based on batches of parallel experiments, as opposed to the conventional approach of sequential experiments. The way to plan the experimentation is therefore different since all experiments need to be designed beforehand. Taking advantage of the lower limitations on experimentation, robust experimental designs can be performed and larger parameter spaces searched. As a consequence of the large batches of experiments performed, large data sets of results are also obtained. Chemometrics provides the tools for efficient experimental planning, data treatment and knowledge extraction. Many stages of the HTE catalysis research can therefore benefit from the aid provided by chemometric methods. The experimental planning of the research workflow can be efficiently performed via Design of Experiments or Global Optimization techniques. Chemometrics data analysis techniques allow to efficiently extract information from the large data sets obtained in HTE research. Analysis of variance (ANOVA) is a statistical set of tools that allow verifying the significance of the information obtained. Methods such as principal component analysis (PCA) allow for outliers and data correlations identification can be used for facilitating the interpretation of the results. When a statistically planned experimental design is performed, regression techniques such as ordinary linear regression (OLR) and principal least squares (PLS) can also be used to obtain the system model.

Correlations between variable descriptors (such as solvents, additives) and activity/performance can be obtained by quantitative structure activity/properties relationships (QSAR, QSPR) methods [11]. These involve the construction of some form of model which enables the observed activity or properties to be related to the molecular structure descriptors; these methods are extensively used in the virtual screening of drug candidates. For catalysis-related studies, mainly heterogeneous catalysis, the importance of study of the composition has led to the development of quantitative composition activity relationship (QCAR) techniques [10]. Via QSPR/QSAR/QCAR models large data sets of libraries can be virtually screened and the combinatorial explosion due to the presence of many categorical variables reduced. Through these models knowledge about catalysis is obtained by extraction of rules and relationships, and this knowledge can be further used in the design of new libraries [15]. These methods belong to the Cheminformatics [16,17] area of research and will not be treated in this thesis in which ordinary linear regression techniques are applied and focus is put on experimental planning techniques.

1.4 - Chemometric experimental planning techniques

The increased throughput that is achieved with HTE technologies has introduced the need for a different strategy in experimental planning and also for new possibilities for robust experimentation. Several experimental planning methodologies exist that can handle different experimentation throughputs and demands. Many of these technologies are also used in other scientific fields that deal with multivariate issues such as economics, social sciences, etc. Their application in chemistry-related issues is relatively recent and in many cases just nowadays being explored (see Chapters 6 and 7).

Two main groups of strategies can be designated for experimental planning: methods based on statistical planning and modelling of the entire parameter space (Design of Experiments); and, methods based on iterative searches in the parameter space (Optimization Algorithms).

Design of Experiments (DoE) is the methodology most commonly used in HTE catalysis. For some catalytic issues, like for instance some cases of heterogeneous catalyst composition optimization, algorithm optimization strategies are nowadays being increasingly studied and applied. These two mathematical aids to experimentation are different in nature and objective, the choice of which to use may depend on the catalytic problem to be solved.

1.4.1 - Design of Experiments (DoE)

Design of Experiments (DoE) uses regression techniques to obtain the relationship between the responses and the system factors. Models for the chemical systems can be obtained via this technique, or just screening information about the importance of the effects of the factors studied. DoE can resemble 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 [18]. 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.

DoE is starting to be increasingly applied in research in the catalysis field. Literature is available reporting screening and optimization studies of catalytic systems. Some of these studies include catalyst formulation and preparation [19-22], catalytic kinetic modelling [23,24], reactor engineering [25,26], and the optimization of catalytic reaction conditions [27-31]. In chapter 2 this methodology is explained in more detail.

1.4.2 - Optimization algorithms

Optimization algorithms are iterative computational routines that guide the experimental procedure (Chapter 6 and 7). There are several algorithm procedures that can be applied, depending on the characteristics of the system requiring optimization. The algorithms can have a heuristic (basing the search procedure on premises) or stochastic (random) character, and can be divided into local or global optimization algorithms. While local optimization algorithms are valuable for fine-tuning optimization, they stop on the first optimum found. Global optimization algorithms in contrast are effective in finding the global optimal solution location but weaker in the fine optimization. Examples of local search algorithms are Simplex and Steepest Ascent; global search algorithms include the Genetic Algorithms (GA) and Simulated Annealing (SA) between others. This thesis will focus on the Global Optimization search algorithms. These are often inspired by natural processes of optimization, like Genetic Algorithms (GA) which simulate biological evolution and survival of the fittest, and Simulated Annealing (SA) which mimics the gradual cooling of a metal to achieve its most stable solid form (Chapters 6 and 7). Optimization algorithms represent alternative optimization methods to DoE, being many times preferred when there is a) a large parameter space, b) non-linear shape of the response surface (where synergistic effects are commonly encountered) and c) the existence of many categorical variables [32,33].

1.5 - Summary

High-Throughput Experimentation, together with chemometric methods of experimental planning and data analysis, allows increased possibilities for catalysis research. The synergy between these two strategies is, however, still nowadays not fully studied and explored. This investigation and mainly the experimental planning techniques applied to HTE catalysis is a focal point from this thesis.

1.6 - References

[1] P. P. Pescarmona, An Exploration of Silsesquioxanes and Zeolites using High-Speed Experimentation, dissertation, Delft, **2003**.

[2] M. A. Gallop, R. W. Barret, W. J. Dower, S. P. A. Fodor and E. M. Gordon, *J. Med. Chem.*, **1994**, 37, 1233.

[3] E. M. Gordon, R. W. Barret, W. J. Dower, S. P. A. Fodor and M. A. Gallop, *J. Med. Chem.,* **1994**, 37, 1385.

[4] X.-D. Xiang, X. Sun, G Briceno, Y. Lou, K-A Lang, H. Chang, W. G. Wallace-Freedman, S.-W. Chen and P. G. Schultz, *Science*, **1995**, 268, 1738.

[5] G. Briceno, H. Chang, X Sun, P G Schultz and X-D Xiang, *Science*, **1995**, 270, 273.

[6] G. Liu and J. A. Ellman, *J. Org. Chem.*, **1995**, 26, 7712.

[7] F. M. Menger, A. V. Eliseev and V. A. Migulin, *J. Org. Chem.*, **1995**, 60, 6661.

[8] P. P. Pescarmona and J. C. van de Waal, *Catal. Lett.*, **1999**, 63, 1.

[9] W. H. Weinberg, B. Jandeleit, K. Self and H. Turner, *Curr. Opin. Solid State & Mater. Sci.*, 31, **1998**, 104.

[10] W. F. Maier, K. Stöwe, and S. Sieg, *Angew. Chem. Int. Ed.*, **2007**, 46, 2.

[11] D. L. Massart, B. G. M. Vandeginste, L. M. C. Buydens, S. De Jong, P. J. Lewi and J. Smeyers-Verbeke, *Handbook of Chemometrics and Qualimetrics: Part A and B*, Elsevier, Amsterdam, **1997**.

[12] R. G. Brereton, *Chemometrics : Data Analysis for the Laboratory and Chemical Plants*, John Wiley & Sons Ltd., Bristol, UK, **2003**.

[13] A. Holzwarth, P. Denton, H. Zanthoff and C. Mirodatos, *Catal. Today*, **2001**, 67, 309.

[14] I. E. Maxwell, P. van den Brink, R.S. Downing, A.H. Sijpkes, S. Gomez and T. Maschmeyer, *Top. Catal.*, **2003**, 24, 125.

[15] A. Corma, J.M. Serra, P. Serna and M. Moliner, *J. of Catal.*, **2005**, 232, 335.

[16] A. R. Leach and V. J. Gillet, *An Introduction to Chemoinformatics*, Springer, Dordrecht, Netherlands, **2007**.

[17] B. A. Bunin, B. Siesel, G. A. Morales and J. Bajorath, *Cheominformatics: Theory, Practice and Products*, Springer, Dordrecht, Netherlands, **2007**.

[18] P. Haaland, *Statistical Design and Analysis of Industrial Experiments*, Chapter 3, Marcel Dekker, New York, **1990**.

[19] E. A. Dawson and P. A. Barnes, *Appl. Catal. A*, **1992**, 90, 217.

[20] B. M. Weckhuysen, A. A. Verberckmoes, J. Debaere, K. Ooms, I. Langhans and R. A. Schoonheydt, *J. Mol. Catal. A*, **2000**, 151, 115.

[21] M. Tagliabue, L. C. Carluccio and D. Ghisletti C. Perego, *Catal. Today*, **2003**, 81, 405.

[22] M. L Bricker, J. W. A. Sachtler, R. D. Gillespie, C. P. McGonegal, H. Vega, D. Bem and J. S. Holmgren, *Appl. Surf. Sci.*, **2004**, 223, 109.

[23] M. M. Barsan and F. C. Thyrion, *Catal. Today*, **2003**, 81, 159.

[24] W. G. Hunter and R. Mezaki, *Can. J. Chem. Eng.*, **1967**, 45, 247.

[25] A. Nordin, L. Eriksson and M. Ohman, *Fuel*, **1995**, 74, 128.

[26] J. Sjöblom, K. Papadakis, D. Creaser and C. U. I. Odenbrand, *Catal. Today*, **2005**, 100, 243.

[27] R. Carlson, *Data Handling in Science and Technology* , Vol. 8 : Design and optimization in organic synthesis, Elsevier, Umeå, Sweden, **1992**.

[28] J. C. McWilliams, D. R. Sidler, Y. Sun and D. J. Mathre, *Applying statistical design of experiments and automation to the rapid optimization of metal-catalysed processes in process development*, Merck, JALA, 2005, 10, 394.

[29] R. Carlson, *Chemom. Intell. Lab. Syst.*, **2004**, 73, 151.

[30] F. Stazi, G. Palmisano, M.Turconi and M. Santagostino, *Tetrahedron. Lett.*, **2005**, 46, 1815.

[31] L. Veum, S. R. M. Pereira, J. C. van der Waal and U. Hanefeld, *Eur. J. Org. Chem.*, **2006**, 7, 1664.

[32] C. Klanner, D. Farrusseng, L. Baumes, Mirodatos and C., Schüth, F. *QSAR Combi. Sci.*, **2003**, 22, 729.

[33] S. Sieg, B. Stutz, T. Schmidt, F. Hamprecht and W. F. Maier**,** *J. Mol. Modeling*, **2006**, 12, 611.

2

DESIGN OF EXPERIMENTS FOR HIGH-THROGHPUT CATALYSIS RESEARCH

Abstract

Design of Experiments (DoE) is a methodology that aims at maximising the amount of information obtained from experimentation while minimising the experimental effort. With DoE a statistically efficient set of experiments can be selected from the parameter space and, by the use of regression techniques, the relationship between the system responses and the influencing factors is obtained. The result is a model from which information about the magnitude of effects can be drawn or a response surface can be constructed. The use of DoE can improve greatly the efficiency of experimental effort and enable more accurate results to be obtained than the classical one-factor-at-a-time research method.

2.1 - Introduction

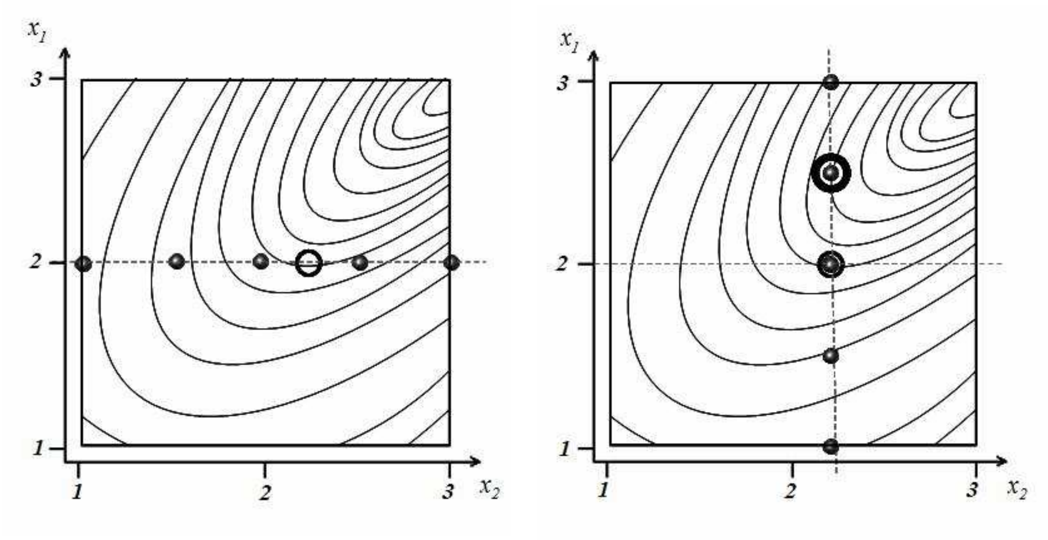
Design of Experiments (DoE) was first developed in the 1920's, by Sir Ronald A. Fisher, a British scientist who studied and proposed a different research approach to conventional one factor at a time (OFAT) research methods in order to maximise the knowledge gained from experimental data [1,2]. Planning a DoE strategy starts with determining the objectives of an experiment and selecting the process factors that are relevant for its study. An experimental design is then chosen in which all the considered factors are simultaneously varied. These designs help to maximise the amount of information obtained from experimentation while minimising the experimental effort required. They are more efficient and lead to more accurate optimization than conventional OFAT research methods. The results obtained can be analysed in a statistical manner, which enables information about their significance and confidence intervals to be gathered. It can be applied to many different scientific fields or basically to any system with measurable inputs and outputs.

In recent years, this methodology has been applied with success in the field of catalyst research. Possible applications in this area include catalyst formulation and preparation [3-6], catalytic kinetic modelling [7,8], reactor engineering [9,10], and the optimization of catalytic reaction conditions [11- 15].

2.2 - Design of experiments (DoE) *vs* one factor at a time (OFAT)

In the conventional research strategy of changing *one factor at a time*, the effect of altering one variable of the studied system is initially measured, the best setting of this variable being then usually chosen for the investigation of the variation of the other system parameters. With this approach a great deal of information can be obtained about how a given variable influences the system under study at the chosen fixed levels. However, the OFAT approach has limitations that can lead to inaccurate or even erroneous conclusions in cases where the influences of the variables on the system are not independent of each other [16], as is usually the case in catalyst optimization studies. Catalytic reactions are commonly characterised by multiple interacting parameter dependencies (like temperature, metal type and composition, pressure, catalyst loading, solvent, etc.) and are therefore not easily optimized by a traditional OFAT approach.

This is demonstrated in Figure 2.1, which represents the difficulty of a proper system optimization via the classical one-variable approach when interaction effects between the parameters are present. In Figure 2.1a variable x_1 is tested at a random fixed level of variable x_2 ($x_2 = 0.0$). The level of the variable x_1 from which the best result obtained is fixed and variable x_2 is then tested at this level (Figure 2.1b). The best result found after testing these two variables is often taken as the optimal solution. In the case of linear dependent variables this methodology can lead to erroneous conclusions, as exemplified in Figure 2.1a-b where the true response surface and optimality area can be observed on the contour-plot graphic background.



a) **b)**

Figure 2.1 - Optimization of the response affected by two factors bearing interaction effects using the conventional one factor at a time approach (OFAT). a) Investigation of the first variable, b) Investigation of the second variable at the best setting of the first.

A multivariate *Design of Experiments* (DoE) approach, on the other hand, is able to take account of these interactions. DoE is an empirical method which directly measures the effects of various changes in the system response. Knowledge about cause and effect relationships can be drawn from a well planned and designed experiment. The parameter space is consistently investigated and interactions between factors can be revealed with this methodology. Multivariate statistically designed experiments are also more economical with respect to time and effort than the classical research approach of altering only one factor at a time.

Using a multivariate DoE approach enables an experimental plan to be constructed that ensures proper experimental coverage of the parameter space and produces representative statistical results (Figure 2.2). The data are fitted to a model that originates a response surface that indicates the location of the optimality area. This model is an approximation of the true response surface, and further close-up experiments in the optimal area region can be performed to find the best optimal solution.

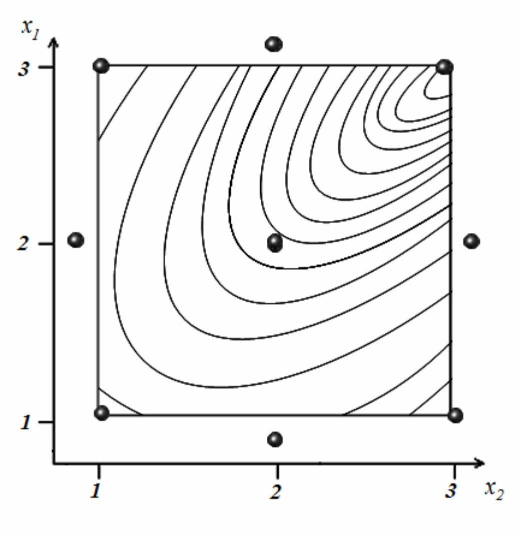


Figure 2.2 - Optimization of the response affected by two factors having interaction effects. Example of a DoE optimization experimental plan using a Central Composite Design (CCD).

2.3 - DoE models

As with most chemical problems, for catalytic chemical systems it is difficult or even impossible to derive a function in an analytical form based on previous knowledge and chemical theory. Experiments are therefore conducted to try to obtain information about the functional relationship between the changing of the factors (x_1, x_2, …, x_k) and the resulting responses (y_i).

$$y = f(x_1, x_2, \ldots, x_k)$$ **Equation 2.1**

In the DoE approach this relationship is established in an analytical form and the Taylor polynomial function is usually employed. This function is called a response model. In most catalytically related issues the inclusion of the second-order terms is considered sufficient for representing the chemical system to optimize [17]. This expression can therefore be written as:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \beta_{12} x_1 x_2 + \ldots + \beta_{ij} x_i x_j + \beta_{11} x_1^2 + \ldots + \beta_{kk} x_k^2 + \varepsilon$$

Equation 2.2

The error term, ε, contains information from the omitted terms of the Taylor expansion and includes both experimental error and the effects of uncontrolled factors in the experiment.

The coefficients of the response model describe how the settings of the experimental variables are linked to the response. The β coefficients of the Taylor function can be obtained via linear regression from the experimental results. In a designed experiment, x_1 and x_2 are systematically manipulated while measuring y, with the objective of estimating the coefficients β_0, β_1, β_2, *and* β_3. The $\beta_k x_k$, terms represent the main effects of the factors, $\beta_{ij} x_i x_j$ the two-way interaction between factor effect terms, and $\beta_{kk} x_k^2$ are quadratic terms accounting for the system’s curvature. The roles played by each factor can be analysed from the estimated magnitude of its coefficient. For obtaining representative coefficient values a consistent design plan needs to be performed.

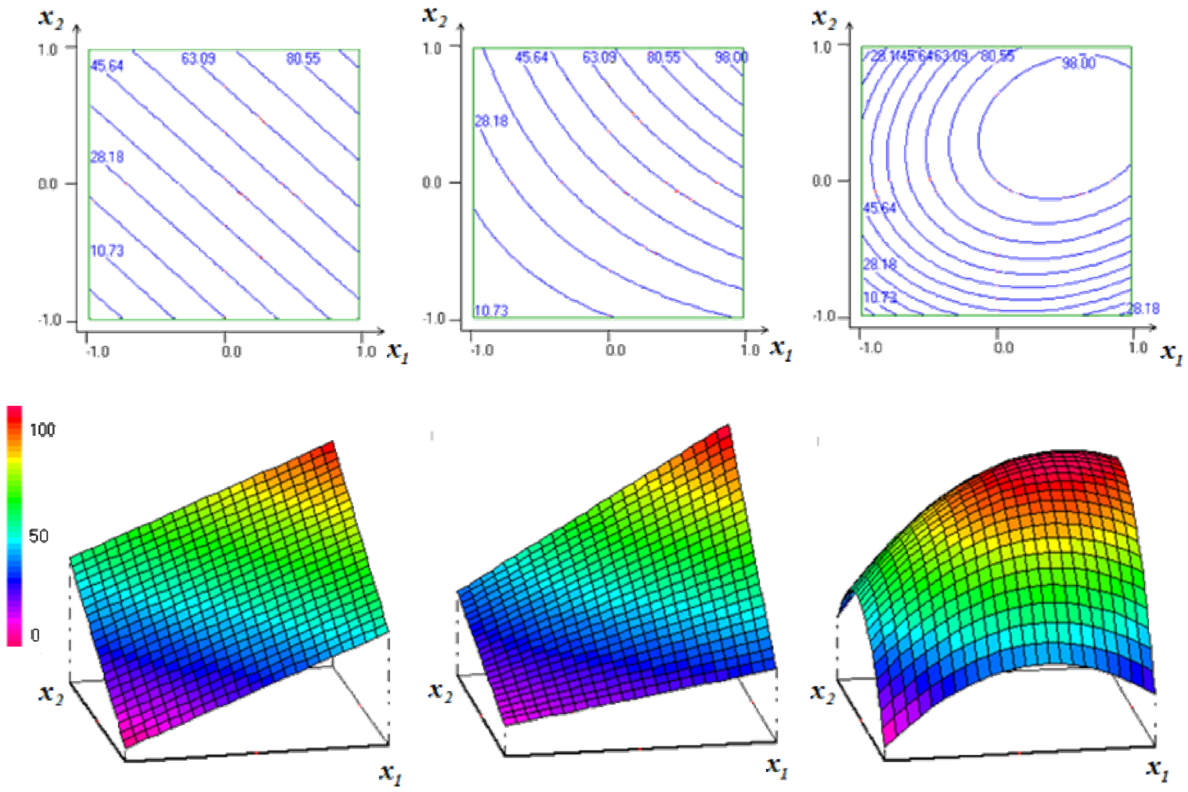
A quadratic order response model is not always necessary to represent the system being studied, or to obtain the desired information. Lower order models are in many cases sufficient. The terms that are necessary to include in the model function are dependent on the objective and complexity of the system under study. The more terms are included in the equation, the more detailed the information can be obtained about the chemical system, but the price to pay is an increase in the experimental effort. A graphical exemplification of the type of information obtained with increased order in the models can be observed in Figure 2.3.

In this figure the increase in information obtained by adding increased order terms into the model function is shown.

In *linear modelling* (Figure 2.3a) the response is approximated by a plane. These models are a simple approximation of the experimental surface, and do not give a perfect description of its variance. But they can be very useful when rough estimates of the influence of the experimental variables are all that is required, as in the screening experiments stage. *Second order interaction models* (Figure 2.3b) allow the twists of the planes to be revealed. This occurs when interactions exist between the variables. And *quadratic models* (Figure 2.3c) allow the curvature of the response surface to be exposed. These surfaces are valuable for locating the final optimum conditions.

In screening experiments, for instance, one is mainly interested in determining, among the many trivial factors, the few parameters of most importance; these can then be investigated further in more detail. The main effects of the factors ($\beta_k x_k$) can be sufficient to obtain this information and a linear model can be simply employed. If interactions between factors are also assumed to be important, the interaction terms ($\beta_{ij} x_i x_j$) should also be considered. In a later stage of research, if more detailed information is necessary and curvature is suspected for the system response, then the quadratic terms ($\beta_{kk} x_k^2$) should then also be adopted. In most cases likely to be met with, third order interactions between the factors are not found to be significant and therefore are often not necessary to be considered for optimization purposes.

A multi-step approach using a sequence of designs, each increasing the degree of information yielded, is a research strategy usually considered to be very efficient (see Chapter 4) [18,19].



a) **b)** **c)**

Figure 2.3 - Example of contour plots and corresponding response surfaces for different order models. a) First order or linear model ($y = \beta_1 x_1 + \beta_2 x_2$), b) Second order with interactions ($y = \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2$) and d) Quadratic model ($y = \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2$)

2.4 - Types of designs

Depending on the type of catalytic problem to be solved, the degree of detailed information wanted, the dimensions of the parameter space and the experimental constraints, a proper design should be chosen that satisfies these demands and limitations. With a proper design the response model coefficients can be obtained in the most efficient and/or accurate way.

Commonly used designs can be divided into screening, response surface or optimization, and mixture design classes. Both classical and non-classical types are available; the classical employ predetermined positions of the experimental points in the parameter space, while in non-classical (computer-based “Optimal” designs) an algorithm selects the experimental points, on the basis of the number of experiments specified by the researcher. This enables also the experimental constraints to be taken into account and so these designs are generally used whenever the classical designs cannot be employed.

The *screening designs* are usually used to select the few important factors that significantly affect the system and discard the many trivial ones. These investigations may include many variables and play an important role in the early stages of an investigation. Their main objective is problem reduction. They allow locating a proper region from the parameter space to perform more detailed investigation via a further optimization design. Common screening designs are the Full and Fractional Factorial designs (Section 2.4.1) and D-Optimal designs (Section 2.4.3).

In *optimization designs,* usually more detailed information is desired from the few significant factors in question. Their main objective is to reveal the optimal values for the experimental factors or to build a mathematical model which can be used to predict the behaviour of the process being investigated. Optimization designs are also called Response Surface designs and some commonly used examples include the Central Composite design, the Doehlert design (Section 2.4.2) and also the D-Optimal design (Section 2.4.3).

Mixture designs are employed whenever it is necessary to optimize the composition of mixtures [20]. In these cases the proportions of the different factors must sum up to

100%, complicating the design and the analysis of the results. An example of a mixture design can be observed in Figure 2.4f. This DoE method is not applied in this thesis.

2.4.1 - Factorial designs

Factorial designs typically place points at regular intervals in the design space, including the highest and lowest values of each factor.

A *Full Factorial* design comprehends the exhaustive performance of all the possible parameter combinations plus replication experiments for error analysis.

In *Fractional Factorial* designs just a part of the total number of runs is performed. The runs are chosen with detail in order to have the most equilibrated design possible [21,22]. The information obtained is dependent on the quantity of experiments and the experimental plan considered. The fractional factorial designs can have different levels of resolution (III, IV and V) and the choice of resolution is dependent on the level of detail of information desired and on the experimental requirements. The price to pay by not doing a full factorial design is that some effect terms of the model are confused by ‘aliases’. This means that the effects of these terms on the system cannot be distinguished. For instance, the main effect of changing one variable is linked to the main effect of changing one other variable and it is not possible to determine which of the variables is indeed responsible for the observed effect. The more reduced the design, the more aliases between the factor effects are usually found, because the number of experiments is insufficient to distinguish between them. More information about fractional factorial designs can be obtained in the literature [11,19].

2.4.2 - Response surface or optimization designs

Depending on the degree of detail necessary (or the model order) the Response Surface or optimization designs can be considerably more experimentally demanding than screening designs. They are usually employed to study the influence of the most important factors after they have been selected via a screening design. Response surface designs (RSM) are used when more than one continuous factor is present. There are several RSM design strategies that can be chosen to model catalytic systems. Figure 2.4

shows geometrical representations of some of the most common strategies for systems with two factors.

The *Pentagon* (Figure 2.4a) and the *Doehlert* design (Figure 2.4b-c) are experimentally economical designs [23]. The Pentagon is more efficient requiring just the amount of experiments necessary to fit a quadratic second order model (in the case of two continuous variables 6 coefficients are needed and so 6 experiments). The reduced amount of experiments may, however, compromise the model prediction qualities. The Doehlert design is more commonly used than the Pentagon design. One of the reasons is the possibility of augmenting it by adding three additional adjacent experimental points in order to continue to explore the parameter space (Figure 2.4b-c).

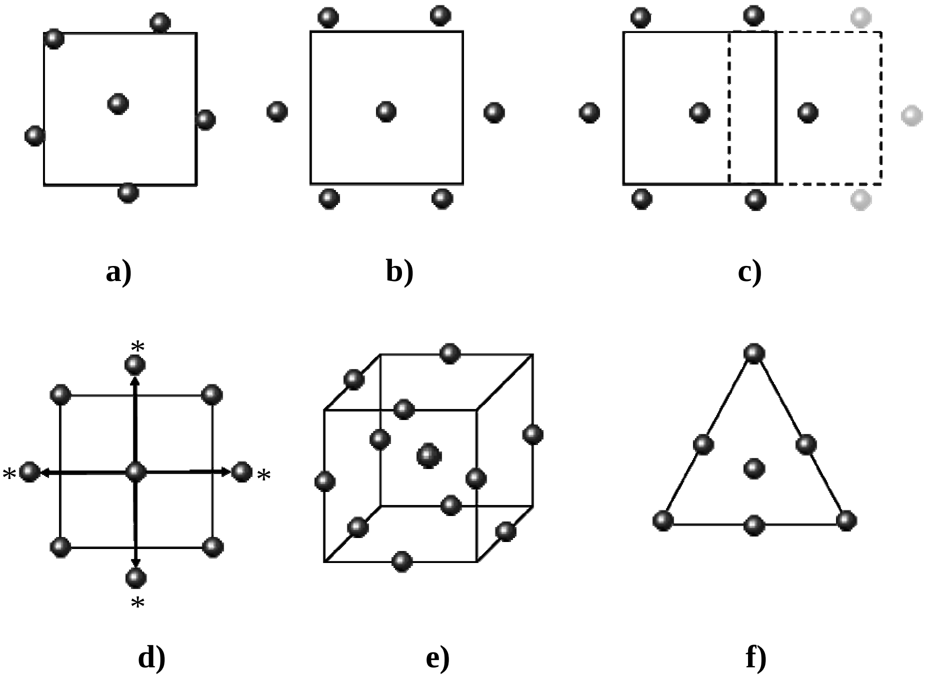


Figure 2.4 - Classical response surface or optimization designs. a) Pentagon design, b) Doehlert design, c) Augmented Doehlert design, d) Central Composite design, e) Box-Behnken design and f) a mixture design. * Star or axial points.

The *Central Composite* design (Figure 2.4d) contains a factorial design with centre points that is augmented with a group of star or axial points [24,25]. The star points establish new extremes for the low and high settings for all factors and allow an estimation of curvature. These designs have very good statistical properties and can be used in demanding simulation issues.

The *Box-Behnken* designs (Figure 2.4e) place points on the midpoints of the edges of the hyper-cubical design region, as well as points at the centre [26]. These designs are used when there are more than three levels and are more economical than either the Central Composite or the Doehlert design.

2.4.3 - D-Optimal designs

Optimal designs differ from those discussed above in being generated from an algorithm based on a particular optimality criterion [27]. The algorithm searches for a sub-set of experiments from the entire parameter space that best satisfies this criterion. All the optimality methods are model-dependent and so the experimenter must specify the model terms to be considered and the final number of design points preferred before the design generation. The computer algorithm usually uses a stepping and exchanging process to select the sub-set of experiments. There are various forms of optimality criterion that can be used which are all based on some function of the information matrix. The most common criterion is the D-Optimal. The D-Optimal design algorithm seeks to maximise the $|X'X|$, the determinant of the information matrix ($X'X$) of the design. This criterion results in minimising the generalised variance of the parameter estimates in a pre-specified model [20].

The Optimal designs are usually applied whenever the classical factorial or RSM designs cannot be used due to constraint limitations of the parameter space or experimental effort, and/or the presence of discrete factors with many levels (a common case in catalysis issues). Computer-based optimal designs can be used both for screening and optimization, depending on the model specified.

2.5 - Multi-design approach

Multi-design DoE approaches resemble 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 set of variables from one stage to the next and detailed information is obtained in the end for the most important factors [28]. 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 common strategy is to start with a screening design to select the most important factors affecting the system. Then, setting the less influential factors to the most advantageous levels, a series of more detailed designs (such as full factorial designs) can be conducted to better understand the effect of the most important factors on the response. The process can, if necessary, be continued with additional optimization steps, enabling the response surface to be further modelled until a satisfactory description of the region of interest is obtained. Such a strategy is usually highly efficient and is exemplified in Chapter 4.

2.6 - Steps for a DoE research strategy

A DoE research strategy should be carefully planned in order to make it possible to obtain in the most efficient way the information required from the experimentation. The planning of a DoE experimental research programme differs from the conventional OFAT strategy, and the more important issues will be briefly explained in this section. Common steps in this planning are: 1) set the objectives and responses, 2) select the parameter space, 3) select a proper experimental design, 4) perform the experimentation, 5) analyse and interpret the results, and prepare the next step based on the previous results obtained.

2.6.1 - Setting objectives (responses)

The set-up of the experimental design depends on the question to be answered. The goal of the experiment needs to be clear to be able to make the right experimental plan. A thoroughly thought-out experimental design is after that essential for being able to obtain the desired answers. For this, the relevant factors that may influence the system need to be determined, and the experimental constraints considered.

In many cases there is the need of multi-responses optimization. The choice of the most relevant response to optimize or its importance in the hierarchy depends on the goal of the catalytic challenge. Often the yield of the desired product is the most important response; however other responses such as selectivity, stereo-selectivity or conversion can be the most important ones to consider. Other responses may be for instance the production cost of the product, consumption of a specific starting material, time for reaching a certain conversion, etc. There exist some mathematical techniques that allow the simultaneous analysis of several responses. If the number of responses to optimize is small, the surfaces for each of the responses can be evaluated by visual inspection of the contour plots, preferably superimposed on the same plot [19]. Another method is to combine the responses together into one criterion, the Desirability function, which is then optimized. If there are many responses to consider, their simultaneous evaluation can become very complex. Multivariate analysis of the response matrix by principal components analysis (PCA) or principle least squares (PLS) can simplify this evaluation [19]. The Desirability function method has been used in Chapter 5.

2.6.2 - Selection of the parameter space

The selections of factors and factor levels that can influence the catalytic system in the pursuit of the set goal have to be carefully considered with the previously defined objectives in mind. Existing knowledge about the system under study should be primarily used at this stage.

A catalytic reaction is a complex system that can be affected by several factors. Common factors influencing a catalytic system are for instance the catalyst type and loading, its stability, the pressure, temperature, and the solvent; this makes it a complex system to study and to optimize.

The variables can be usually divided into continuous or discrete factors. Variations such as: substrates, reagents, catalysts, solvents, etc. (the categorical variables) are discrete and constitute what is sometimes called the reaction system. The experimental conditions, such as: concentrations of a substrate, reagent or catalyst; temperature, pressure, stirring rate, etc. are usually continuous and can be set to any value in their range of variation.

Well-planned measurement of continuous variables allow response model fitting and interpolation estimates of the non-measured variable levels. While among the discrete variables, measurement estimates of non-tested levels cannot be obtained.

2.6.3 - Choice of experimental design

Depending on the objective of the experiment and on the experimental constraints (time, limited number of experiments, etc.) there are several possible types of design that can be used to plan the experimental effort. An overview of these designs can be seen in Section 2.4.

2.6.4 - Performance of the experimental plan

When performing the experiments a high degree of attention should be paid to avoid experimental errors (random but mainly systematic). Known, but unwanted, sources of variation caused for instance by changes in reactants, equipment or personnel can be ‘blocked’. With *blocking* the variation caused by this change can be accounted for and not taken into consideration for the true estimation of the effect of the factor on the response. Furthermore, results can be revealed that otherwise would be obscured by the variation.

With *randomisation* the run sequence of the experimental units is determined randomly. This decreases the possibility of errors to be associated with specific factor levels.

Blocking is used to remove the effects of important uncontrollable but known noise sources and randomization is used to reduce the contaminating effects of the other uncontrollable variables. A rule of thumb found valuable in DoE is: ‘block what we can and randomize what we cannot’ [29].

2.6.5 - Analysis and interpretation of the results

The analysis and interpretation of the experimental results can be performed in an initial stage simply by a preliminary visual inspection of the data. Most of the available software packages possess, however, graphical and statistical aid tools for their analysis and interpretation which largely improves the data interpretation and information knowledge.

With a well designed DoE experimental plan the effect that varying a factor can have on the system response can be computed in the following manner:

$$Effect_{(Factor_A)} = \frac{\sum {y_A}^{+}}{{n_A}^{+}} - \frac{\sum {y_A}^{-}}{{n_A}^{-}}$$ **Equation 2.3**

Where n_A^+ and n_A^- refer to the number of data points collected at the + and - level of factor A respectively and (${y_A}^+$) and (${y_A}^-$) to the associated responses.

The effects of changing experimental factors are transmitted via the corresponding model parameters into the systematic variation of the response y. To be significant this variation must be above the system noise level. Hence, to determine the influence of experimental variables we must determine the values of their model parameters and then compare these values to an estimate of the experimental error. In this thesis a graphical representation of the effects magnitude is usually presented with the significance limits at the 95% confidence interval (Chapters 3, 4 and 7). Other statistical tests can be performed based on the statistical method of analysis of variance (ANOVA). All the previous tests can be automatically performed via the common DoE software.

2.7 - Designs of experiments in HTE catalysis research

Automated equipment in conjunction with statistical design of experiments (DoE) can accelerate both the screening and the optimization phase of catalytic reactions. The large numbers of factors that affect a catalytic system usually implies a large experimental effort. DoE enables the maximum of information to be extracted with an efficient experimental plan, while High-Throughput methodology makes possible faster experimentation with greatly reduced non-reproducibility and human error. The simultaneous experimental planning mode of the DoE strategy combines well with the batch parallel mode of operation of the HTE equipment used in catalysis research. In this way an entire design can be performed simultaneously in the HTE equipment. The use of DoE and HTE are therefore two highly compatible methodologies for both catalysis screening and optimization projects.

2.8 - Summary

Design of experiments is a better optimization methodology than one factor at a time optimization. With the DoE methodology the parameter space is consistently searched and the probability of locating the system optimal conditions is largely increased. The multivariate nature of the DoE methodology necessitates a different strategy of experimental planning, leading to the need of the scientist to change his/her approach to optimization. The DoE methodology presents the versatility to deal with the most common catalytic chemistry issues and its batch design output is very well suited for the parallel High Throughput Experimentation.

2.9 - References

[1] R. A Fischer, *Statistical methods for research workers*, Oliver and Boyd, Edinburgh, **1925.**

[2] R. A. Fischer, *Design of Experiments*, Oliver and Boyd, Edinburgh, **1935**.

[3] E. A. Dawson and P. A Barnes,. *Appl. Catal. A*, **1992**, 90, 217.

[4] B. M. Weckhuysen, A. A. Verberckmoes, J. Debaere, K. Ooms, I. Langhans and R. A. Schoonheydt, *J. Mol. Catal. A*, **2000**, 151, 115.

[5] M. Tagliabue, L. C. Carluccio, D. Ghislett and C. Perego, *Catal. Today*, **2003**, 81, 405.

[6] M. L. Bricker, J. W. A. Sachtler, R. D. Gillespie, C. P. McGonegal, H. Vega, D. Bem and J. S. Holmgren, *Appl. Surf. Sci.*, **2004**, 223, 109.

[7] M. M. Barsan and F. C. Thyrion, *Catal. Today***, 2003**, 81, 159.

[8] W. G. Hunter and R. Mezaki, *Can. J. Chem. Eng.*, **1967**, 45, 247.

[9] A. Nordin, L. Eriksson and M. Ohman, *Fuel*, **1995**, 74, 128.

[10] J. Sjöblom, K. Papadakis, D. Creaser and C. U. I. Odenbrand, *Catal. Today*, **2005**, 100, 243.

[11] R. Carlson, *Data Handling in Science and Technology* , Vol. 8 : Design and optimisation in organic synthesis, Elsevier, Umeå, Sweden, **1992**.

[12] J. C. McWilliams, D. R. Sidler, Y. Sun and D. J. Mathre, *Applying statistical design of experiments and automation to the rapid optimisation of metal-catalysed processes in process development*, Merck, JALA, **2005**, 10, 394.

[13] R.,Carlson, *Chemom. Intell. Lab. Syst.*, **2004**, 73, 151.

[14] F. Stazi, G. Palmisano, M. Turconi and M. Santagostino, *Tetrahedron Lett.*, **2005**, 46, 1815.

[15] L. Veum, S. R. M. Pereira, J. C. van der Waal and U. Hanefeld, *Eur. J. Org. Chem.*, **2006**, 7, 1664.

[16] L. Davies, *Efficiency in research, development, and production: The statistical design and analysis of chemical experiments*, Royal Society of Chemistry, Cambridge, 1993.

[17] R. Carlson, *Chemom. Intell. Lab. Syst.*, **2004**, 73, 151.

[18] R. Marchetti and M. E. Guerzoni, *Cerevisia Biotechnol*, 16, **1991**, 24.

[19] R. Carlson, Design and optimization in organic synthesis, Elsevier, Amsterdam, **1992**.

[20] D. L. Massart, B. G. M. Vandeginste, L. M. C. Buydens, S. De Jong, P. J. Lewi and J. Smeyers-Verbeke, *Handbook of Chemometrics and Qualimetrics: Part A and B*, Elsevier, Amsterdam, **1997**.

[21] G. E. P. Box and J. S. Hunter, *Technometrics*, **1961**, 3.

[22] G. E. P. Box, J. S. Hunter, W. G. Hunter, *Statistics for experimenters*, Wiley, New York, **1978**.

[23] D. H. Doehlert, Appl. Stat., **1970**, 13, 231.

[24] G. E. P. Box and K. B. Wilson, *J. R. Stat. Soc*., B, **1951**, 13, 1.

[25] W. Spendley, G. R. Hext, F. R. Himsworth, *Technometrics*, **1962**, 4, 441.

[26] G. E. P. Box and D. W. Behnken, *Technometrics*, **1960**, 2, 455.

[27] V. V. Federov, *Theory of Optimal Experiments*, Academic Press, New York, **1972**.

[28] P. Haaland, *Statistical Design and Analysis of Industrial Experiments*, Chapter 3, Marcel Dekker, New York, **1990**.

[29] M. J. Anderson and P. J. Whitcomb, *DoE simplified: practical tools for effective experimentation*, Productivity Press, Portland, **2000**.

3

HIGH-THROUGHPUT METHODOLOGY COMBINED WITH DESIGN OF EXPERIMENTS APPLIED TO THE EXPLORATION OF A COMPLEX CHEMICAL SYSTEM: TERPENE ADDITION REACTIONS

Abstract

A Design of Experiments strategy in combination with high-throughput experimentation methods was employed for the exploration of addition reaction, namely the hydration reaction of terpenes. Several starting materials: *α*-pinene, *β*-pinene, camphene, limonene, carene and myrcene were studied simultaneously in relation to the factors determining activity and selectivity. The effects of variation of acid catalyst and solvent on the reactions were investigated and the product selectivities rationalized in terms of the available reaction pathways towards *α*-terpinyl, isobornyl and bornyl products. In terms of competing nucleophiles, the highest selectivities (91%) were obtained for the isobornyl methyl ether product, whereas isoborneol was the only alcohol derivative obtained with selectivities above 73%. An increase in activity was observed by using stronger acids, especially a heteropolyacid (HPA), but acid strength alone was not found to be a factor exerting a major influence on the final product mixture. The multi-design DoE technique proved to be highly valuable in the elucidation of the main patterns and trends of this complex chemical reaction system.

3.1 - Introduction

With the arrival of Combinatorial Chemistry and High-Throughput-Experimentation (HTE) equipment, today's chemist has available powerful tools for studying problems with a much wider scope than previously [1,2]. Due to the higher throughput a more comprehensive approach is possible where, in principle, most of the parameters thought to influence a chemical problem can be studied in a more consistent and uniform manner. Optimization methodologies such as Design of Experiments (DoE) [3,4], Genetic Algorithms (GA) [5-7], Simulated Annealing (SA)[8] and others (see Chapter 6, 7), are being adopted in order to explore the chemical parameter space in a more rational and efficient way [9]. Design of Experiments (DoE) is considered to be one of the most suited methodologies for the application of HTE in synthetic organic chemistry research [10] because this methodology allows system optimization together with a good understanding of the problem investigated and a considerable reduction of the experimental effort. An overall knowledge of the response surface is acquired, in contrast to other screening techniques (like GA and Simulated Annealing) that focus on the subspace where it is hoped to find the optimal solutions.

In this chapter a complex chemical reaction system was chosen and studied using a Design of Experiments strategy. Terpene chemistry has been intensively investigated since the beginning of the twentieth century and has contributed to the conception of several fundamental organic chemistry theories (Wagner-Meerwein rearrangement, Woodward-Hoffman rules of cycloaddition reactions, considerations about the non-classical carbenium ion, etc.) [11,12]. Numerous synthetic terpene derivatives are in industrial production as flavor and fragrance chemicals. It was chosen to study the addition of water to unsaturated terpenes, namely the hydration reaction in the presence of competing nucleophiles, as an example of a chemically complex and rich system to investigate [13-16]. Acid-catalysed reactions of terpenes are usually not very selective due to the existence of alternative carbenium ion pathways [17]. It is clear that finding a selective reaction will not be an easy task, because the carbenium ions are not very stable and can rearrange easily without the further aid of or interaction with a catalyst. In addition, possible competition between water and protic solvents or catalysts for the nucleophilic role further increases the complexity of the system and the diversity of selective product candidates. An explorative research of a chosen parameter space is intended in this chapter in order to obtain an initial understanding of the system studied. This is performed by using a systematic approach involving experimental design schemes in combination with High-Throughput-Experimentation techniques.

The reactivity of several terpene substrates is investigated simultaneously. This option allows the study of existent common carbenium ion reaction paths and has also a financial and economic driver. Despite the fact that terpene-derived products are widely used in many applications, their volumes and/or prices are typically low and would not justify a focused research effort. However, by combining more than one starting material and thus inherently looking for a wider variety of possible feeds and products within one design approach, the costs can be shared.

The search for the main patterns and trends of a complex terpene chemical reaction system was performed in this Chapter by DoE techniques.

3.2 - Results and Discussion

3.2.1 - Initial Screening

The parameter space chosen for investigation is summarised in Table 3.1. The six substrates (shown in Figure 3.1) are all inexpensive and derived from renewable resources.

These commonly available monoterpenes were selected with regard to their chemical and commercial interest. As catalysts, several commonly used Brønsted acids were chosen. The solvent library comprised both protic and aprotic types. A single temperature of 50 ºC, 18 h reaction time, 0.5 N catalyst and 1 M terpene concentration was employed. And the water concentration was studied at two levels: 1.5 and 3 M.

Table 3.1 - Initial screen parameter space.

Substrate	Catalyst	Solvent	[Water] (M)
α-Pinene	Boric acid	Acetone	1.5
β-Pinene	Acetic acid	IPA	3
Camphene	Oxalic acid	Methanol	
3-Carene	PTSA	DMF	
Limonene	H_2SO_4	THF	
Myrcene	HCl		
	HPA		

Abbreviations: IPA, Isopropyl alcohol, DMF *N*,*N*-Dimethyl formamide, THF Tetrahydrofuran, PTSA p-Toluenesulfonic acid, HPA Heteropolyacid - phosphotungstic acid ($H_3PW_{12}O_{40}$).

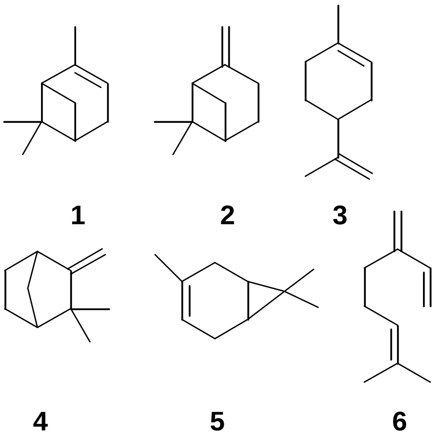
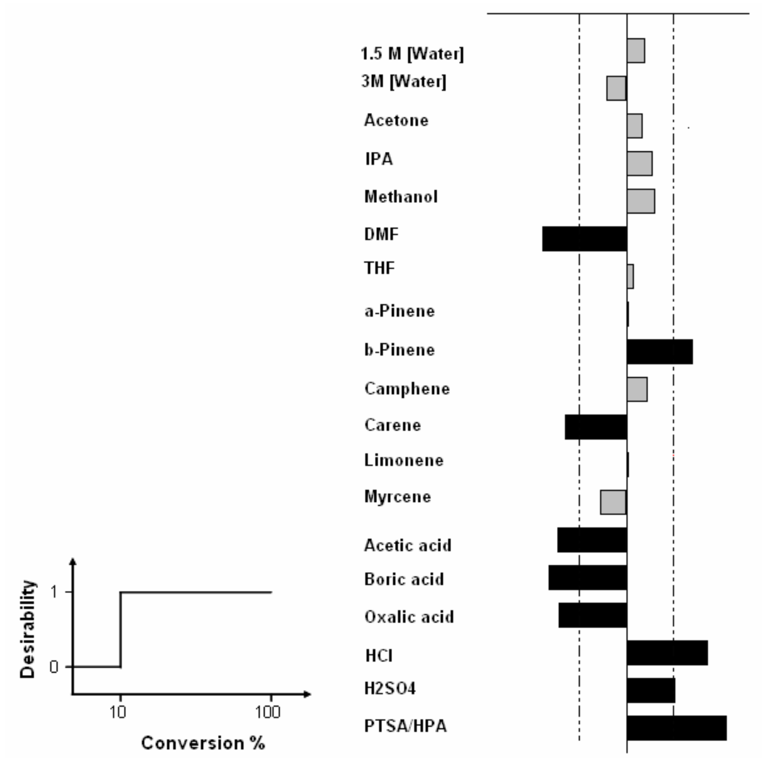


Figure 3.1 - Monoterpene substrates. 1) α-pinene, 2) β-pinene, 3) limonene, 4) camphene, 5) 3-carene and 6) myrcene.

The object of the initial screen was to obtain basic information about the selected parameter space, such as whether the experimental conditions would lead to acceptable levels of activity with the chosen substrates. A D-optimal criterion was used to choose part of the total combinations of variables (70 of the 420 possible combinations). The D-optimal criterion chooses a sub-set of the total combination of experiments which is ‘best-distributed’ (the experiments are situated as far from each other as possible). The main effects of these factors (presented in Table 3.1) on the reactivity of the system are summarised in Figure 3.2. In this representation (and in subsequent similar figures) the lengths of the bars indicate the magnitude of the effects relative to the average value. The bars directed to the right mean 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.



a) **b)**

Figure 3.2 - Main parameter effects on conversion for the initial screen. a) Activity response transformation (Desirability) considered for calculating the main parameter effects on the response in regard to reactive or non reactive systems. b) Main effects of the parameter space in relation to the desired Desirability. The bars directed to the right mean 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.

It could be observed that the substrates differ markedly in reactivity. The catalyst is the factor showing the highest influence on activity, followed by the solvent. Under the conditions employed, the weak acids: acetic, boric and oxalic acid, showed only very low activities and were consequently eliminated from consideration in the second screen. A trend of increased reactivity with increased acidic strength is observed. Of the five solvents, activities in DMF were in general low and so this solvent was likewise excluded from the set investigated in the second screen. The reason for the low activity in DMF may be due to its unstable character in an acidic medium and/or its electron donor properties that can compete for the catalytic protons. Low reactivities were also obtained when 3-carene and myrcene were used as the starting substrate. Due to these low conversions and in addition miscibility problems with myrcene under the reaction conditions employed, these two substrates were not considered for further screening. The water concentration was found to have a smaller influence, without any discernable trend in this initial design.

On the basis of this initial screening, a subspace of the initial parameter space with desirable reactivity levels was allocated for further study.

3.2.2 - Second Screening

The parameter space for the second screen is summarised in Table 3.2. In this new parameter space four of the six original substrates were retained and four catalysts, four solvents and two water concentrations were investigated. A D-optimal design algorithm is employed to select 100 experiments from the possible 128 experimental reactions (75% of the total combination of variables) including 4% replicates for error analysis. With this design, information about the major effects and 2-level interactions between the parameters can possibly be obtained.

Table 3.2 - Second screen parameter space (128 possible combinations)

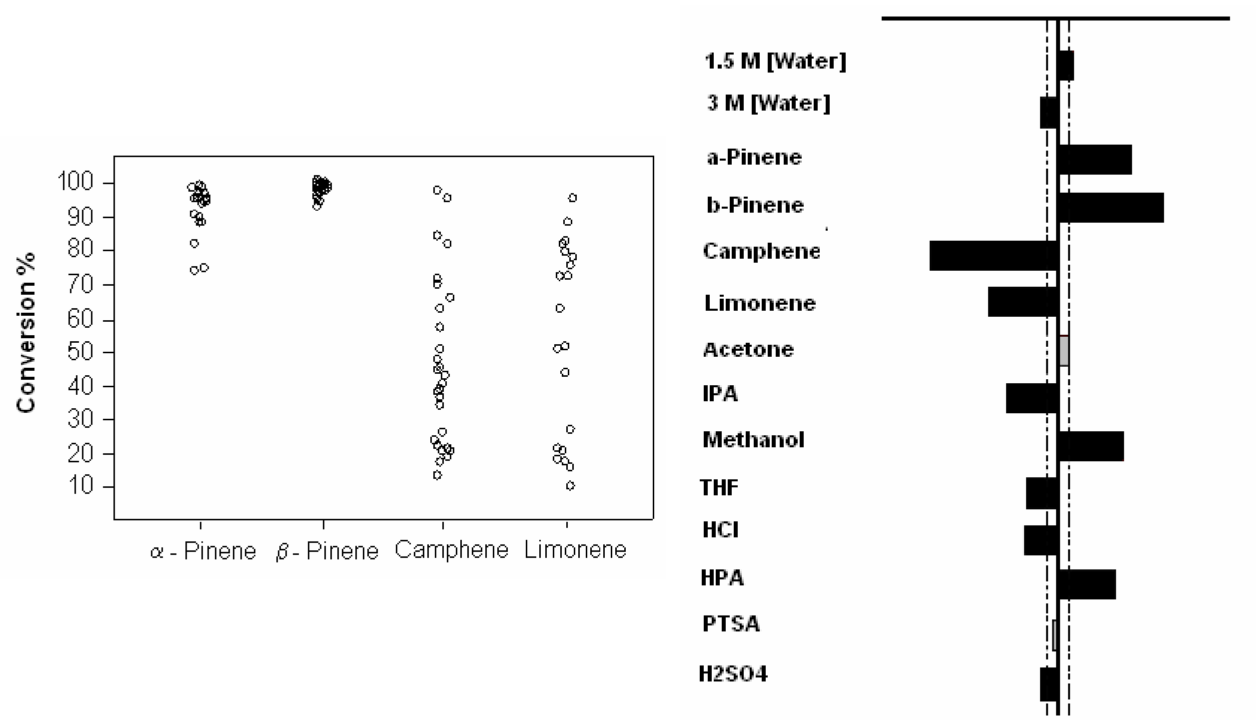
Substrates	Catalyst	Solvents	[Water] (M)
α-Pinene	PTSA	Acetone	1.5
β-Pinene	H_2SO_4	IPA	3
Camphene	HCl	Methanol	
Limonene	HPA	THF	

Reaction temperature 50°C; reaction time 18 h.

3.2.2.1 - Parameter effects on activity

A schematic representation of the reactivity trends versus the substrate, obtained in all experiments (100) from this second screen, is shown in Figure 3.3a. A statistical analysis scheme of the main effects of the various factors on the conversion response is shown in Figure 3.3b.

The conversion trend (Figure 3.3) in relation to the substrate is now clear, with *β*-pinene, in general, being the most reactive one, followed by *α*-pinene, limonene and then camphene. The higher activity of *α*- and *β*-pinene is easily explained by stereochemical stress release of their 4-membered rings via protonation and subsequent skeletal rearrangement, while the slightly lower reactivity of *α*-pinene compared to *β*-pinene is the result of its endo-cyclic double bond.



a) **b)**

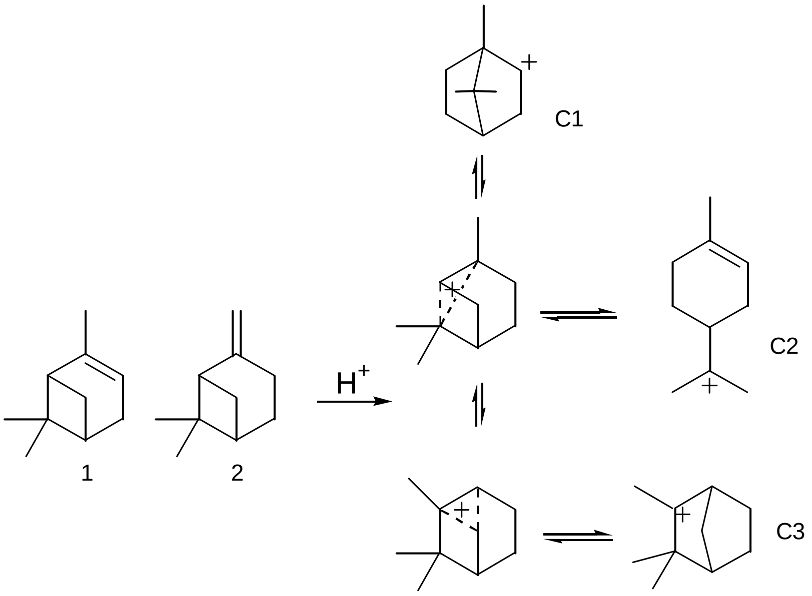
Figure 3.3 - Effects on conversion. a) Substrate conversion as function of the substrate for all the reaction conditions. b) Parameter effects with respect to the substrate conversion.

Since α- and β-pinene give almost complete conversions in most cases (see Figure 3.3a), it follows that the influence of the other parameters will be most discernible in the conversion response of camphene and limonene. The strong acids selected from the first screening show also a trend of reactivity according to their acidity. This effect is especially evident for the strongest acid HPA which consistently gives rise to the highest conversions. The solvent is also an influential parameter. The highest activities are generally found in methanol, which is mainly due to its additional role as a good nucleophile in the attack on the carbenium ions. The effect with respect to conversion of changing the water concentration is again found to be small.

3.2.2.2 - Parameter effects on selectivity

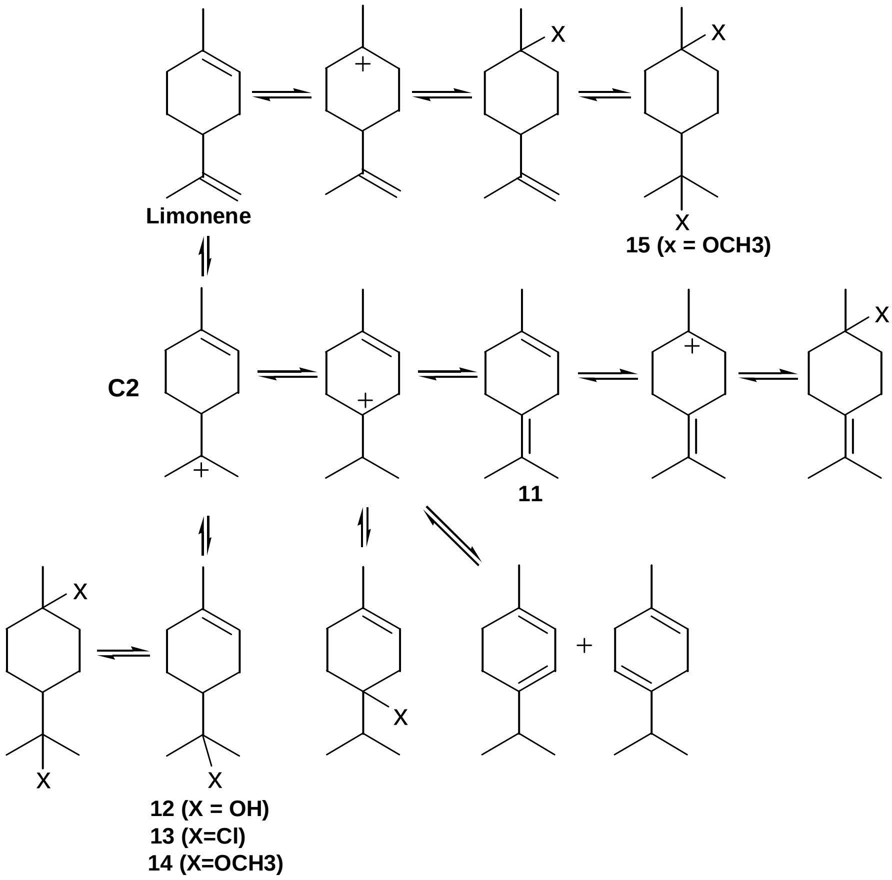
Rearrangement pathways

The main reaction pathways undergone by the substrates after protonation are shown in Schemes 3.1-3.

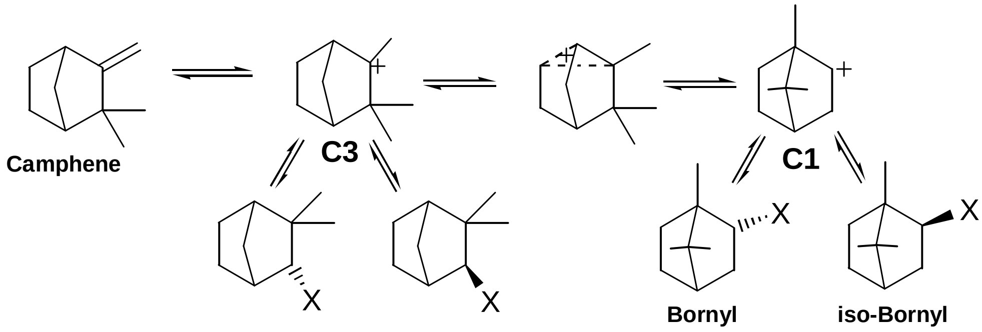


Scheme 3.1 - Main pathways of the acid-catalysed rearrangement of α-pinene (**1**) and β-pinene (**2**). **C1**: bornane skeleton carbenium ion, **C2**: p-menthane skeleton carbenium ion, **C3**: camphene skeleton carbenium ion.

For *α*- and *β*-pinene the products resulting from the rearrangement path of carbenium ion **C2** are in the majority, but substantial amounts of products obtained via rearrangements of carbenium ions **C1** and **C3** (see Schemes 3.1-3) are also observed.



Scheme 3.2 - Main pathways of the acid-catalysed rearrangement of limonene and **C2** carbenium ion.



Scheme 3.3 - Main pathways of the acid catalysed rearrangement of camphene and **C1** and **C3** carbenium ions.

Scheme 1 shows the theoretically possible conversion between the **C1**, **C2** and **C3** carbenium ions. From the analysis of the resulting reaction mixture (see Figure 3.5) it is observed that in the case of limonene the compounds follow almost exclusively the rearrangement path of the carbenium ion **C2** (Scheme 3.2) showing that in practice the rearrangement of carbenium **C2** into a strained bi-cyclic structure is not favoured. In the case of camphene (Scheme 3) it is found that compounds obtained via rearrangement of carbenium ion **C1** represent the majority of products formed.

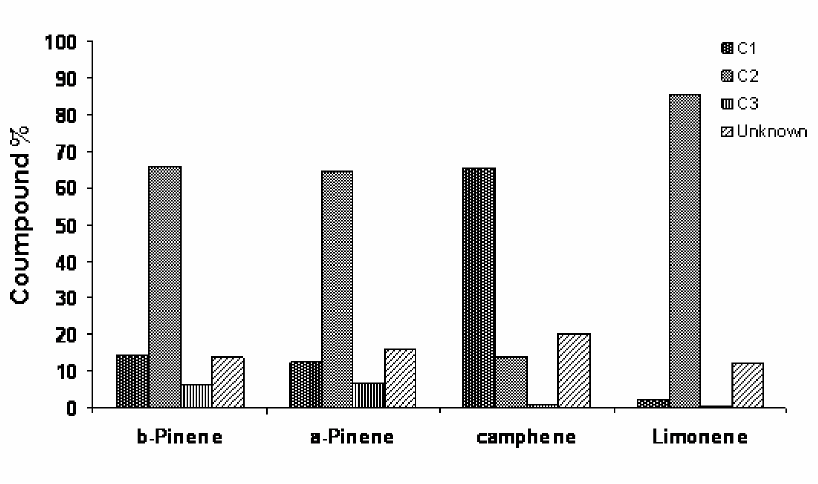


Figure 3.5 - Percentage of compounds present in the reaction mixture with the **C1**, **C2** and **C3** carbenium ions skeleton.

Selectivity: general

The selectivities observed are mainly determined by the carbenium-ion route following protonation of the substrates. The carbenium ion with a p-menthane skeleton (**C2**) has a larger range of possible rearrangements available compared to carbenium ions **C1** and **C3**, implying an increased difficulty in obtaining high selectivities towards any specific product.

As shown in Figure 3.6, camphene with its more limited rearrangement possibilities is the substrate that usually gives the highest selectivities towards the main reaction product, followed by limonene (the products of which follow mainly the path of carbenium ion **C2**). The lowest selectivities are obtained with the highly reactive *α*- and *β*-pinene, the products of which are obtained mainly via both the **C1** and **C2** carbenium-ion pathways (see Scheme 1 and Figure 3.5 - 6). **C3** products are also theoretically possible but were not found in significant amounts.

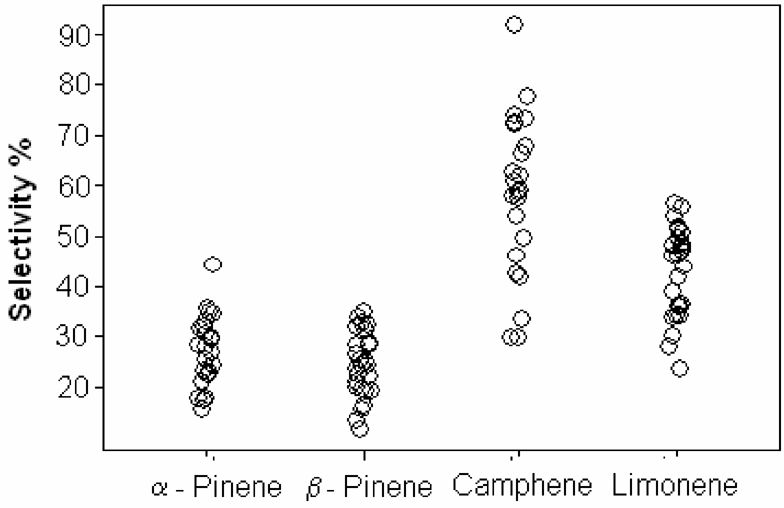


Figure 3.6 - General results on selectivity towards the main product obtained in each reaction: % selectivity *vs*. substrate.

Products

When the aprotic solvents THF and acetone, and catalysts other than HCl were used, alcohol derivatives were the main products obtained in the reaction. It was also observed that some acid counter-ions and protic solvents can also act as nucleophiles, leading to competition with water resulting in a more complex product mixture. With respect to the addition of the acids, this addition was only significant in the case of HCl where substantial amounts of chlorinated products were found. The small size and Lewis-basic properties of the Cl^- anion make it a good nucleophile for attacking the protonated substrates. In fact, even in the presence of less than stoichiometric amounts and in the presence of an excess of water, chlorinated derivatives are usually the main products obtained when HCl is present (see Figure 3.7). Regarding the addition of protic solvents, it is clear from Figure 3.7 that methoxy compounds prevail whenever methanol is present. For IPA, even if in much larger excess than water or the acids, ether formation is less marked and water addition to give alcohols, or chloro-products in the case of HCl catalyst, is the prevailing reaction. From Figure 3.7 it is clearly seen that under the actual reaction conditions the order of preference with respect to nucleophile addition is MeOH > HCl > H_2O > IPA. Rearrangement to the diene product terpinolene is also observed, this being the favoured product when water is present as the only nucleophile at its lower concentration level of 1.5 M.

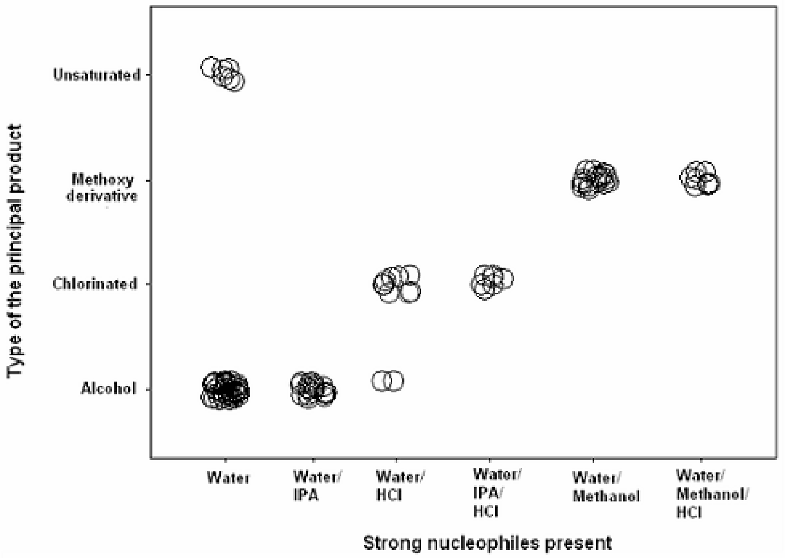


Figure 3.7 - Main product types obtained according to the presence of possible nucleophiles in the reaction mixture. Only the nucleophiles from which corresponding addition products were detected in reasonable amounts (water, HCl, IPA, methanol) are considered.

3.2.2.3 - Selectivities and yields towards the main reaction products obtained

The principal reaction product (defined as the major compound present in the product mixture) obtained for each reaction, was in general that known from the large volume of published work on monoterpene hydration/isomerisation reactions [11-16]. The principal products are shown in Figure 3.8.

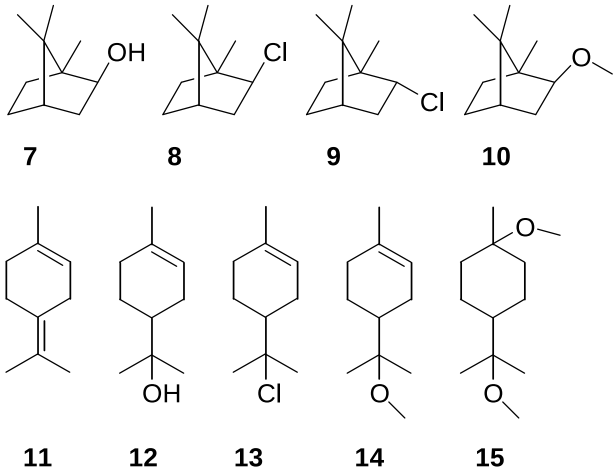
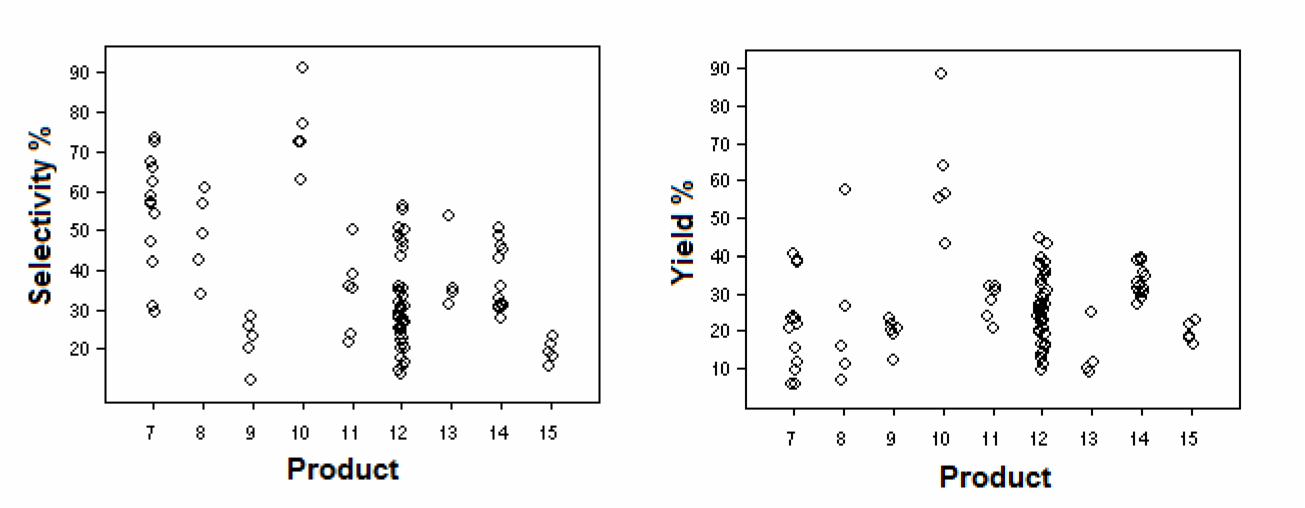


Figure 3.8 - Main products obtained: 7) isoborneol, 8) isobornyl chloride, 9) bornyl chloride, 10) isobornyl methyl ether, 11) terpinolene, 12) *α*-terpineol, 13) *α*-terpinyl chloride, 14) *α*-terpinyl methyl ether, 15) terpin dimethyl ether

The principal products obtained from *α*-pinene, *β*-pinene and limonene are formed via the **C2** carbenium-ion path. These are the rearranged product terpinolene (**11**) together with the alcohols *α*-terpineol (**12**) and/or their chloro- (**13**) and methoxy- (**14**) substitution products, depending on the reaction conditions employed. When HCl is present, bornyl chloride (**9**) is the predominant product from the two pinenes and is formed via the carbenium ion **C1**. Camphene yielded mainly isoborneol (**7**) and its chloro- (**8**) and methoxy- (**10**) derivatives.

In Figure 3.9 the selectivities and yields of the main products obtained for all the reactions are summarised. It can be observed that the only product obtained in close to quantitative yields is isobornyl methyl ether (**10**) from camphene. The chlorinated products (**8-9**) are usually obtained with lower selectivities than the corresponding methoxy derivative (**10**) due to competition between the water and HCl nucleophiles, compared to the competition between methanol, present as solvent, and water.

Figure 3.9 presents the main results obtained per product. The *α*-Terpinyl derivatives *α*-terpineol (**12**), *α*-terpinyl chloride (**13**) and *α*-terpinyl methyl ether (**14**) have responses below 50% for yield, *α*-terpinyl chloride being obtained as main product only in the reaction of limonene. High selectivities to isobornyl derivatives: isoborneol (**7**), isobornyl chloride (**8**) and isobornyl methyl ether (**10**) are obtained from camphene, up to 73, 60 and 91 % respectively. The corresponding yields are lower due to low camphene conversions (Figure 3.9). Bornyl derivatives are formed as primary products only when α- and β-pinene are used; moreover, only bornyl chloride (**9**) was obtained in significant amounts, borneol and bornyl methyl ether being in all cases just smaller secondary products. The effect of the factors influencing the synthesis of the main product derivatives is discussed below.



a) **b)**

Figure 3.9 - Selectivities and yields of the principal products, a) Selectivity, b)Yield. 7) isoborneol, 8) isobornyl chloride, 9) bornyl chloride, 10) isobornyl methyl ether, 11) terpinolene, 12) α-terpineol, 13) α-terpinyl chloride, 14) α-terpinyl methyl ether, 15) terpin dimethyl ether.

α-Terpinyl derivatives

From Figure 3.10 we see that the largest selectivities to α-terpinyl products are usually obtained from limonene, followed by α- and β-pinenes. For the synthesis of α-terpineol (see Figure 3.10a) THF gives the best results, followed by IPA and acetone. The water concentration is a relevant factor, higher concentrations being preferential. HPA seems to be the best catalyst and the presence of either HCl or methanol as competing nucleophiles is prejudicial for the synthesis of α-terpineol.

For the formation of *α*-terpinyl chloride (Figure 3.10 b) and *α*-terpinyl methyl ether (Figure 3.10c), besides the obvious necessity for the presence of the corresponding nucleophiles, it was found that the difference in the water concentrations considered does not have a significant influence. Acetone was the preferred solvent followed by THF in the *α*-terpinyl chloride synthesis. The best catalyst for the *α*-terpinyl methyl ether synthesis was H_2SO_4, which was slightly better than the stronger acid HPA. The latter forms considerable amounts of terpin dimethyl ether (**15**) obtained by consecutive reaction. In Table 3.3 the best results obtained in the screening are summarised.

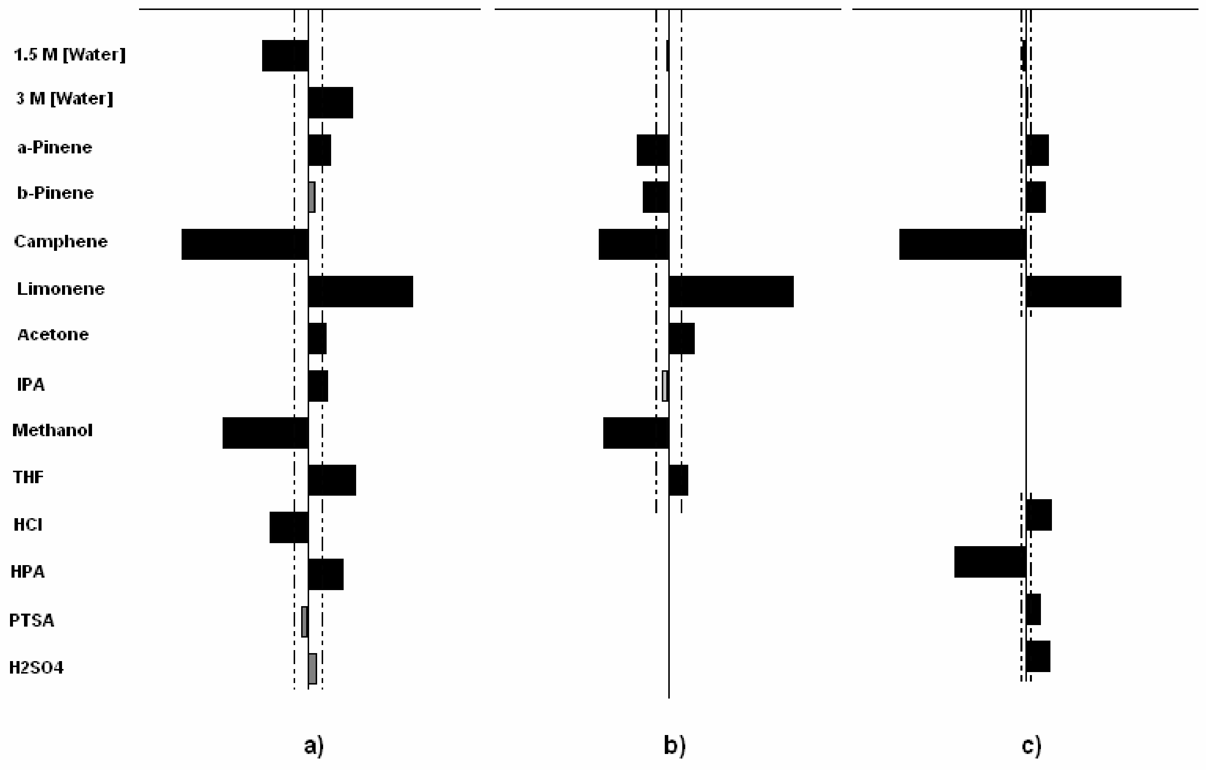


Figure 3.10 - Selectivity towards: a) *α*-terpineol (**12**), b) *α*-terpinyl chloride (**13**) HCl being the only acid present, c) *α*-terpinyl methyl ether (**14**) methanol being the only solvent applied.

Table 3.3 - Best results obtained for the terpenyl derivative products **12**, **13**, **14** and **15**.

Product	**Water (M)**	**Substrate**	**Solvent**	**Catalyst**	**Conv (%)**	**Select (%)**	**Yield (%)**
12	3	Limonene	IPA	HPA	73	55	43
12	3	Limonene	THF	HPA	63	56	38
12	3	α-Pinene	THF	H_2SO_4	99	45	45
13	1.5	Limonene	Acetone	HCl	43	54	26
14	1.5	Limonene	Methanol	PTSA	93	42	40
14	3	Limonene	Methanol	H_2SO_4	62	49	33
15	1.5	α-Pinene	Methanol	HPA	100	22	22
15	1.5	Limonene	Methanol	HPA	97	23	23

Conversion, selectivity and yields calculated according to GC analysis.

Isobornyl products

Camphene is the only substrate to produce isobornyl derivatives in relevant amounts. For the synthesis of isoborneol the water concentration seems to have only a small influence on selectivity; the preferred solvents are the aprotic acetone and THF and the preferred catalyst is HPA, followed by H_2SO_4. For the synthesis of isobornyl chloride a lower water concentration and THF as solvent seems to be preferential. In the case of isobornyl methyl ether the high excess of the strong nucleophile methanol obscures any influence of the water concentration effect and HPA appears to be, even if not markedly, the preferred catalyst. In Figure 3.11 these trends are summarised whereas in Table 3.4 the best results obtained in the screen are presented.

Table 3.4 - Best results obtained for the Isobornyl derivative products **7**, **8** and **10**.

Product	**Water (M)**	**Substrate**	**Solvent**	**Catalyst**	**Conv (%)**	**Select (%)**	**Yield (%)**
7	3	Camphene	THF	HPA	50	73	40
7	1.5	Camphene	Acetone	HPA	57	62	38
8	1.5	Camphene	THF	HCl	44	57	25
8	1.5	Camphene	Acetone	HCl	37	60	26
10	1.5	Camphene	Methanol	PTSA	86	81	73
10	1.5	Camphene	Methanol	HPA	97	91	89

Conversion, selectivity and yields calculated according to GC analysis.

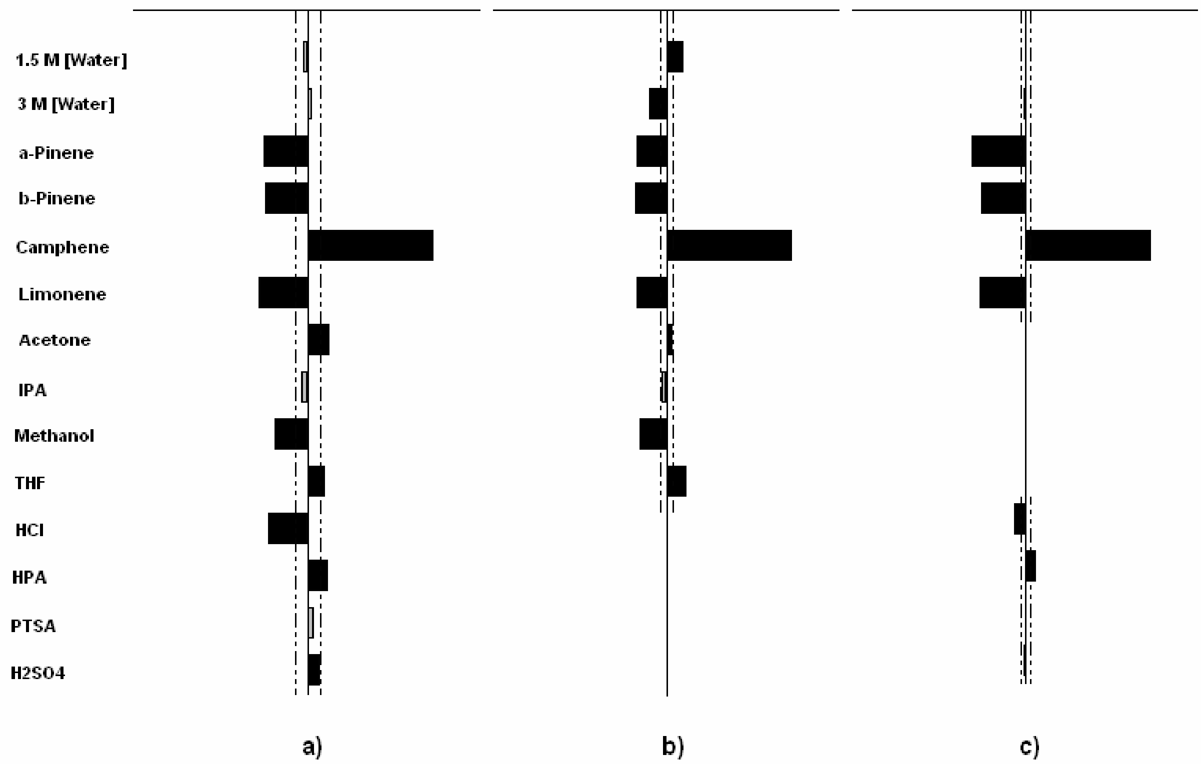


Figure 3.11 - Selectivity towards: a) Isoborneol, b) Isobornyl Chloride, HCl being the only acid present c) Isobornyl methyl ether, MeOH being the only solvent applied.

Bornyl products

Bornyl chloride is generally produced as the major reaction product when *α*- and *β*-pinene react in the presence of HCl. It seems that, in the presence of HCl, ring opening does not occur readily and reaction of these substrates proceeds via carbenium ions C3 and especially C1, in the latter case the product with endo configuration is enhanced. This phenomenon of HCl addition giving the endo-product was already reported in the work of Hanack in 1960 but its reason is even nowadays not fully understood [12, 14]. The fact that mainly the endo product (bornyl) is obtained in contrast to the more common exo configuration in other addition reaction, e.g. to isoborneol in hydrations, is probably due to a certain degree of concertedness and steric constraint in the attack of the nucleophile.

Table 3.5. Best results obtained for bornyl chloride **9**.

Product	Water (M)	Substrate	Solvent	Catalyst	Conv (%)	Select (%)	Yield (%)
9	1.5	α-Pinene	Acetone	HCl	88	23	20
9	3	α-Pinene	IPA	HCl	76	26	21
9	1.5	α-Pinene	IPA	HCl	83	28	24

Conversion, selectivity and yields calculated according to GC analysis.

As bornyl chloride is only formed from the less selective α- and β-pinene substrates and since, in the presence of water as competing nucleophile, products of carbenium ion **C2** are also obtained, the selectivities and yields of bornyl products are below 30% for both substrates. For the synthesis of bornyl chloride low concentrations of water are of course preferential and IPA seems to be the preferred solvent followed by acetone. These trends are summarised in Figure 3.12 and the best results are presented in Table 3.5.

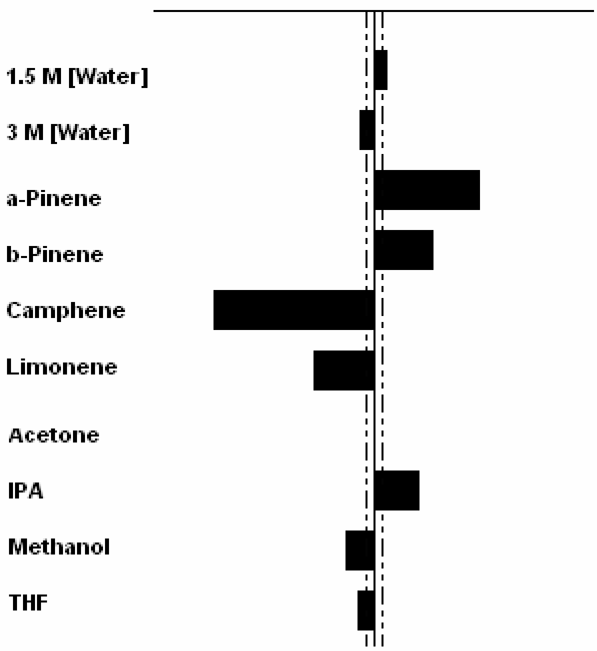


Figure 3.12 - Selectivity towards bornyl chloride. The bars directed to the right have a positive relative effect and those to the left a negative one. The dotted lines represent the 95% confidence interval calculated from the estimated experimental variance. Effects higher than this confidence are considered statistically significant and are represented in black.

3.3 - Conclusions

From this introductory investigation of the terpenyl parameter space we have obtained a great deal of insight in the main reactivity trends and we are able to develop guide-lines for further optimization: we have observed that the dominant factor determining the selectivity is the variety of carbenium ion rearrangement pathways usually undergone by a particular substrate. Thus in this case the highest selectivities are usually obtained with camphene and the lowest with the two pinene substrates. Activity is increased by using stronger acids, especially HPA, but, acid strength was not found to be a factor exerting a major influence on the product mixture. A special case is formed by HCl which, owing to the unique action of Cl^- as nucleophile, yields substantial amounts of bornyl chloride from *α*- and *β*-pinene via the C1, bornyl carbenium ion pathway, whereas, in other cases, products from the C2, p-menthyl carbenium ion, dominate. Besides the effect of solvents due to their nucleophilic properties, some differences in the final products obtained are found between aprotic acetone and THF and IPA. In the case of hydration products the effect of water concentration is significant but not generally pronounced.

The highest selectivities (91%) were obtained for the isobornyl methyl ether product, whereas isoborneol was the only alcohol derivative obtained with selectivities above 73%. On the basis of the knowledge obtained from this sort of explorative search new, more product-specific, optimizations can be designed to obtain the desired products.

HTE combined with Experimental Design proved to be a fast and fruitful strategy to explore and discover the underlining trends in a challenging chemical parameter space. The simultaneous study of several substrates that can share similar reactivity characteristics proved to be advantageous at this level of explorative information.

3.4 - Experimental

3.4.1 - General

The monoterpenes used here were of commercial grade. (1S)-(-)-α-Pinene (98%), (1S)-(-)-β-pinene (98%), myrcene (90%) and (+)-limonene (97%) from Acros, and, (+)-3-carene (90%) and camphene (85%) from Aldrich.

For the high-throughput set-up, an automated workstation was used consisting of a Hamilton Micro Lab Duo liquid-transfer system and a purpose-designed working platform, coupled with a temperature-controlled Zinsser Analytic Desyre mixer. On this platform a throughput of 24 reactions/day was obtained. The reactor rack comprised 24 x 1.5 ml HPLC reaction vessels which could be heated and agitated simultaneously.

^{1}H- and ^{13}C-NMR spectra were recorded on a Varian VXR-400S (400 and 100 MHz, respectively) or Varian Unity Inova 300 (300 and 75 MHz, respectively), instruments. GC-MS was measured by means of a VG 250 SE instrument equipped with a CP Sil 8 CB column of 25 m × 0.25 mm and 0.4 μm DF. A Varian Star 3600 GC, equipped with a CP Sil 5CB column with 50 m × 0.55 mm and 1 μm DF was used to determine the conversions and yields in the crude reaction mixtures. Column chromatography was carried out with silica gel packing of 0.060-0.200 mm, pore diameter ca. 6 nm and with mixtures of petroleum ether (PE) and ethyl acetate (EtOAc) as solvent. TLC was performed on 0.20 mm silica gel plates.

3.4.2 - General Procedure A: Experimental designs and data analysis

For the first and second screenings a D-Optimal Design formed the basis of the experimental plan proposed for studying the major effects in the first screening, and the major and interaction effects in the second screening, with the reaction parameters presented in Table 3.1 and Table 3.2, respectively. In the first screening 70 experiments out of a possible 420 candidates were selected using Design Expert 6.0.3 software, while in the second screening 100 experiments out of 128 candidates were selected, including 4% replicate experiments for random error analysis. Overall, a design containing three categorical factors with multiple levels was employed: catalyst type, substrate type and solvent type. Water concentration was the only continuous factor studied. Other reaction conditions such as temperature (50 °C), catalyst concentration (0.5 N), substrate concentration (1 M), and mixing speed (750 rpm) were kept constant for this exploratory search and so were not included in the design strategy. The GC data obtained was worked out to obtain the desired responses (conversion, yield, selectivity) using a purpose-made Visual Basic program. Statistical data treatment of the responses to obtain the parameter effects was realised using NemrodW 2000 and further representation of the results via the Spot Fire 6.0.0 software.

3.4.3 - General Procedure B: Screening in the HTE platform

The reactions of the planned design were performed, in a random order, in batches of 24 x 1.5 ml screw-top HPLC reaction vessels. In each reaction vessel the concentrations were: catalyst 0.5 N, substrate 1 M, and the internal standard (PCBTF, 4-chlorobenzotrifluoride) 0.5 M. The solid catalysts were weighed manually and the appropriate amount of water (1.5-3 M), catalyst solution and solvent (IPA, THF, methanol or acetone) were dispensed to the reactors using the automated workstation. The vessels were then closed and the reaction temperature regulated to 50 ºC. After approximately 10 minutes, the mixed solution of the substrates and internal standard was transferred to the vials, initiating the reaction. A special rack equipped with needles was inserted in the upper part of the vessels to avoid problems of over-pressure during the liquid transfer; after this transfer a normal rack without needles was used. After allowing the reaction to proceed for 18 h, 300 μl of a saturated solution of $NaHCO_3/NaCO_3$, pH ≅ 7, was added to quench the reaction, the temperature cooled down, and the reaction rack mixed until no more release of CO_2 was noticed. 500 μl of 2:7 THF/toluene solution was added to dilute the reaction samples and extract the products by forming a two-layer system. The samples were centrifuged and the upper layer analysed by GC.

3.4.4 - General Procedure C: Scale-up of selected screening reactions

For purposes of product identification and verification of the screening results, some reactions were scaled-up to 30 ml reaction volume. The desired products were then purified by column chromatography using an ethyl acetate/hexane solvent system (ratio adjusted according to the desired product separation). The collected fractions were evaporated and analysed by GC-MS and NMR for identification of the resulting reaction products.

3.5 - References

[1] I. E. Maxwell, P. van der Brink, R. S. Downing, A.H. Sijpkes, S. Gomez and T. Maschmeyer, *Topics. Catal*., **2003**, 24, 125.

[2] P.P. Pescarmona, J. C. Van der Waal, I. E. Maxwell and T. Maschmeyer, *Catal. Lett*., **1999**, 63, 1.

[3] R. Carlson, J. Carlson and A. Grennberg, *J. Chemom*., **2001**, 15, 455.

[4] L. Veum, S. R. M. Pereira, J. C. van der Waal and U. Hanefeld, *Eur. J. Org. Chem*., **2006**, 7, 1664.

[5] D. B. Hibbert, *Chemom. Intell. Lab. Syst*., **1993**, 19, 277.

[6] D. Wolf, O. V. Buyevskaya and M. Baerns, *Appl. Catal,, A*, **2000**, 200, 63.

[7] S. R. M. Pereira; F. Clerk, D. Farrusseng, J. C. van der Waal, C. Mirodatos and T. Maschmeyer, *QSAR Comb. Sci.,* **2005**, 24, 45.

[8] A. S. McLeod, and L. F. Gladden, *J. Chem. Inf. Comput. Sci*., ***2000***, 40, 981.

[9] A. Holzwarth, P. Denton, H. Zanthoff and C. Mirodatos, *Catal. Today*, **2001**, 67, 309.

[10] R. Carlson*, Design and Optimization in Organic Synthesis, Data Handling in Science and Technology,* vol. 8, Elsevier, Amsterdam, **1992**.

[11] T. Norin, *Studies on the Chemistry of Terpenes*, Dissertation, Stockholm, **1964**.

[12] A. A. Newman, Ed., *Chemistry of Terpenes and Terpenoids*, Academic Press, London and New York, **1972**.

[13] C. M. Williams and D. Whittaker, *J. Chem. Soc., B*, **1971**, 668.

[14] M. Hanack, *Chem. Ber*., **1960**, 93, 844.

[15] D. H. Sheffield, Brunswick, Hercules Powder Company, US Patent Application. 2178349, **1939**.

[16] N. A. Milas, Belmont, Union Bay State Chemical, US Patent Application. 2467330, **1949**.

[17] J. C. van der Waal, H. van Bekkum and J. M. Vital, *J. Mol. Catal., A*, **1996**, 105, 185.

4

CATALYTIC HYDROGENATION OF CYANOHYDRIN ESTERS AS A NOVEL APPROACH TO *N*-ACYLATED *β*-AMINO ALCOHOLS: REACTION OPTIMIZATION EMPLOYING A DESIGN OF EXPERIMENT APPROACH

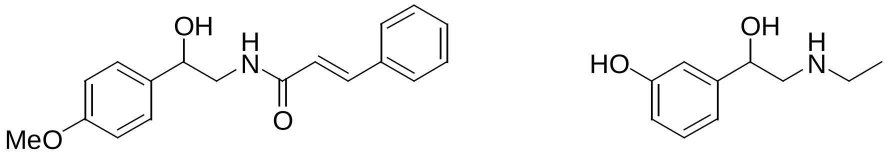
Abstract

The catalytic hydrogenation of acylated cyanohydrins followed by an intra-molecular migration of the acyl group to yield pharmaceutically interesting *N*-acyl *β*-amino alcohols, is shown to be a successful one-pot preparation method. The combination of a multistep DoE approach and High-Throughput methodology proved to be an effective strategy for the optimization of the reaction. With the favoured catalyst/solvent combination nickel-on-alumina in dioxane, both hydrogenation and acyl-group migration proceeded smoothly, giving the *N*-acyl *β*-amino alcohols in yields of up to 90 % for aliphatic, and up to 50 % for benzylic substrates, the latter being more prone to side reactions. When enantiopure cyanohydrin esters were used, no racemisation was found to occur at the chiral centre of an aliphatic molecule, though a minor decrease in ee was observed for a benzylic substrate.

The contents of this chapter have been published in:
L. Veum, S. R. M. Pereira, J. C. van der Waal, U. Hanefeld, *Eur. J. Org. Chem*, **2006**, 7, 1664.

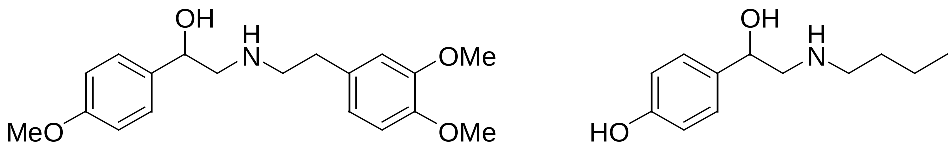
4.1 - Introduction

N-Acylated *β*-aminoalcohols, such as aegeline (Figure 4.1), occur in nature and can readily be converted into *β*-*sec*-amino alcohols, an important class of compounds in the pharmaceutical and agrochemical industries. Some representative examples of the numerous biologically active *β*-*sec*-amino alcohols are etilefrine, bamethane and denopamine (Figure 4.1). An established route to the *N*-acyl *β*-amino alcohols is the reduction of the free cyanohydrin, followed by acylation of the amino group [1-3]. The reduction is usually performed using stoichiometric amounts of either $LiAlH_4$ or BH_3, but it can also be achieved by catalytic hydrogenation under strongly acidic conditions [4-8]. If enantiopure substrates are used, the stereocentre remains intact during these reduction reactions. Given the low atom efficiency of aluminium and boron hydride reductions and the strongly acidic conditions required for the catalytic hydrogenations, a different approach has been investigated. The overall aim of this investigation was to integrate the reduction and acylation steps in a one-pot procedure under mild conditions.



Aegeline

Etilefrine

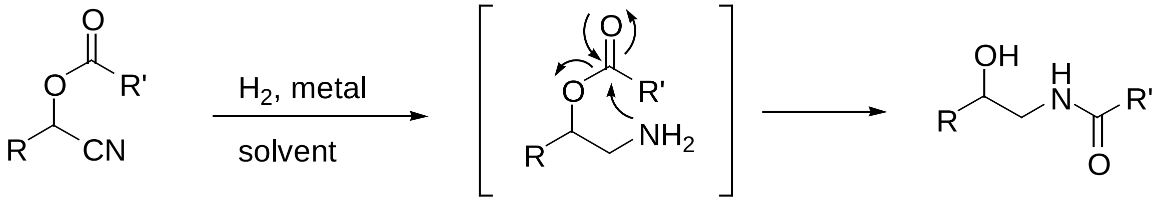


Denopamine

Bamethane

Figure 4.1 - *N*-Acyl-*β*-amino alcohols and *β*-*sec*-amino alcohols showing high biological activity.

The unprotected cyanohydrins, that are commonly used as starting materials, are relatively unstable and racemise easily. In contrast to this, cyanohydrin esters are stable and do not racemise. Moreover they are readily prepared, both in their racemic [9] and enantiopure forms [10-15]. In addition it was realised that the acyl group of the protected cyanohydrins is a potential intra-molecular acyl-donor (see Scheme 4.1). Following the catalytic hydrogenation of the nitrile group the newly formed amine, as a strong nucleophile, can immediately react with the neighbouring acyl group via a five-membered transition state to yield the *N*-acyl *β*-amino alcohol. This type of intramolecular acyl migration has previously been described in the $NaBH_3(OCOCF_3)$ reduction of an acylated cyanohydrin to yield denopamine [16], suggesting that it should proceed equally well following catalytic hydrogenation of the nitrile.



1 **2**

Scheme 4.1 - The hydrogenation of acylated cyanohydrins with subsequent acyl-migration.

Earlier reports of the catalytic hydrogenation of acylated cyanohydrins, in particular of mandelonitrile esters, describe the application of Pd/C or PtO_2 under strongly acidic conditions [17-18] The primary product obtained was not the *N*-acyl *β*-amino alcohol but *β*-phenylethylamine, owing to the facile hydrogenation of the benzylic C-O bond over platinum or palladium catalysts in acidic medium. In this case the amine was the desired product [17]. In the present work the objective is to maximise the yield of the *N*-acylated-*β*-amino alcohol and the reductive cleavage of the benzylic C-O bond needs to be avoided. This investigation of a selective, catalytic route for the direct conversion of acylated cyanohydrins (**1**) into *N*-acylated-*β*-amino alcohols (**2**) employed High-Throughput methods for the screening of catalysts, solvents and reaction conditions.

The large number of parameters to be investigated suggested a Design of Experiments (DoE) approach. DoE methodologies (Chapter 2) [19-21] are superior to traditional methods involving the consecutive optimization of the various parameters; they make it possible to maximise the amount of information that can be obtained from the results, while minimising the number of experiments, and increase the possibility of establishing the true optimum within the search space. The experiments to be performed may be chosen in order to cover the whole search space as efficiently as possible. The size of the design (selected number of reactions) depends on the kind of information that is desired. In the present case a strategy of three sequential small designs was adopted. This enables the information obtained from the first to be used to improve the subsequent designs [22]. Preliminary screening designs (typically less than 25% of the possible number of reactions) are sufficient to distinguish between significant and insignificant parameters and are therefore well suited at the early stages of the research effort to reduce the search space. By continuous refinement of the conditions in subsequent optimization designs, the most influential parameters can then be studied with higher detail, and further optimization achieved. Since the number of parameters to investigate is usually reduced, the number of experiments per parameter can be increased. In this way, more information on the main effects and especially the interactions of the parameters can be obtained. Such a sequence of DoE’s is generally a better strategy than one large one, because the information obtained from one design is used to improve the following (Chapter 2).

4.2 - Results and discussion

As shown by Hartung [17], the hydrogenation of benzylic cyanohydrin acetates easily yields products such as β-phenylethylamines by reductive cleavage of the benzylic C-O bond. In aliphatic substrates, on the other hand, the C-O bond is more stable and resistant to cleavage even under drastic conditions. Different conditions are likely to be required for the selective hydrogenation of aliphatic and benzylic cyanohydrins and it was therefore chosen to optimize the reactions for mandelonitrile acetate (**1a**), representative for the benzylic substrates, and heptanonitrile acetate (**1c)** (see Scheme 4.2), representative for the aliphatic substrates, separately.

4.2.1 - Initial screening

The use of a DoE strategy requires as the first step the compilation of all potentially important parameters, based on previous experience, the literature or chemical intuition. The initial search space should be broad enough to assure that the optimal settings of the reaction are included but not too big to be still experimentally feasible. In the hydrogenation of nitriles the main factor influencing the reaction rate and the product distribution is the metal of the hydrogenation catalyst. Most commonly, Raney nickel, Raney cobalt, Pd/C, Pt/C, Ru/C, and Rh/C are used [23]. The same metals but supported on SiO_2 and Al_2O_3 are also often reported [23]. Normally Rh, Pd and Pt tend to give more secondary and tertiary amines than Co, Ni and Ru. As the migration of the acyl group might suppress the formation of secondary and tertiary amines, these metals were still included in this investigation. For the initial screening Ni, Pd, Rh, Pt and Ru, on carbon and Al_2O_3 as carriers were selected as potential catalysts.

The solvent forms a second important parameter. The most commonly used solvents for the hydrogenation of nitriles are protic solvents such as methanol and ethanol. However, since the envisaged reaction involves a migration of the acyl-group, solvents with a broader range of properties were selected: 2-propanol, a protic but less polar solvent than methanol; dioxane, an aprotic, polar ether; and toluene, a relatively apolar solvent.

It is known that the addition of ammonia and of water can change the distribution ratio of the products of nitrile hydrogenation [23]. Ammonia is a commonly used additive, favouring the formation of primary amines, though in the present case reaction with the ester group is a possible side reaction. Reports on the effect of water are conflicting; several cases have been reported where water is added to promote the formation of both primary and secondary amines [24-25] but it was also claimed that water does not change the product distribution but instead increases the reaction rate [26]. The effects of both these additives were studied in the initial screening.

The parameter space for the initial screen is summarised in Table 4.1. The reaction temperature was varied over two levels. The small reactors of the High-Throughput unit (Appendix A.3) did not permit independent variation in pressure, which was kept constant at 20 bar H_2. Using a D-Optimal algorithm a selection of 24 reactions out of the total of 320 possible combinations was made for each of the two substrates [27]. This minimal design is sufficient to obtain information about the main effects of each parameter and a more detailed subsequent study would then allow for further optimization. Acidic conditions were not included since any formation of the amino salts would prevent the intramolecular migration of the acyl group. Furthermore, in contrast to the free cyanohydrins, the cyanohydrin esters (**1**) are more stable towards possible basic side products like the secondary amine.

Table 4.1 - The parameter space to be investigated for substrates **1a** and **1c**.

Temperature (°C)	Reaction time (h)	Support	Solvent	Additive	Metal
90	3	Alumina	2-propanol	No additive	Ni
120	24	Carbon	Toluene	H_2O	Pd
			Dioxane	NH_3	Pt
			Methanol	$H_2O + NH_3$	Rh
					Ru

After executing the 2×24 reactions, *N*-acyl *β*-amino alcohols **2a** and **2c** were identified among the products in two of the experiments for each substrate, showing the hydrogenation indeed to have been followed by intra-molecular acyl-migration in a one-pot procedure. The conditions for the four successful reactions are given in Table 4.2. This result already shows the advantage of using DoE with successive small designs as an approach towards the optimization of a new reaction. It made it possible to investigate a large parameter space and identify the region of interest for further exploration, even though only 8 % of the possible number of reactions was executed. If a single large DoE design had been chosen, a good number of unnecessary reactions would have been performed.

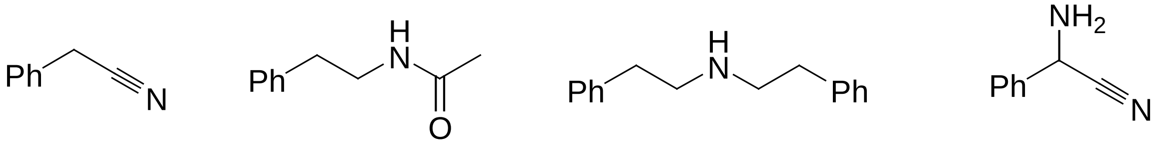
When using such a small design it is important to realise that each result is extremely influential for the calculation of the main effects of the parameters. These calculations will become increasingly inaccurate with a growing number of “zero-yield” reactions or failed experiments. In this case the number of reactions leading to the desired product is so low, i.e. 2 per design, that a statistical evaluation of the effect of the parameters on the yield would not be meaningful. The results do, however, enable the identification of unfavourable factors and their exclusion from the next screening phase.

Table 4.2 - Conditions for the successful hydrogenation in the initial screening.

Substrate	Metal	Support	Temp [°C]	Solvent	Additive	Conversion of 1 [%]	Yield[a] 2 [%]
1a	Ni	C	120	Dioxane	NH_3	100	33
1a	Rh	Al_2O_3	120	Dioxane	NH_3+H_2O	100	24
1c	Ni	Al_2O_3	120	Dioxane	H_2O	100	65
1c	Rh	C	120	2-propanol	H_2O	100	48

[a] According to GC

The reactions in Table 4.2 were all run for 24 h at 120 ºC, with either dioxane or 2-propanol as the solvent, and ammonia or water as the additive. The successful metals were Ni and Rh, supported on either carbon or alumina. From the reactions which did not yield the desired product, the following trends could be observed: the reactions using Ru or toluene gave low conversions, while the reactions performed in MeOH in all cases gave complete conversions, but, with a wide range of side-products. Since the intention of the first screening was to reduce the parameter space, none of the side-products of the reactions were isolated. However, GC-MS enabled the identification of several side-products (**3** to **6**).



3 **4** **5** **6**

Figure 4.2 - Identified side-products in the hydrogenation of **1a**.

The presence of **5** shows that the secondary amine is formed in some cases. Products **3** to **5**, in which the benzylic alcohol group has been removed, were particularly dominant when platinum was used as the catalyst, which is to be expected from the application of platinum catalysts for the cleavage of this type of bond. Equivalent by-products could also be identified in the case of the aliphatic substrate, though in much smaller amounts in accordance with the greater stability of the C-O bond. The reason that even the aliphatic C-O bond can be cleaved can be attributed to the stabilising effect of the nitrile group on the intermediate radical or carbenium ion formed during the cleavage.

In the successful reactions of **1a** (see Table 4.2) ammonia was present as additive, but analysis of the results did not show unambiguously that the presence of ammonia was essential. Since the formation of 6 indicates that ammonia also reacts with the substrate it was chosen, in order to avoid this side-reaction, to further optimize the conditions in the absence of ammonia. Samples taken after 3 hours showed only low conversions and there was no formation of **2a** or **2b** in any other reaction than under the conditions reported in Table 4.2. The long reaction time could be due to an initial activation period for the catalyst but the study of this was deferred to a later stage and a reaction time of 24 h was maintained for the second design.

4.2.2 - Second Screening

Based on the results of the first screening design, the second design was conducted with the parameters indicated in Table 4.3. In order to further study the effect of the carrier silica was included in this design. Since the parameter space was now considerably reduced, a Full Factorial design (Chapter 2), i.e. 36 combinations, became feasible for each substrate. All the reactions were performed at 120 °C and 20 bar H_2, with a reaction time of 24 hours.

Table 4.3 - Conditions and parameters in the second screening round.

Additive	Solvent	Metal	Support
No additive	2-propanol	Ni	Alumina
H_2O	Dioxane	Rh	Carbon[a]
		Ru	Silica

[a] In the case of nickel, Raney-nickel was used instead of nickel on carbon.

The results of this screening are presented in Figure 4.3 and Figure 4.4. The conversions of **1a** and **1c** were in most cases 100 %, except for the reactions where Ru-alumina and Ru-silica was used. For those two catalysts no conversion was observed. In sharp contrast to the first screening, where only a few reactions yielded the *N*-acyl *β*-amino alcohols, all active catalysts now yielded the desired products. A statistical evaluation of the results from Figure 4.3 and Figure 4.4, with respect to the main effects of the parameters and their interactions, is presented in Figure 4.5 and Figure 4.6.

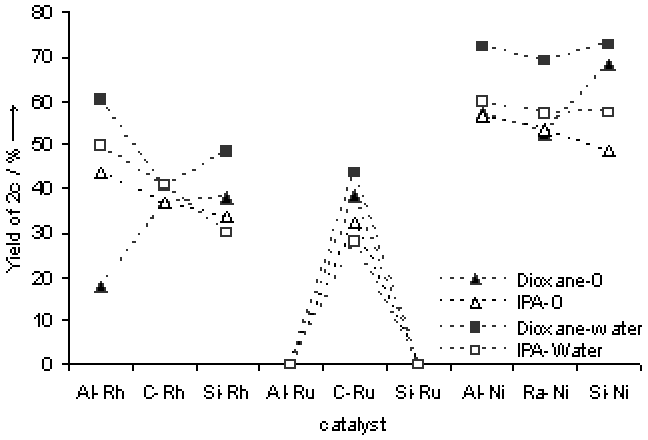


Figure 4.3 - Graphical representation of the results of the second screening with substrate **1c.** Note: dashed lines are intended to group individual point results at similar reaction conditions.

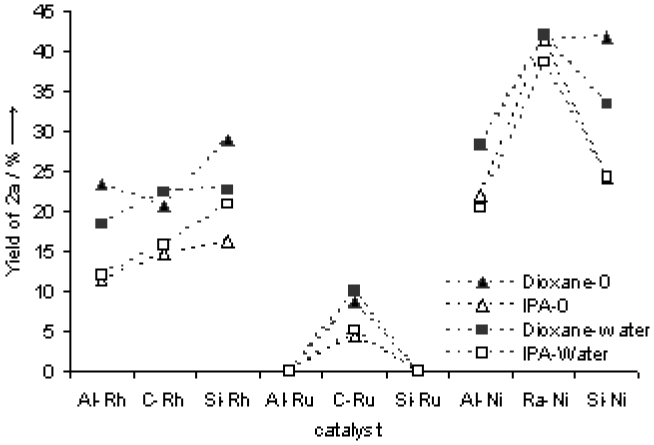


Figure 4.4 - Graphical representation of the results of the second screening with substrate **1a**. Note: dashed lines are intended to group individual point results at similar reaction conditions.

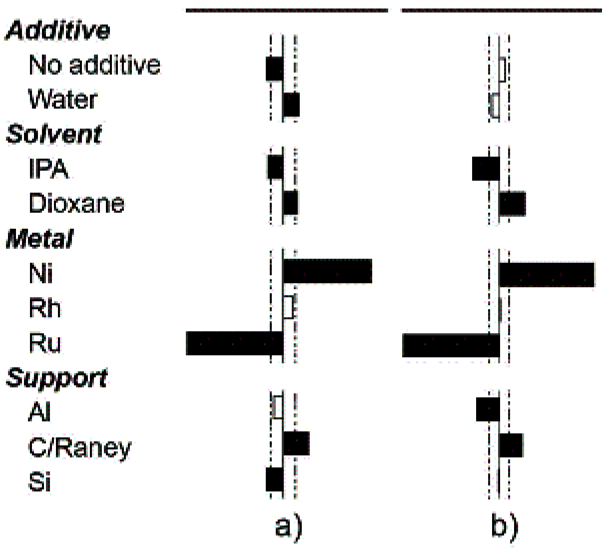


Figure 4.5 - Effect of the main parameters in the second screening: a) aliphatic substrate **1c**; b) benzylic substrate **1a**. The lengths of the bars show the relative influence of the main parameters on the yield of **2a** and **2c**. The bars directed to the right have a positive relative effect and to the left a negative. The dotted lines represent the 95% confidence interval calculated from the estimated experimental variance. Effects higher than this confidence interval are considered significant and are represented in black (‘Al’ = alumina, ‘Si’ = silica).

When evaluating the main parameters in Figure 4.5, many similarities for the two different substrates are noted. The most important parameter in both cases is the type of metal, with nickel being the best followed by rhodium. For ruthenium, the poor results from the initial screening are confirmed. In the initial screening, the successful reactions included those in which water was used as an additive. A small but statistically significant positive effect of water is indeed observed in the case of the aliphatic substrate, though not with the benzylic substrate. Although the absolute difference between the two solvents is small, dioxane is statistically significantly better than 2-propanol for both substrates. With respect to the effects of the catalyst carriers, the apparent superiority of carbon is based solely on the fact that ruthenium gives the product only in combination with carbon while, in addition, nickel on carbon was not available and Raney-Ni was used instead, so that no definite conclusion on carrier effects can be drawn from this second design.

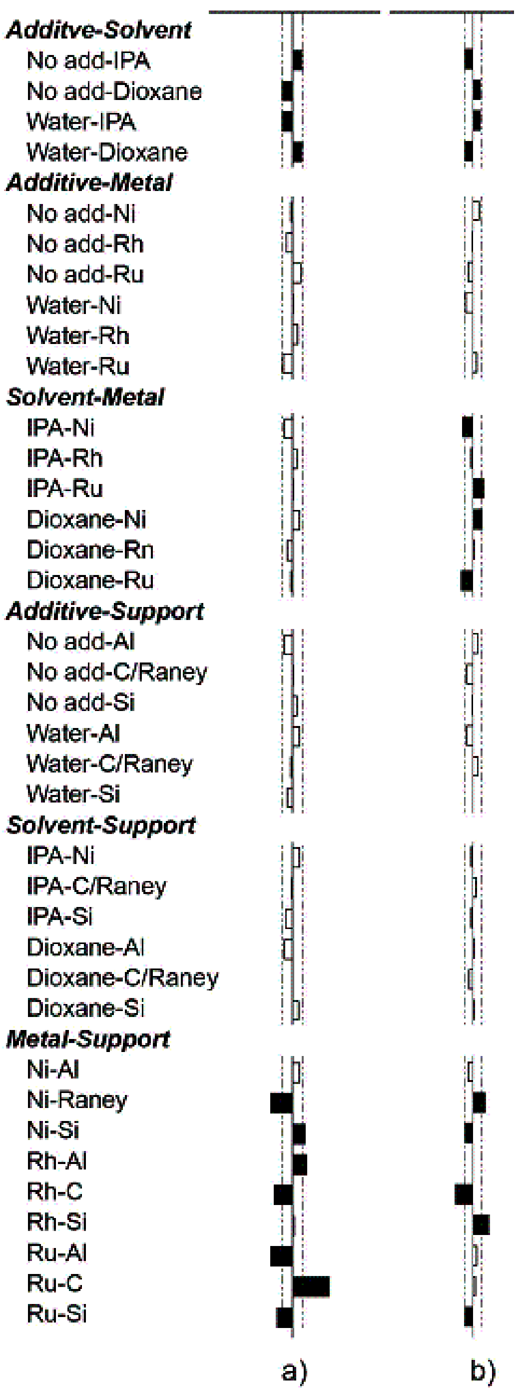


Figure 4.6 - Interaction effects between the parameters of the second screening: a) aliphatic substrate **1c**; b) benzylic substrate **1a**. The lengths of the bars show the relative influence of the interaction effects, between the different parameters, on the yield of **2a** and **2c**. The bars directed to the right have a positive relative effect and 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 interval are considered significant and are represented in black (‘Al’ = alumina, ‘Si’ = silica).

With respect to the 2-level interactions between the parameters presented in Figure 4.6 it was noticed that for both substrates there are significant additive/solvent and metal/support interactions, while in the case of the benzylic substrate a solvent/metal interaction also exists. However, the effects of these interactions are relatively small in comparison with the main effect of the metal itself. For the aliphatic substrates, nickel on silica, using dioxane as the solvent and water as additive, is the combination of choice, while for the benzylic substrate Raney-nickel is the indicated catalyst when used in dioxane without addition of water.

All the reactions from these two screenings were performed in the "Quick Catalyst Screening 96" platform (Appendix A.3). This equipment has a maximum pressure limit of 20 bar and no individual temperature control for the reactors. Further optimization regarding pressure and temperature was for that reason performed in a conventional autoclave.

In preparation for this, a test was conducted of whether the results with the nickel catalysts could be improved by activation with H_2 prior to the catalytic test. By activating the catalysts at 140 ˚C for 12 h at 40 bar H_2, nickel on alumina gave similar yields to those of Raney-nickel and it was found to be more practical to use this activated catalyst for further optimization. In the case of the aliphatic substrate the reaction time was reduced to two hours and for the benzylic substrate to three hours. Once again, dioxane proved to be slightly superior to 2-propanol. As a result of this it was chosen as the solvent to perform the third round of screening: the optimization of temperature and pressure using activated nickel on alumina in dioxane. In the case of the aliphatic substrate **1c** water was used as an additive.

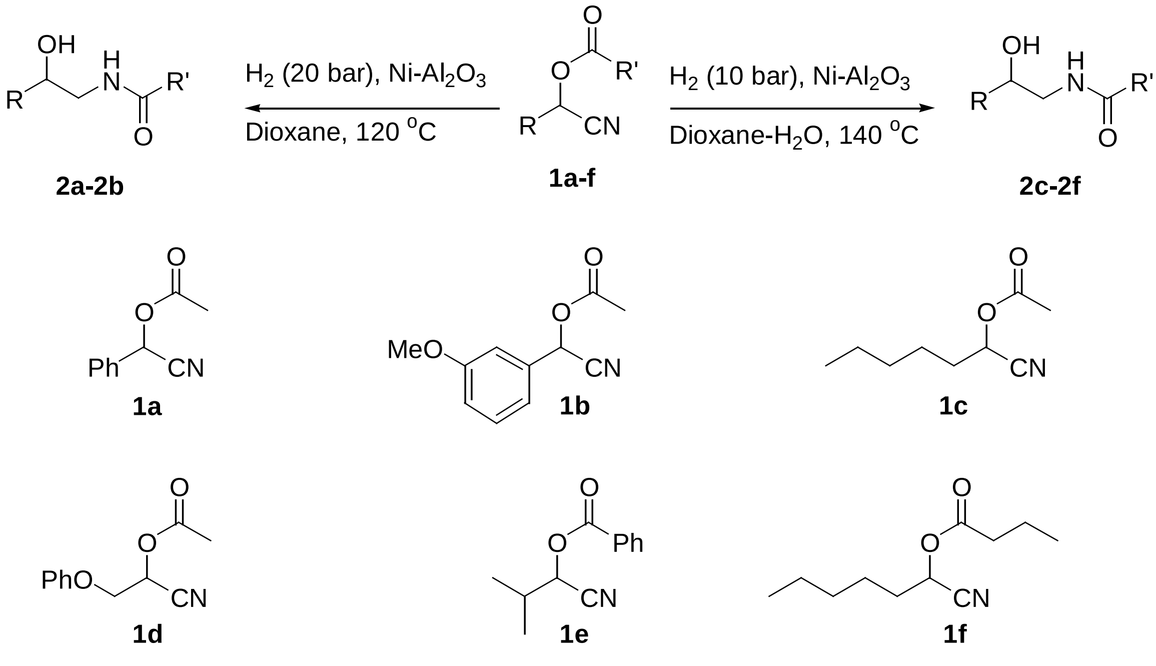
4.2.3 - Optimization design

A temperature range of 80 to 160 ˚C and a pressure range from 5 to 40 bar was tested. Only minor differences in the yields (± 7 % for the aliphatic, ± 5 % for the aromatic) were observed, except for reaction temperatures below 90 ˚C, where hardly any reaction occurred. Despite the small differences in the observed yields an optimum of 10 bar H_2 at 140 ˚C is found for the aliphatic substrate **1c** and 20 bar H_2 at 120 ˚C is found for the benzylic substrate **1a**.

4.2.4 - Other substrates

In order to establish the versatility of the reaction the optimized conditions were applied to a number of other acylated cyanohydrins: a substituted benzylic ester (**1b**), an aliphatic substrate with an aromatic side chain (**1d**), and aliphatic substrates with a variety of acyl groups (**1e**, **1f**) (Scheme 4.2).

All these substrates were successfully hydrogenated to yield the desired *N*-acyl *β*-amino alcohols **2a-f**. The conversion of the substrates was in all cases 100 %. The benzylic substrates **1a** and **1b** gave more side products than the aliphatic **1c-f**. This difference between the substrates is in accordance with the less stable benzylic C-O bond. In the case of **1c** and **1f**, the products were isolated by crystallisation from the reaction mixtures.



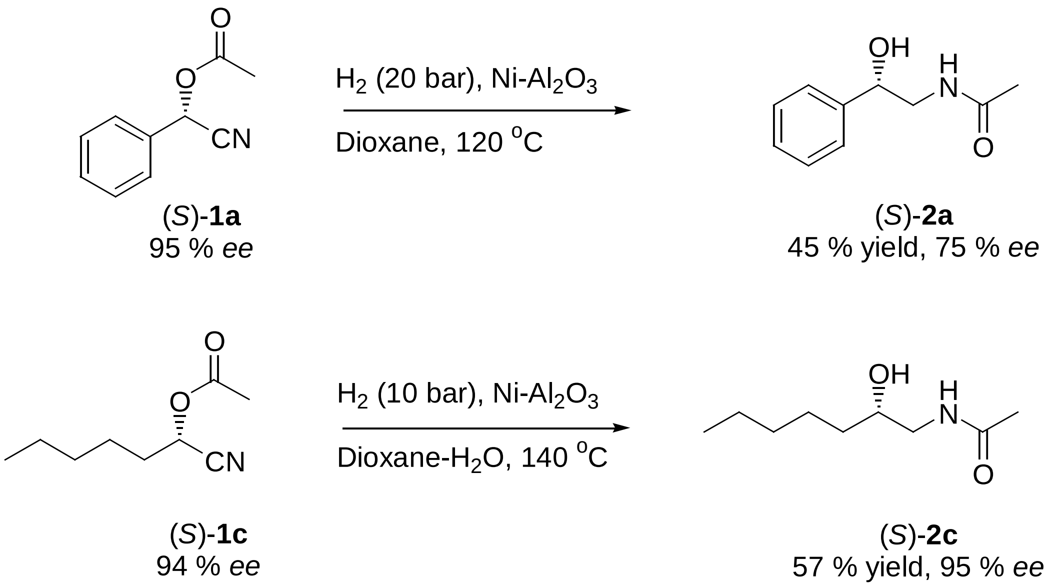
Scheme 4.2 - The catalytic hydrogenation of cyanohydrin esters. (**1a**, R = -Ph, R’= $-CH_3$; **1b**, R = $-OCH_3$, R’ = $-CH_3$; **1c**, R = $-C_5H_{11}$, R’= $-CH_3$; **1d**, R = $-CH_2OPh$, R’ = $-CH_3$, **1e**, R = $-CH(CH_3)_2$, R’= -Ph; **1f**, R = $-C_5H_{11}$, R’ = $-C_2H_5$)

Table 4.4 - Conversion and yield from the hydrogenation of acylated cyanohydrins **1a-f**.

Substrate	Conversion of 1 [%]	NMR-Yield of 2 [%]	Isolated yields of 2 [%]
1a	100	n.d	49[a]
1b	100	n.d	50[a]
1c	100	74	57[b]
1d	100	91	72[a]
1e	100	~75	58[a]
1f	100	83	30[b]

[a] Isolated by column chromatography. [b] Isolated by recrystallisation from ethyl acetate, not optimized.

The hydrogenation was also performed on the optically active substrates (*S*)-**1a** (95 % ee) and (*S*)-1c (94 % ee). As expected, the chiral centre of (*S*)-**1c** was found to remain unchanged during both the hydrogenation and the intra-molecular migration. This was not the case with (*S*)-**1a**. The isolated (*S*)-**2a** had an ee of only 75 % (see Scheme 4.3).



(*S*)-**1a**
95 % *ee*

(*S*)-**2a**
45 % yield, 75 % *ee*

(*S*)-**1c**
94 % *ee*

(*S*)-**2c**
57 % yield, 95 % *ee*

Scheme 4.3 - Catalytic hydrogenation of enantiopure acylated cyanohydrins

This decrease in ee might be explained by a base-catalysed racemisation of the substrate; the base being either ammonia released in the formation of the secondary amine side-product, or the secondary amine itself.

4.3 - Conclusions

A multi-step DoE approach proved an efficient method for the optimization of the reaction. From more than 2000 possible combinations of the parameters requiring to be studied, it proved possible to effect the optimization using only 70 experiments for each substrate. This shows the great advantage of the DoE approach towards the optimization of a new reaction, enabling a large parameter space to be investigated and the most interesting range within the parameter space to be identified.

The catalytic hydrogenation of acylated cyanohydrins (**1**) with subsequent intra-molecular migration of the acyl group constitutes a valuable one-pot route to the pharmaceutically important *N*-acyl *β*-amino alcohols (**2**). The nickel-on-alumina catalyst in dioxane as solvent proved to be preferable to the traditional catalysts (Pd/C and PtO_2) that are used under acidic conditions [8]; both the hydrogenation and the migration proceeded smoothly and the desired product could be obtained in yields of up to 90 % for the aliphatic substrates and up to 50 % for the more sensitive benzylic substrates. The application to a range of aliphatic and aromatic substrates with different acyl groups was demonstrated. When enantiopure substrates are employed the stereocentre remains unaltered for aliphatic substrates and only a small amount of racemisation is observed for benzylic substrates. Given the straightforward access to the (chiral) starting materials and the mild, catalytic reaction conditions this one-pot sequence represents a significant step forward.

4.4 - Experimental

4.4.1 - General

^{1}H and ^{13}C-NMR spectra were recorded on a Varian VXR-400S (400 and 100, MHz, respectively) or a Varian Unity Inova 300 (300 MHz and 75, MHz, respectively), instrument. Chemical shifts are expressed in parts per million (δ) relative to tetramethylsilane. Abbreviations are as follows: s (singlet), d (doublet), t (triplet), q (quadriplet) and m (multiplet).

Mass spectra were determined on a VG 70 SE spectrometer operating at 70 eV. GC-MS was measured by means of a VG 250 SE instrument equipped with a CP Sil 8 CB column of 25m × 0.25mm and 0.4 μm DF. A Varian Star 3600 - GC equipped with a CP Sil 5CB column with 50m × 0.55mm and 1 μm DF, was used to determine the conversions in the crude reaction mixtures. Optical rotations were obtained using a Perkin-Elmer 241 polarimeter. Melting points are uncorrected. Column chromatography was carried out with silica gel packing of 0.060-0.200 mm, pore diameter ca. 6 nm and with mixtures of petroleum ether (PE), methanol (MeOH) and ethyl acetate (EtOAc) as solvent. TLC was performed on 0.20 mm silica gel.

The nickel catalysts were all activated at 140 ˚C for 12 hours at 40 bar H_2 before use in the general procedures B, C and D described below. All other catalysts, and the solvents employed, were used as received from commercial sources. For all the supported catalysts the metal loading was 5%, except for Rh-Silica (1%), Ni-Alumina (50%) and Ni-Silica (66%). Racemic [9] and enantiopure cyanohydrin acetates [11,28] were synthesised according to literature procedures. The optical purity of **2a** was determined by HPLC employing a Waters 510 pump, a 4.6 × 250 mm 10 μ Chiracel OJ column and a Waters 486 UV detector. The eluant was a mixture of hexane and 2-propanol (90:10) with a flow of 0.8 ml min^{-1}. The optical purity of **1a**, **1c** and **2c** was determined by chiral GC using a Shimadzu Gas Chromatograph GC-17A equipped with a *β*-cyclodextrin column (CP-Chirasil-Dex CB 25m × 0.25mm). A Shimadzu Auto-injector AOC-20i and FID detector were employed, and He with a linear gas velocity of 75 cm/s formed the carrier gas.

The Avantium "Quick Catalyst Screening 96" platform was used to perform the reactions of the first and second experimental designs. This equipment has a maximum pressure limit of 20 bar and the temperature is controlled for all reactors simultaneously. Otherwise, a 100 ml Parr autoclave was used. The elemental analysis was performed on a Elementar Vario EL III analyser.

4.4.2 - General procedure A: Screening in the Avantium "Quick Catalyst Screening 96" platform

The various supported metal catalyst (5 mg) were weighed into the autoclaves and added to a 1.7 M solution of the substrate in the desired solvent (1.5 ml). When water was used as an additive, 10 μl was added. In the case ammonia was used as additive, the concentration of ammonia in the reaction mixture was 0.5 M. After stirring the reaction at 90 or 120 ºC and 20 bar H_2 for 3 or 24 h, the reaction mixture was centrifuged and the supernatant liquid analysed by GC and GC-MS.

4.4.3 - General procedure B: Screening for temperature and pressure in the Parr autoclave

The pre-activated 50% Ni on alumina (100 mg) was added to a solution of **1a** or **1c** (5.7 mmol) in dioxane (30 ml). In the case of **1c**, water (0.2 ml) was also added. After stirring the reaction at 80, 100, 120, 140, or 160 ºC, and 5, 10, 20, 30 or 40 bar H_2 for 2 hours, the reaction mixture was filtered. The filtrate was analysed by GC.

4.4.4 - General procedure C. Reductions in the Parr autoclave with optimized conditions for substrates prepared from aromatic aldehydes

Activated 50 % Ni on alumina (100 mg) was added to a solution of the substrate (5.7 mmol) in dioxane (30 ml). After stirring the reaction at 120 ºC and 20 bar H_2, the reaction mixture was filtered. A sample of 2 ml was taken from the filtrate and the solvents from this sample were removed under vacuum. The sample was then analysed by ^{1}H-NMR. The combined filtrate and NMR-sample was then evaporated to dryness to yield the oil or solid products.

***N*-(2-Hydroxy-2-phenylethyl) acetamide 2a**: The solid prepared from **1a** according to general procedure C was purified by column chromatography (silica, EtOAc/MeOH, 95:5, R_f = 0.27). Yield of (*S*)-**2a**: 503 mg (49 %) as a white solid; m.p. 125-126 ºC;

^{1}H-NMR (300 MHz, CDCl$_3$, 25 ºC, TMS): δ = 2.01 (s, 3H, C*H*$_3$-C=O), 3.32 (ddd, *J* = 5.0, 7.9, 14.1 Hz, 1H, C*H*$_2$-N), 3.70 (ddd, *J* = 3.3, 7.0, 14.1, 1H, C*H*$_2$-N), 4.85 (dd, *J* = 3.3, 7.9, 14.1 Hz, 1H, C*H*-O), 5.92 (s, 1H, N*H*), 7.28-7.38 (m, 5H, aromatic); ^{13}C-NMR (75 MHz, CDCl$_3$, 25 ºC, TMS): δ = 23.1 (*C*H$_3$), 47.6 (*C*H$_2$-N), 73.6 (*C*H-O), 125.8, 128.8, 128.5 and 141.8 (aromatic), 171.6 (*C*=O); IR (KBr): ν = 3300, 3080, 1648, 1547, 1295 cm^{-1}; MS (70 eV, EI): *m/z* (%): 179 (1) [M$^+$], 161 (3) [M$^+$-H$_2$O], 120 (14), 107 (21), 79 (31), 77 (31), 73 (100); elemental analysis calculated (%) for $C_{10}H_{13}NO_2$ (179.22): C 67.02, H 7.31, N 7.82; found: C 67.00, H 7.49, N 7.81.

(*S*)-*N*-(2-Hydroxy-2-phenylethyl) acetamide (*S*)-2a: The solid prepared from **(*S*)-1a** (95 % ee) according to general procedure C was purified by column chromatography (silica, EtOAc/MeOH, 95:5, R$_f$ = 0.27). Yield of (*S*)-**2a**: 0.454 mg (45.4 %) as a white solid; *ee* = 75 %, $[\alpha]_D^{20}$ = + 8.1 (*c* = 1.0 in MeOH); other spectroscopic data as for **2a**.

***N*-[2-Hydroxy-2-(3-methoxyphenyl)ethyl] acetamide 2b**: The solid prepared from *rac*-**1b** according to general procedure C was purified by column chromatography (silica, EtOAc/MeOH, 95:5, R$_f$ = 0.25). Yield of **2b**: 0.570 mg (57 %) as a white solid; m.p. 123-124 ºC; ^{1}H-NMR (300 MHz, CD$_3$OD, 25 ºC, TMS): δ = 1.93 (s, 3H, C*H*$_3$-C=O), 3.28 (dd, *J* = 7.9, 13.7 Hz, 1H, C*H*$_2$-N), 3.45 (dd, *J* = 4.6, 13.5 Hz, 1H, C*H*$_2$-N), 3.78 (s, 3H, OC*H*$_3$), 4.71 (dd, *J* = 4.6, 7.9 Hz, 1H, C*H*-O), 6.81 (ddd, J = 0.9, 2.6, 8.2 Hz, 1H, C4-*H*), 6.95 (m, 2H, C2-H, C6-*H*),), 7.24 (apparent t, J = 7.9 Hz, 1H, C5-*H*); ^{13}C-NMR (75 MHz, CD$_3$OD, 25 ºC, TMS): δ = 22.5 (*C*H$_3$-CO), 48.3 (*C*H$_2$-N), 55.6 (O*C*H$_3$), 73.5 (*C*H-O), 112.6 (*C*2), 114.1 (*C*4), 119.4 (*C*6), 130.3 (*C*5), 145.5 (*C*1), 161.2 (*C*3), 173.6 (*C*=O); IR (KBr): ν = 3290, 1634, 1596, 1552, 1259, 1066 cm^{-1}; MS (70 eV, EI): *m/z* (%): 209 (7) [M$^+$], 191 (3) [M$^+$-H$_2$O], 150 (31), 109 (25), 73 (87), 62 (46), 45 (100); elemental analysis calcd (%) for $C_{11}H_{15}NO_3$ (209.24): C 63.14, H 7.23, N 6.69; found: C 61.41, H 7.57, N 6.50.

4.4.5 - General procedure D. Reductions in the Parr autoclave with optimized conditions for substrates prepared from aliphatic aldehydes

Activated 50 % Ni on alumina (100 mg) was added to a solution of the substrate (5.7 mmol) in dioxane (30 ml) and water (0.2 ml). After stirring the reaction at 140 ºC and

10 bar H_2, the reaction mixture was filtered. A sample of 2 ml was taken from the filtrate and the solvents from this sample were removed under vacuum and the sample was then analysed by ^{1}H-NMR. The combined filtrate and NMR-sample was then evaporated to dryness to yield the oil or solid products.

***N*-(2-Hydroxyheptyl) acetamide 2c**: The oil prepared from **1c** according to general procedure D was purified by recrystallisation from EtOAc. Yield of **2c**: 454 mg (56 %) as a white solid; m.p. 75-76 ºC; ^{1}H-NMR (300 MHz, $CDCl_3$, 25 ºC, TMS): δ = 0.89 (m, 3H, CH_3-CH_2), 1.29-1.47 (m, 8H, CH_3-CH_2-CH_2-CH_2-CH_2), 2.00 (s, 3H, CH_3-C=O), 3.08 (ddd, *J* = 5.0, 7.9, 13.7 Hz, 1H, CH_2-N), 3.45 (ddd, *J* = 2.9, 6.6, 13.9 Hz, 1H, CH_2-N), 3.45 (s, 1H, O*H*), 3.69 (m, 1H, C*H*-O), 6.49 (s, 1H, N*H*); ^{13}C-NMR (75 MHz, $CDCl_3$, 25 ºC, TMS): δ = 14.0 (CH_3-CH_2), 22.6 (CH_3-CH_2), 23.2 (CH_3-CO), 25.2 (CH_3-CH_2-CH_2), 31.8 (CH_2-CH_2-CH), 35.0 (CH_2-CH), 45.9 (CH_2-N), 71.2 (*C*H-O), 171.4 (*C*=O); IR (KBr): ν = 3425, 3279, 1661, 1627, 1586, 1569, 1136 cm^{-1}; MS (70 eV, EI): *m/z* (%): 174 (3) [M^++1], 102 (10), 73 (100); elemental analysis calcd (%) for $C_9H_{19}NO_2$ (173.25): C 62.39, H 11.05, N 8.08; found: C 62.01, H 11.67, N 8.04.

(*S*)-*N*-(2-Hydroxyheptyl) acetamide (*S*)-2c: The oil prepared from **(*S*)-1c** (5.4 mmol, 94 % *ee*) according to general procedure D was purified by recrystallisation from EtOAc. Yield of **(*S*)-2c**: 533 mg (57 %) as a white solid; *ee* = 95 %; m.p. 75-76 ºC; $[\alpha]_D^{20}$ = + 14.1 (*c* = 1.0 in MeOH); other spectroscopic data as for **2a**.

***N*-(2-Hydroxy-3-phenoxypropyl) acetamide 2d**: The oil prepared from *rac*-**1d** according to general procedure D was purified by column chromatography (silica, EtOAc/MeOH, 95:5, R_f = 0.29). Yield of **2d**: 967 mg (72 %) as a white solid; m.p. 49-50 ºC; ^{1}H-NMR (300 MHz, $CDCl_3$, 25 ºC, TMS): δ = 1.99 (s, 3H, CH_3-C=O), 3.36 (ddd, *J* = 5.5, 6.8, 6.8 Hz, 1H, CH_2-N), 3.59 (ddd, *J* = 3.3, 6.1, 14.0 Hz, 1H, CH_2-N), 3.92 (d, *J* = 5.5 Hz, 2H, CH_2-O), 4.09 (m, 1H, C*H*-O), 4.18 (s, 1H, O*H*), 6.58 (s, 1H, N*H*), 6.87 (m, 2H, aromatic), 6.95 (m, 1H, aromatic), 7.26 (m, 2H, aromatic); ^{13}C-NMR (75 MHz, $CDCl_3$, 25 ºC, TMS): δ = 23.0 (CH_3-C=O), 43.0 (CH_2-N), 69.5 (*C*H-OH and CH_2-O), 114.5, 121.2, 129.6, and 158.4 (aromatic), 171.9 (*C*=O); IR (KBr): ν = 3384, 3299, 1630, 1601, 1571, 1284, 1118, 751 cm^{-1}; MS (70 eV, EI): *m/z* (%): 209 (3) [M^+], 191 (32) [M^+-H_2O], 148 (7), 116 (100); elemental analysis calcd (%) for $C_{11}H_{15}NO_3$ (209.24): C 63.14, H 7.23, N 6.69; found: C 61.99, H 7.22, N 6.46.

***N*-(2-Hydroxy-3-methylbutyl) benzamide 2e**: The solid prepared from *rac*-**1e** according to general procedure D was purified by column chromatography (silica, EtOAc/PE, 45:55, R_f = 0.30). Yield of **2e**: 681mg (58 %) as a white solid; m.p. 116-117 ºC; ^{1}H-NMR (300 MHz, $CDCl_3$, 25 ºC, TMS): δ = 0.97 (dd, J = 6.8, 9.0 Hz, 6H, 2 C*H*$_3$), 1.73 (m, 1H, C*H*-(CH_3)$_2$), 3.06 (s, 1H, O*H*), 3.30 (ddd, J = 4.6, 8.61, 13.7 Hz, 1H, C*H*$_2$-N), 3.50 (m, 1H, C*H*-O), 3.72 (ddd, J = 2.8, 6.8, 13.7 Hz, 1H, C*H*$_2$-N), 6.86 (s, 1H, N*H*), 7.38 (m, 2H, aromatic), 7.46 (m, 1H, aromatic), 7.77 (m, 2H, aromatic); ^{13}C-NMR (75 MHz, $CDCl_3$, 25 ºC, TMS): δ = 17.9 (*C*H_3), 18.6 (*C*H_3), 32.3 (*C*H-(CH_3)$_2$), 44.1 (CH_2-N), 76.3 (*C*H-O), 127.0, 128.5, 131.5, and 134.3 (aromatic), 168.5 (*C*=O); IR (KBr): ν = 3398, 3319, 1633, 1578, 1541, 1057, 697 cm^{-1}; MS (70 eV, EI): *m/z* (%): 207 (1) [M^+], 189 (3) [M^+-H_2O], 164 (16), 134 (89), 122 (29), 105 (100); elemental analysis calcd (%) for $C_{12}H_{17}NO_2$ (207.27): C 69.54, H 8.27, N 6.76; found: C 68.75, H 8.61, N 6.68.

***N*-(2-Hydroxyhepyl) butanamide 2f**: The oil prepared from *rac*-**1f** according to general procedure D was purified by recrystallisation from EtOAc. Yield of **2f**: 345 mg (30 %) as a white solid; m.p. 62-63 ºC; ^{1}H-NMR (300 MHz, $CDCl_3$, 25 ºC, TMS): δ = 0.89 (m, 3H, pentyl-C*H*$_3$), 0.95 (t, J = 7.5 Hz, 3H, propyl-C*H*$_3$), 1.29-1.44 (m, 8H, CH_3-C*H*$_2$-C*H*$_2$-C*H*$_2$-C*H*$_2$), 1.67 (sextet, J = 7.4 Hz, 2H, C*H*$_2$-CH_2-C=O), 2.18 (t, J = 7.4 Hz CH_2-C*H*$_2$-C=O), 2.94 (s, 1H, O*H*), 3.11 (ddd, J = 4.9, 7.7, 13.0 Hz, 1H, C*H*$_2$-N), 3.47 (ddd, J = 2.7, 6.2, 13.7 Hz, 1H, C*H*$_2$-N), 3.70 (m, 1H, C*H*-O), 6.15 (s, 1H, N*H*); ^{13}C-NMR (75 MHz, $CDCl_3$, 25 ºC, TMS): δ = 13.8 (propyl-*C*H_3), 14.0 (pentyl-*C*H_3), 19.2 (*C*H_2-CH_2-C=O), 22.6 (CH_3-*C*H_2-CH_2-CH_2), 25.2 (CH_3-CH_2-*C*H_2-CH_2), 31.8 (*C*H_2-CH_2-CH), 35.0 (*C*H_2-CH), 38.6 (*C*H_2-C=O), 45.7 (*C*H_2-N), 71.5 (*C*H-O), 174.2 (*C*=O) ; IR (KBr): ν = 3418, 3283, 2964, 2919, 1657 ,1624, 1566 cm^{-1}; MS (70 eV, EI): *m/z* (%): [M^+] could not be identified, 130 (17), 101 (100); elemental analysis calcd (%) for $C_{11}H_{23}NO_2$ (173.25): C 65.63, H 11.52, N 6.96; found: C 64.78, H 12.03, N 6.83.

4.5 - References

[1] M. North, *Tetrahedron: Asymmetry* **2003**, *14*, 147.

[2] H. Griengl, H. Schwab and M. Fechter, *Trends Biotechnol*. **2000**, *18*, 252.

[3] J. Brussee, A. van der Gen, in *Stereoselective Biocatalysis* (Ed.: P. N. Ramesh), Marcel Dekker, Inc., New York, **2000**, pp. 289.

[4] L. T. Kanerva, *Acta Chem. Scand.* **1996**, *50*, 234.

[5] K. Tanaka, A. Mori and S. Inoue, *J. Org. Chem* **1990**, *55*, 181.

[6] T. Ziegler, B. Hörst and F. Effenberger, *Synthesis* **1990**, 575.

[7] T. Ooi, M. Kameda, J. Fujii and K. Maruoka, *Org. Lett.* **2004**, *6*, 2397.

[8] W. J. Greenlee, J. P. Springer and A. A. Patchett, *J. Med. Chem,* **1989**, *32*, 165.

[9] A. Fishman and M. Zviely, *Tetrahedron: Asymmetry* **1998**, *9*, 107.

[10] M. Inagaki, J. Hiratake, T. Nishioka and J. Oda, *J. Org. Chem.* **1992**, *57*, 5643.

[11] L. Veum, L. T. Kanerva, P. J. Halling, T. Maschmeyer and U. Hanefeld, *Adv. Synth. Catal.* **2005,** *347*, 1015.

[12] C. Paizs, P. Tähtinen, M. Toşa, C. Majdik, F. D. Irimie and L. T. Kanerva, *Tetrahedron* **2004**, *60*, 10533.

[13] S. Lundgren, E. S. Lundgren, E. Wingstrand, M. Penhoat, C. Moberg, *J. Am.Chem. Soc.* **2005**, *127*, 11592.

[14] Y. N. Belokon, P.Carta, A. V. Gutnov, V. Maleev, M. A. Moskalenko, L. V. Yashkina, N. S. Ikonnikov, N. V. Voskoboev, V. N. Khrustalev, M. North, *Helv. Chim. Acta* **2002**, *85*, 3301.

[15] Y. N. Belokon,P. Carta, M. North, *Lett. Org. Chem.* **2004**, *1*, 81.

[16] M. Ikezaki, N. Umino, M. Gaino, K. Aoe, T. Iwakuma and T. Oh-Ishi, *Yakugaku Zasshi*, **1986**, *106*, 80.

[17] W. H. Hartung, *J. Am. Chem. Soc.* **1928**, *50,* 3370.

[18] J. S. Buck, *J. Am. Chem. Soc.* **1933**, *55*, 2593.

[19] D. L. Massart, B. G. M.Vandeginste, L.M.C. Buydens, S. de Jong, P. J. Lewi and J. Smeyers-Verbeke, *Handbook of Chemometrics and Qualimetrics, Part A*, Elsevier, Amsterdam, **1997**.

[20] E. W. Kirchhoff , D. R. Anderson, S. Zhang, C. C. Cassidy and M. T. Flavin, Automated Process Research and the optimization of the synthesis of 4(5)-(3-Pyridyl)imidazole, *Organic Process Res. & Dev.*, **2001**, 5, 50.

[21] R. Marchetti and M. E. Guerzoni, *Cerevisia Biotechnol*, 16 (1), **1991**, 24.

[22] P. D. Haaland, *Biotechnology Experimental Design* in: *Statistical Design and analysis of Industrial experiments*, (Ed.: S. Ghosh), Marcel Dekker, New York, **1990**, pp. 73-108.

[23] S. Gomez, J. A. Peters and T. Maschmeyer, Adv. *Synth. Catal.* **2002**, *344*, 1037.

[24] J. Volf and J. Pasek. *in Catalytic Hydrogenation Vol 27* (Ed.: L. Cerveny), Elsevier, Amsterdam **1986**, pp. 105.

[25] H. Greenfield, *Ind. Eng. Chem., Prod. Res. Dev.* **1976**, *15*, 156.

[26] Y. Huang and V. Adeeva, W. M. H. Sachtler, *Appl. Catal. A: Gen.* **2000**, *196*, 73.

[27] P. F. Aguiar, B. Bourguignon, M. S. Khots, D. L. Massart and R. Phan-Than-Luu, *Chemom. Intell. Lab. Syst.*, **1995**, 30, 199.

[28] H. Griengl, N. Klempier, P. Pöchlauer, M. Schmidt, N. Shi and A. A. Zabelinskaja-Mackova, *Tetrahedron*, **1998**, *54*, 14477.

5

SYNTHESIS OF TWO CATALYST RESPONSE SURFACE BENCHMARKS FOR CO OXIDATION: COOX AND SELOX

Abstract

Two catalytic response surface benchmarks were derived based on the performance of libraries of heterogeneous mixed-oxide catalysts in the CO oxidation (COOX) and selective CO oxidation (SELOX) reactions for fuel cell applications. Almost 200 catalysts were synthesised and tested by means of High-Throughput Experimentation (HTE). The catalyst library was investigated under different reaction conditions and more than 1000 catalyst performance results were obtained and used to build the SELOX and COOX benchmark response surfaces. A Design of Experiments (DoE) approach was applied for the experimental plan and modelling of the resulting response surfaces. COOX and SELOX constitute realistic catalytic benchmarks that are used in the later chapters to test the performance of Global Optimization algorithms.

The contents of this chapter have been published in:
S. R. M. Pereira, F. Clerc, D. Farrusseng, J. C. van der Waal, T. Maschmeyer, C. Mirodatos, *QSAR and Comb. Sci.*, 24, **2005**, 45.

5.1 - Introduction

Since a few years, High-Throughput Experimentation (HTE) enables fast synthesis and testing of large libraries of materials [1-4]. However, a systematic investigation of the entire parameter space is usually inefficient and often experimentally unaffordable since the high number of parameters to investigate leads to a too large number of sample candidates to be prepared and tested. A rational selection of the compounds must be considered to limit the size of this library while ensuring an optimum rate of discovery and optimization. Because of the large number of variables and experiments, library design of materials cannot be efficiently performed manually and the assistance of computer science is required for this purpose [5-7]. In drug discovery, explorative data analysis, data mining and artificial intelligence are intensively used since several years in order to efficiently reduce the experimental effort in discovery and optimization processes. Because the description of materials is more complex than for molecules, an approach employing Quantitative Structure Activity/Property Relationships (QSAR/QSPR) that allows for a complementary virtual exploration of large libraries of individual compounds is in general not easily applicable to heterogeneous catalysts [8]. Therefore other design strategies and associated algorithms need to be developed to explore large parameter spaces. Several algorithms like Genetic Algorithms (Chapter 6), Simulated Annealing, Taboo Search methodologies etc. (Chapter 7) are in this way being applied to try to solve this issue. However, a general lack of knowledge exists about the applicability of these algorithms and their parameter settings in order to obtain an efficient optimization performance when applied to catalytic case studies.

The optimization of algorithm configurations and settings is generally performed on so-called virtual or synthetic benchmarks, which are non-chemistry-related custom-made mathematical response surfaces. These functions are usually far from what one can expect for real case studies in material optimizations. Since the algorithm optimization performance is very dependent on the type of surface to optimize, they can be non-representative for testing algorithm performance and finding its best parameters configuration. Ideally, real data collections on heterogeneous catalysis performance would be employed in performing these simulations. However, data banks on heterogeneous catalysts and their performance that could be used as a real application benchmark are usually kept confidential, while data collection from literature usually provides data sets which are too small and often have inconsistent or poor data accumulated over decades. Validation of GA configurations by performing multiple experimental optimizations at different configuration settings is usually not an alternative because it would lead to a too large number of runs and experimental effort in order to get statistically significant results. Benchmarks based on experimental and simulation catalytic results using Artificial Neural networks strategies have been used by Rodemerk *et al.* to test algorithm performances [9]. In this chapter a methodology for constructing benchmarks is discussed, based on DoE modelling of catalytic data obtained from an experimental design. The considered catalytic search space was systematically mapped using a Design of Experiments approach that allowed the construction of a response surface described by a set of mathematical functions. These will be later used as benchmarks to test the influence of algorithm settings (Chapter 6) and different algorithms (Chapter 7) in the optimization performance of the algorithms.

The reactions chosen to obtain the catalytic benchmarks are the oxidation of carbon monoxide to carbon dioxide (COOX) and the same reaction but in the presence of hydrogen (SELOX).

In order to supply proton-exchange membrane fuel cells (PEMFC) for on-board or domestic electricity generation, ideally a pure hydrogen fuel should be used at the anode and oxygen/air at the cathode. Due to the numerous current problems related to hydrogen infrastructure, alternative fuel solutions are currently being investigated, namely hydrocarbon fuels. In this solution, the hydrocarbon fuel is first transformed to a hydrogen-rich but CO-containing gas, which is then conducted to the anode catalyst of the PEMFC where the hydrogen undergoes electro-oxidation. The power efficiency of the PEMFC is substantially reduced by the presence of CO in the hydrogen stream, due to the poisoning of the anode catalyst by the adsorption and surface-bonding of CO and consequent deactivation. Anode catalysts tolerant to CO adsorption have been investigated but the issue is still one of the main challenges in PEMFC technology. An enormous effort has also been made in methods to remove the CO from the CO/H_2 gas mixture in order to feed the catalytic anode of the PEMFC with a cleaner gas stream. The SELOX and COOX reactions are representative of these present efforts in finding catalytic solutions to promote new clean and miniaturised processes for generating hydrogen from hydrocarbon fuels.

The SELOX and COOX benchmark response surfaces will be used in later chapters (Chapter 6 and 7) to study the performance of several algorithms in the exploration and optimization of catalytic response surfaces.

5.2 - COOX and SELOX benchmarks planning

5.2.1 - Parameter/Search Space

The catalyst search space was constructed according to generally accepted principles in catalysis, namely that heterogeneous catalysts are usually multi-component systems, and that each component has specific functions and interacts with the others. For the present case study, the composition search space was restricted to 3 components per catalyst, which were considered representative elements for studying the reaction of CO oxidation in the absence or in the presence of H_2. The three components chosen were: i) one metal (NM) acting as the main active phase (Au, Cu, Pt), ii) one transition metal (TM) acting as modifier/promoter of the active phase and support (Mo, Nb, V), and iii) one metal oxide acting as support (CeO_2, TiO_2, ZrO_2). The elemental composition of NM and TM varies from 0.1 to 2.1 % and from 1 to 5%, respectively. It was considered that the bulk composition of the support does not change significantly with the concentration of TM and NM. Therefore, the catalyst composition search space can be considered as made of 3^3 distinct continuous sub-spaces. Temperature is also a key parameter for the working of the CO oxidation reactions. Therefore, the catalyst library was tested at 3 different temperatures namely 200, 225, and 250°C for the COOX, and 200, 250, and 300°C for the SELOX benchmarks. This results in a total search space consisting of 81 distinct continuous response surfaces or sub-spaces for each benchmark. Each sub-space represented the reactivity of a different catalyst ternary system in respect to the variation of its NM and TM concentration and at the corresponding tested temperature.

5.2.2 - Design of Experiments for the COOX and SELOX benchmarks

For mapping the search space a design of experiments strategy based on the optimization Doehlert DoE design (see Chapter 2) has been chosen. The sub-spaces were considered to be able to be modelled by linear regression, assuming that non-linear effects do not prevail for this type of reaction in the parameter space chosen. The 81 sub-spaces can therefore be modelled using an optimization Doehlert Design as experimental planning technique that allows the fitting of a quadratic surface model [10].

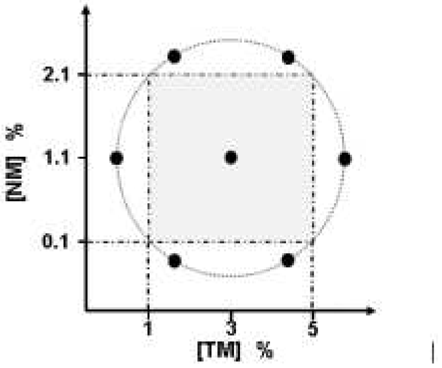


Figure 5.1 - Doehlert Experimental Design for a NM, TM and Support ternary system. The Support amount is fixed. The dots represent the experimental points.

This design (Figure 5.1), which requires seven independent experiments, enables to reach an efficient and robust model for the main effects and second order parameters accounting for interaction and quadratic effects (Equation 5.1) [11].

$$y = \beta_0 + \beta_1[TM] + \beta_2[NM] + \beta_{12}[TM].[NM] + \beta_{11}[TM]^2 + \beta_{22}[NM]^2 \qquad \textbf{Equation 5.1}$$

Where [TM] and [NM] represents the TM and NM mass percentage, respectively.

Therefore, the construction of the response surface benchmark can be performed by collecting the 81 distinct linear models. This experimental plan was applied to all 27 ternaries resulting in a library of 189 (3^3x7) catalysts, which were synthesised in batches of 24-48 catalysts using the automated equipment described in Appendix A.2. All catalysts were tested for both reactions (COOX and SELOX) with a HTE 16-vessel gas reactor (Appendix A.4) at 3 different temperatures, yielding 1134 catalytic data point results (27x7x2) plus replicates.

5.2.3 - Weighted surface response - Desirability

For obtaining cleaner H_2 gas streams to feed the PEMFC catalyst anode it is crucial to obtain the highest possible conversion of the CO present. For the COOX reaction conditions the important response to consider is therefore the conversion of CO (Equation 2a). In the SELOX reaction, due to the possibility of the parallel oxidation of H_2 (Equation 5.2.b), both CO and H_2 conversion and CO_2 and H_2O selectivities need also to be considered. There are therefore multi-responses that need to be taken into consideration.

COOX: $CO + 0.5\ O_2 \rightarrow CO_2$ **Equation 5.2a**

SELOX: $CO + 0.5\ O_2 \rightarrow CO_2$ **Equation 5.2a**
$H_2 + 0.5\ O_2 \rightarrow H_2O$ **Equation 5.2b**

$X(CO) = pCO_2/pCO^o$ **Equation 5.3a**
$S(CO_2) = pCO_2/(2.pO_2^o - 2.pO_2)$ **Equation 5.3b**

where, p denotes the partial pressure of the respective gases and the superscript ° means the inlet partial pressure, whereas no superscript means the outlet partial pressure.

For multi-response analysis or optimization several methods can be applied (see Chapter 2). The desirability function method weighs all the responses according to their relevance into a single criterion, the Desirability function [4].

Since the highest possible CO conversion (X(CO)) is the key criterion for fuel-cell applications (Equation 5.3a), a larger weight shall be given to CO conversion with respect to other criteria. As stressed above, for the SELOX reaction, CO_2 selectivity - $S(CO_2)$ (Equation 5.3b) - has also to be considered as a key criterion because the oxidation has to be selective towards CO while limiting parallel H_2 oxidation. In addition, since low-temperature oxidation is preferred, in connection with the requirements of other units of the PEMFC system, the temperature was also chosen as a criterion for the performance assessment. The weights of three system responses were defined with mathematical Desirability function curves as given in Figure 5.2.

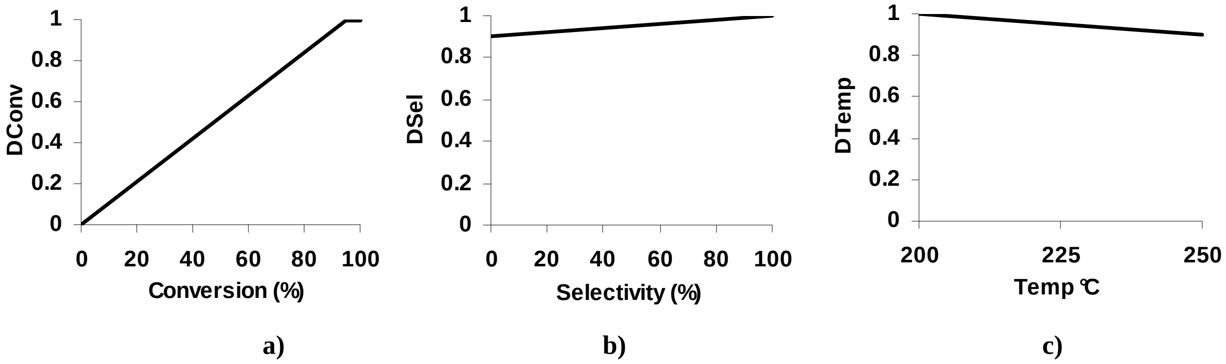


Figure 5.2 - Desirability Functions for: a) CO Conversion (D_{Conv}), b) CO_2 Selectivities (D_{Sel}) and c) Temperature (D_{Temp})

D_{Conv} = 0.0105 X(CO) if X(CO) < 95% and D_{Conv} = 1 if X(CO) $\geq$ 95%

D_{Sel} = 0.001 S(CO_2) + 0.9

D_{Temp} = - 0.002 Temp + 1.4

The steeper the slope of the Desirability curve, the larger the weight. The Desirability function for CO conversion (denoted D_{Conv}) is approximately proportional to the conversion itself and works basically as a scaling function. Due to limitations in quantifying accurately CO and CO_2 traces, the Desirability for catalysts exhibiting conversion higher than 95% was set to 1. Two Desirability functions with low weights were defined for CO_2 selectivity and temperature, which are denoted D_{Sel} and D_{Temp}, respectively. The global Desirability functions, which represent the benchmarks fitness function, combine all sub-Desirabilities and are defined as $D_{COOX} = D_{Conv}.D_{Temp}$ and $D_{SELOX} = D_{Conv}.D_{Sel}.D_{Temp}$, for the COOX and SELOX benchmarks, respectively. The combination of weighted performance criteria in a single evaluation function enables a more refined optimization, making it possible to discriminate between catalysts in the same conversion range.

5.3 - Results and Discussion

5.3.1 - Description of the surface responses for COOX and SELOX reactions

After synthesising (section 5.4) and testing the catalysts library, the conversion and selectivity values of all 7x27 experiments were collected. The coefficients of Equation 5.1 for each of the 81 sub-spaces were obtained by linear regression and are presented in Appendix B. The obtained equations for the quadratic models show in general good fitting in relation to the experimental points, having in the majority of the cases an R-Squared value higher than 0.8 (Appendix B). As an example, the resulting sub-response surface for the ternary system Pt-V-Ti at 250°C for SELOX conditions is shown in Figure 5.3.

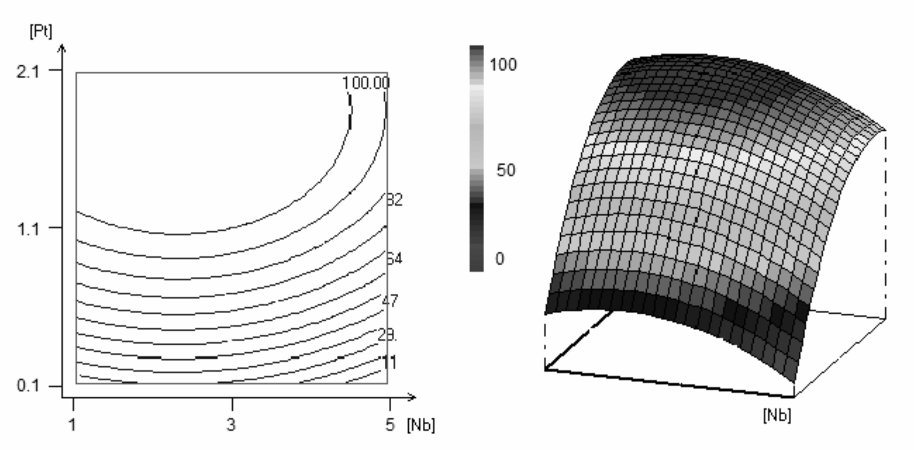
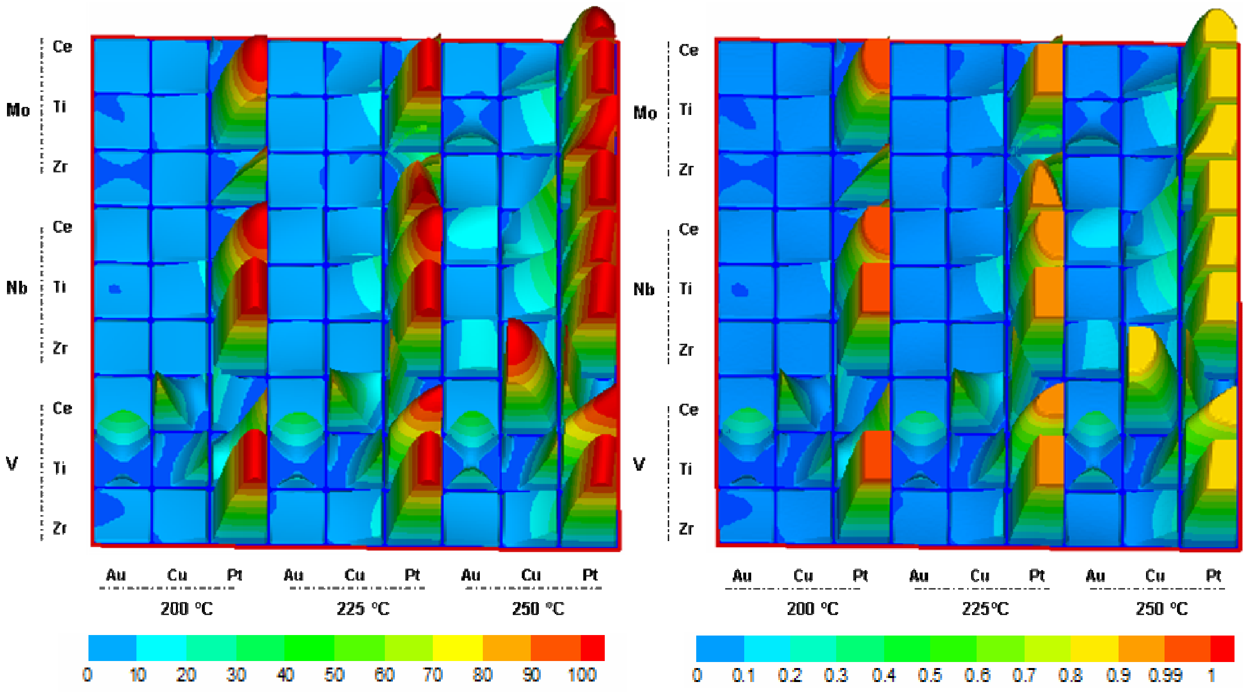


Figure 5.3 - Sub-response surface for the ternary Pt-Nb-Zr at 200°C for the conversion response at SELOX reaction conditions. $y = 100 - 11[TM] + 49[NM] - 16[NM]^2 - 32[TM]^2$

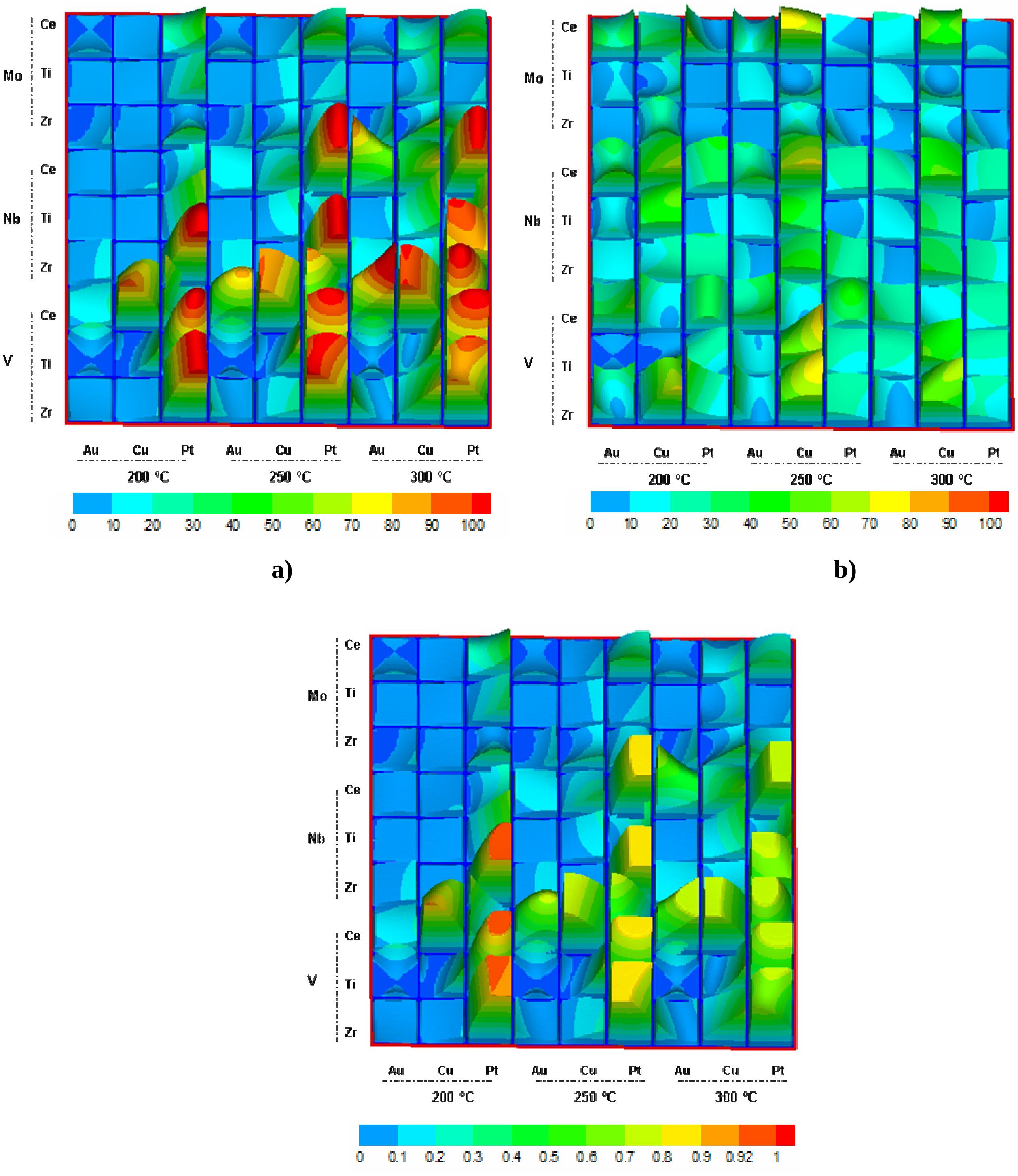
All sub-response surfaces were gathered, and the desirability functions were applied, yielding the two COOX and SELOX benchmarks. A visualisation of the resulting benchmark response surfaces is depicted in Figure 5.5 and 5.6 respectively.

For the COOX conversion response surface (Figure 5.4a), it is obvious that Pt containing catalysts are the best performing ones, whatever the testing temperature, except when associated with CeO_2. The composition Cu-V-CeO_2 also shows significant conversion, particularly at 250°C, and V-TiO_2 are always active compositions, whatever the temperature and metal. It can be noted that the conversion usually increases with temperature as expected for activated processes. All other combinations result in inactive or poorly active systems, even at a high temperature such as 250° C. Figure 5.5b shows the effect of surface response transformation by the Desirability function. Because the Desirability function for the temperature penalises high temperatures (negative slope), final optima of the COOX benchmark are located at the lower temperature of 200°C.



a) **b)**

Figure 5.4 - COOX response surface a) Conversion: X(CO), b) COOX Fitness landscape after applying the Desirability functions(D_{COOX}).



a) **b)**

c)

Figure 5.5 - SELOX response surface. a) Conversion: X(CO), b) Selectivity: S(CO_2) and c) SELOX Fitness landscape after applying the Desirability functions (D_{SELOX}).

When performing the reaction in the presence of H_2 (SELOX) a slightly different landscape emerges (Figure 5.5a). Even if Pt is still the metal with the highest influence on performance, not all compositions containing Pt are active, especially at low temperature. In addition, Cu-V/CeO_2 performs much better with increasing conversion values when temperature increases in contrast with Pt based catalysts.

For the SELOX benchmark, the nature of the promoter (TM) was found to have more influence on the catalyst performance, giving the ranking V>Nb>Mo. Cerium-based catalysts are on average the best systems in comparison with the other supports. However, at low temperature synergy with Pt is not favourable, since the ternary Pt-V-Ce catalysts compositions are poorly performing. Response variance for CO_2 selectivity is relatively low with respect to CO conversion. Nevertheless, Cu-based catalysts exhibit generally higher selectivity.

Through the desirability function, response surfaces are weighted with respect to temperature and selectivity. The global response emphasises the search space areas of larger interest and enables discrimination between catalysts with very similar conversion values. The top surfaces of response hills are flattened due to the Desirability settings that all catalysts with a conversion higher than 95% get the maximum desirability value (D_{Conv} =1). These flat surfaces are then discriminated between themselves due to the D_{temp} and D_{sel} functions (Figure 5.5b, Figure 5.5c). In Table 5.1 the catalyst formulations that contribute for the global optimum response in the COOX (D_{COOX} >0.99), and SELOX benchmarks (D_{SELOX}>0.92) are listed.

Table 5.1 - Global Optimal catalyst solutions for the COOX and SELOX benchmarks.

Optimum	COOX	SELOX
1	Pt/V/Zr/200 ° C	Pt/V/Zr/200 ° C
2	Pt/Nb/Zr/200 ° C	Pt/Nb/Zr/200 ° C
3	Pt/Nb/Ti/200 ° C	Pt/V/Ti/200 ° C
4	Pt/Mo/Ti/200 ° C	Cu/V/Ce/200 ° C
5	Pt/Mo/Zr/200 ° C	

In the presence of H_2, the CO conversion does not always increase with temperature. For instance for the ternary Pt-V-Zr system under SELOX conditions (Figure 5.5a), raising the temperature leads to a decrease of activity. This can be explained by assuming a reverse Water Gas Shift (WGS) mechanism occurring in the downstream catalyst bed during CO oxidation. In this mechanism the CO_2 formed from the initial CO oxidation reaction is in turn reduced to CO and H_2O by the hydrogen present [13]. The preferential oxidation of hydrogen at high temperature may also contribute to this effect. This study has confirmed the beneficial effect of V doping of CeO_2 on a broad range of compounds. This synergetic effect was reported for both SELOX and Water Gas shift reaction with catalysts prepared by impregnation [13].

The Au/TiO_2 Catalysts are known in the literature to be among the best formulations for the CO oxidation reaction [12] but were found in the present work to be inactive. This may come from the synthesis procedure used in this study. It might lead to a low dispersion of the gold particles. Despite the high surface area of the TiO_2 support, high-resolution-electron-microscopy (HREM) results denote the presence of large sized gold particles (Figure 5.6).

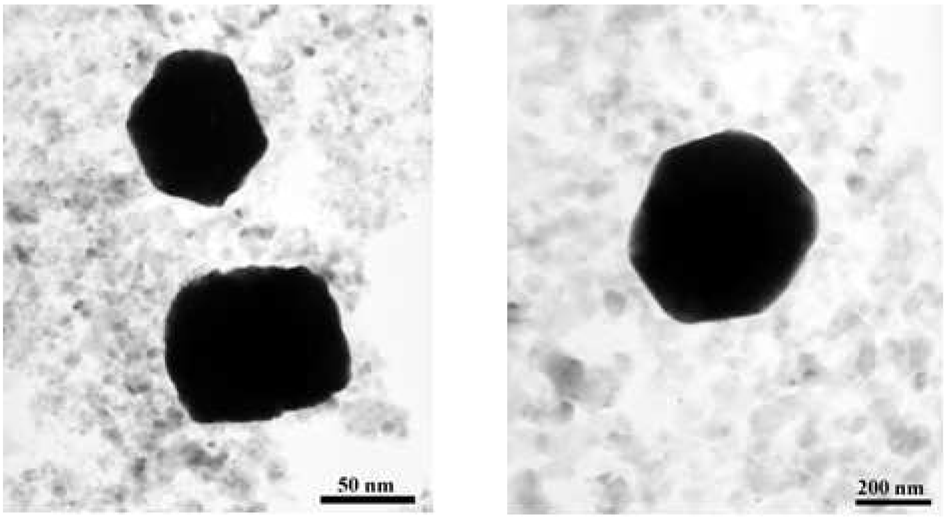


Figure 5.6 - High Resolution Electron Microscopy (HREM) pictures of Au-TiO_2 sample catalysts.

As a matter of fact, the good performances of Au-based catalysts are usually related to the high metal dispersion that ensures a partially ionic state of Au atoms at the particles-support interface. This effect is lost when a considerable quantity of large gold particles are present, resulting in a less active catalyst. In addition, rather low activity for Cu-CeO_2 samples is noticed, compared to catalysts prepared by other methods as investigated by Tibiletty *et al.* [13]. Here also the explanation may lie in a poor Cu dispersion.

The low activity obtained with the Au-TiO_2 and Cu-CeO_2 catalysts indicates a possible pitfall of applying general catalyst preparation procedures typically used in a HTE strategy. Catalyst performance is usually highly sensitive to synthesis procedure. The preferred use of non-laborious and generally applicable preparation methods in HTE for catalyst libraries synthesis can result in the use of non-optimal procedures for the synthesis of certain catalyst formulations in the library. The influence of preparation methods should not be underestimated and different procedures should be taken into account and optimized in order to obtain good performing catalyst solutions. This can be however extremely laborious or even impracticable when the goal is the screening of a highly diverse catalyst library. The synthesis of the catalyst libraries may need to become more labour-intensive or versatile and new technologies that can reproduce complex catalyst recipes at the smaller scale required for HTE may need to be further developed.

Due to the handling of small amounts of compounds in the preparation of HTE catalyst libraries, a larger experimental error can be induced leading to a lower accuracy in the resulting catalyst composition. This uncertainty can be further increased by the limited or absent monitoring of the catalyst performance during time and reduced by extensive heterogeneous catalyst characterisation. The latter is in many cases only performed for the successful or a few chosen unsuccessful catalyst from the synthesised library.

Despite these possible drawbacks of heterogeneous catalyst library synthesis via HTE methods, this is nowadays the only experimentally affordable way to produce highly diverse catalyst libraries and catalyst activity screenings. The amount of information gained via the typical approach used in HTE allows focusing the attention on the most promising catalyst solutions in further screening and optimization steps, in which additional criteria such as the dispersion of the active metal could be introduced into the search space.

5.4 - Conclusions

The COOX and SELOX catalytic benchmarks were successfully obtained based on modelling of the performance of 189 synthesised catalysts at different reaction conditions. The models obtained allow the simulation of the performance of intermediary catalyst compositions in the composition parameter space considered. The benchmarks obtained can be used to navigate the parameter space and serve as realistic catalytic benchmarks for the testing of optimization performance of optimization algorithms.

Considering the chemical information gained during this study, formulations containing an element known to be active like Pt were confirmed as ranking among the best. However, other active elements, like Au, were not found to ensure high catalytic performance. This apparent discrepancy demonstrates the sensitivity of synthesis parameters and procedures on catalyst performance and shows that, in a combinatorial discovery program, this need to be taken into account in the further refinement of a catalyst formulation optimization procedure.

The results from the selected catalyst libraries obtained do not, nevertheless, invalidate the effectiveness of the approach chosen. The advantage of binning the experimental space in small zones using a Design of experiments approach, which enables one to model the whole surface by simple linear regression with good confidence, has been, from a methodological point of view, clearly demonstrated. It ensures that all the search space can be modelled with about the same confidence level in contrast to other methods such as Artificial Neural Networks. In addition, it allows one also to obtain an explicit model in a mathematical form that can be visualised and further used in Excel or other software as a benchmark for catalyst optimization performance assessment.

5.5 - Experimental

5.5.1 - General

An automated catalyst preparation workstation (Sophas - Zinsser Analytics, see Appendix A.2) was used for the catalyst library synthesis, having a typical throughput of 24-48 catalysts/day.

Materials: Activated carbon Darco® KB-B, titanium isopropoxide, zirconyl nitrate hydrate and copper chloride dihydrate were purchased from Sigma Aldrich. Molybdenum chloride, hexa-ammonium heptamolybdate tetrahydrate, copper nitrate pentahemihydrate and niobium chloride were purchased from Riedel de Haën. Cerium nitrate hexahydrate, chloroplatinic acid hexahydrate and hydrogen tetrachloroaurate hydrate were purchased from Strem, ammonium metavanadate from Fluka. DMSO, acetyl acetone and 2-propanol were of reagent grade.

5.5.2 - General Procedure A: Ce and Zr based catalysts prepared by Impregnation on Carbon

General principles of synthesis by the “carbon route” can be found elsewhere [14]. Solutions of 1 M cerium and zirconyl nitrate, 0.05 M TM (ammonium vanadate, niobium and molybdenum chloride), and NM (0.0025 and 0.025 M hydrogen tetrachloroaurate, 0.05 and 0.001 M chloroplatinic acid, and 0.1 M and 0.01 M copper nitrate) were dispensed in the appropriate volumes to accomplish the required catalysts composition according to the design depicted in Figure 5.1. The final volume was adjusted by adding water until a total volume of 4 ml was reached. The solutions were stirred and transferred to vials containing 1.3 g of activated carbon. Subsequently, the impregnated carbon was left to dry for 5 h at 120 °C while shaking. All the previous procedures have been performed by the automated catalyst synthesiser platform presented in appendix A.2. The catalysts were then calcined for 3 h at 550 °C with an initial heating rate of 2 °C/min.

5.5.3 - General Procedure B: Ti supported catalysts prepared by SolGel

To each of the reactor vials 1.5 M titanium isopropoxide with 2 M acetyl acetone in 2-propanol solution was added manually. Subsequently, the TM solutions (0.05 M ammonium vanadate, 0.05 M niobium chloride in DMSO, 0.05 M molybdenum chloride in DMSO) and NM (0.0025 and 0.025 M hydrogen tetrachloroaurate, 0.05 and 0.001 M chloroplatinic acid, and 0.1 M and 0.01 M copper chloride in DMSO) were dispensed in the appropriate volumes to accomplish the required catalysts composition. The reactor vials were heated at 120 °C and allowed to dry under shaking during the night. The catalysts were then calcined as mentioned above.

5.5.4 - General procedure C: High-Throughput Catalysts Testing

All catalysts were tested in a parallel reactor setup (Switch 16 - Amtec, see Appendix A.4) constituted of 16 channels loaded with 150 mg of catalyst. The analyses were performed on a micro-GC (Agilent), which enables complete quantification of all products within 3 min. More details about the reactor system can be found elsewhere [13,15]. When placed in the parallel reactor the 16 catalysts were tested first under COOX (no H_2 present) and then under SELOX conditions (H_2 present). Detailed reaction conditions are given in Table 5.2. After a temperature ramp (1°C/min), the reactor was allowed to stabilise for 15 min before starting the sequential analysis. The analyses were always duplicated implying a throughput of 192 analyses/day accounting for 16 (catalysts) x 3 (temperatures) x 2 (reaction conditions) x 2 (duplicates).

Table 5.2 - Reaction conditions (feed composition, temperature, catalyst loading and total flow) for COOX and SELOX

	H_2 (%)	O_2 (%)	CO (%)	N_2 (%)	Temp $_1$ (ºC)	Temp $_2$ (ºC)	Temp $_3$ (ºC)	Catalyst (mg)	Total Flow (ml/min)
COOX	-	2	1	97	200	225	250	150	30
SELOX	10	2	1	87	200	250	300	150	30

5.6 - References

[1] W. F. Maier, *Angew. Chem. Int. Ed.,* **1999**, *38*, 1216.

[2] B. Jandeleit, D. J. Schaefer, T. S. Powers, H. W. Turner and W. H. Weinberg, *Angew. Chem., Int. Ed.,* **1999**, *38*, 2494.

[3] M. Baerns and C. Mirodatos, *NATO Science Series, II: Mathematics, Physics and Chemistry,* **2002**, *69*, 469.

[4] F. Schüth, C. Hoffmann, A. Wolf, S. Schunk, W. Stichert and A. Brenner, *Comb. Chem.*, **1999**, 463.

[5] J. Cawse, Wiley-VCH, Weinheim, *Experimental design for combinatorial and high throughput materials development*, **2003**.

[6] L. A. Harmon, A. J. Vayda and S. G. Schlosser, *Abstr. Pap. - Am. Chem. Soc.*, **2001**, *221st*, BTEC.

[7] D. Farrusseng, L. Baumes and C. Mirodatos, in *High-Throughput Analysis: A Tool For Combinatorial Materials Science* (Eds.: R. A. Potyrailo., E. J. Amis.), Kluwer Academic/Plenum Publishers, **2003**, pp. 551.

[8] C. Klanner, D. Farrusseng, L. Baumes, C. Mirodatos and F. Schüth, *QSAR Comb. Sci.*, **2003**, *22*, 729.

[9] U. Rodemerck, H. Kosslick, G. U. Wolf, S. Kolf, M. Baerns, *Chem. Ing. Tech.*, **2002**, *74*, 559

[10] D. L. Massart, B. G. M. Vandeginste, L. M. C. Buydens, S. de Jong, P. J. Lewi and J. Smeyers-Verbeke, *Handbook of Chemometrics and Qualimetrics*; Elsevier, **1997**.

[11] G. Renner and A. Ekart, *Comput. Aided Des.*, **2003**, *35*, 709.

[12] A. Wolf and F. Schüth, *Appl. Catal. A: General*, **2002**, *226*, 1.

[13] D. Tibiletti, E. A. Bart de Graaf, S. P. Teh, G. Rothenberg, D. Farrusseng and C. Mirodatos, *J. Catal.* **2004**, *225*, 489.

[14] M. Schwickardi, T. Johann, W. Schmidt, O. Busch and F. Schüth, *Stud. Surf. Sci. Catal.*, **2002**, *143*, 93.

[15] D. Farrusseng, L. Baumes, I. Vauthey, C. Hayaud, P. Denton and C. Mirodatos, in *Principles and methods for accelerated catalyst design and testing* (Eds.: E. G. Derouane, V. Parmon, F. Lemos, F. R. Ribeiro), Kluwer Academic Publishers, Dordrecht, the Netherlands, **2002**, pp. 101.

6

EFFECT OF GENETIC ALGORITHM PARAMETERS ON THE EFFICIENCY OF HETEROGENEOUS CATALYST OPTIMIZATION

Abstract

A study of the effect of Genetic Algorithm (GA) configurations on the performance of heterogeneous catalyst optimization is reported in this Chapter. The GA optimization procedure is validated on the COOX and SELOX benchmarks obtained in Chapter 5. Because of the typical limitations in the number of parallel experimentations which can usually be carried out in heterogeneous catalysis, the effects of the population size on the robustness and convergence speed were investigated. From this study, general considerations about the algorithm settings (crossover, selection and mutation) to use for the optimization of similar heterogeneous catalyst issues are addressed.

The contents of this chapter have been published in:

- S. R. M. Pereira, F. Clerc, D. Farrusseng, J. C. van der Waal, T. Maschmeyer, C. Mirodatos, *QSAR Comb. Sci.*, **2005**, 24,45.
- F. Clerc; M. Lengliz; D. Farrusseng;, C. Mirodatos, S. R. M. Pereira; R. Rakotomalala, *Rev. Sci. Instrum.*, **2005**, 76, 062208.

6.1 - Introduction

The theoretical basis for the optimization strategy called Genetic Algorithms (GA) was conceived by Holland in the early 1970's [1-3] employing the natural evolution rules of selection and survival of the fittest as postulated by Darwin.

A GA tries to mimic the evolutionary process of living species by using similar genetic operators - mating, crossover and mutation - to single out the individuals with the "genetic information" that leads to the best performance [4,5]. An Evolutionary Algorithm was successfully applied by Wolf *et al.* for optimizing a catalyst formulation for the reactions of oxidative dehydrogenation of ethane and propane [6-12]. Nevertheless, despite proof of concept validation for other reaction systems [13-15], the use of GA strategy is still scarce in this domain. A possible explanation of the above statement is the lack of confidence in the "black-box" optimization processes for the chemists who are trained only for rational catalysts design and trial and error approach.
The use of non-optimal implementation of GAs would result either in a tedious and time-consuming optimization process or could even entirely jeopardise a discovery project. As a matter of fact, the number of parameters available for a GA implementation makes the finding of a robust and optimal GA configuration difficult. The task is further complicated due to the stochastic behaviour of GA, in which many different runs are required to get a statistically significant quality assessment of a certain GA architecture [15]. Considering that the sample preparation and testing in heterogeneous catalysis implies that typically only a few generations can be screened per week, several months of work can be lost if a wrong GA configuration is set.
The optimal generation size and the total number of generations which are required to find the global optimum are the two most frequent issues addressed in the community [16]. Generally speaking, the answer to these questions depends on the shape of the surface response, namely on its size and complexity [17]. The use of population sizes that match with the commercially available HTE equipment capacities, usually consisting of 8, 16, 24, 48 or 96 parallel vessels or racks, would result in the best experimental throughput (performing 1 generation per day). The optimization efficiency when using these population sizes is therefore studied for the COOX and SELOX case

studies. The next issue concerns the choice of the best algorithm settings, considering the practical limitations of dealing with a low population size capacity.

Optimization of algorithm configuration and settings is usually performed on so-called virtual or synthetic benchmarks which are custom-made mathematical response surfaces usually non-related to chemistry science. These functions are typically distinct from what one can expect for material optimizations case studies posing the question of their suitability for the optimization of GA parameters, since this is largely dependent on the response surface to optimize. Relevant GA optimization investigations would ideally be performed on extensive experimental data collections on heterogeneous catalysis. However, historical data collections from scientific literature usually provide data sets which are too small and include inconsistent and often poor data accumulated over decades. And on the other hand, comprehensive catalysis data banks that could be used for this purpose are mostly performed and kept confidential in industry research. Direct validation of GA configurations by directly performing experimental library design and analysis is usually not considered an alternative because it would lead to a too large experimental effort in order to get statistically significant results. A strategy based on experimental and simulation results using Artificial Neural Networks (ANN) has been applied, by Rodemerck *et al*, to construct a non-linear benchmark for the oxidative dehydrogenation of propane [16]. In this study, the effect of the GA population size was investigated but the results of the study relied on a very limited number of runs and key parameters such as the type of selection were not considered. In chapter 5 a Design of Experiments approach is adopted to build two catalysis benchmarks (the COOX and SELOX) which will be used in this chapter to perform a thorough validation of the influence of the GA configuration settings on the optimization performance on these benchmarks. General guidelines for optimum configuration settings are obtained and give indications about which parameter settings to use in similar optimizations of heterogeneous catalysts employing a GA approach.

6.2 - Genetic Algorithm implementation

A software platform named *Opticat* - “Optimization for catalysis” was used for applying the Genetic algorithms mentioned in this study. This platform has been designed and implemented by the CNRS institute (Lyon, France) and a brief description can be found in Appendix C. This platform enables to build a custom-made GA configuration by a versatile drag-and-drop system based on a diverse library of possible operators.

The data workflow describing a usual GA architecture is shown in Figure 6.1. An initial random population of catalyst candidate solutions is chosen and evaluated. Some candidate solutions of this generation are selected to undergo the crossover and mutation operations leading to a new generation of candidate solutions that will be tested and initiate a new iterative optimization loop.

.

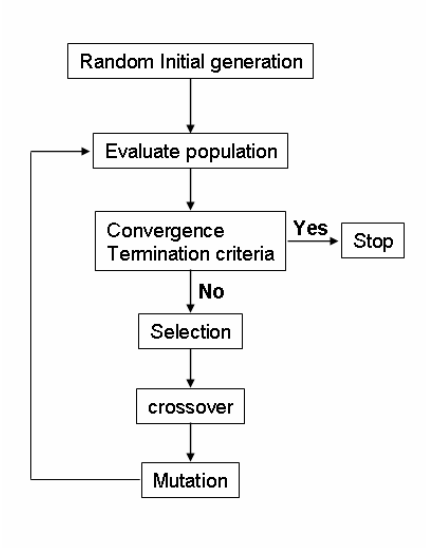


Figure 6.1 - Framework of a basic Genetic algorithm procedure

In this study the evaluation of the candidate solutions is made employing the COOX and SELOX benchmark fitness functions obtained in Chapter 5 and presented in Appendix B. Examples of the procedure of different modalities of the selection and crossover operators are illustrated in Figure 6.2 and 6.3 respectively.

The *selection operator* is usually based on probabilities. In the Wheel selection type, the probability of a catalyst to be selected as parents for the next generation is proportional to its fitness or performance (Figure 6.2a,c) whereas in the Ranking selection it relates to its fitness rank (Figure 6.2a,d). The Threshold selection value defines the percentage of low fitness performing catalysts that are accepted for the breeding process of the next generation of catalysts (Figure 6.2b). In Tournament selection, the individuals that are allowed to be reproduced, are chosen among a fixed random pool of catalysts from which the best one is selected. The selectiveness can be further tuned by the chosen *Selective Pressure (SP)*. The *Elitism operator* is a method, which forces a given number of the best performing catalysts to be always selected.

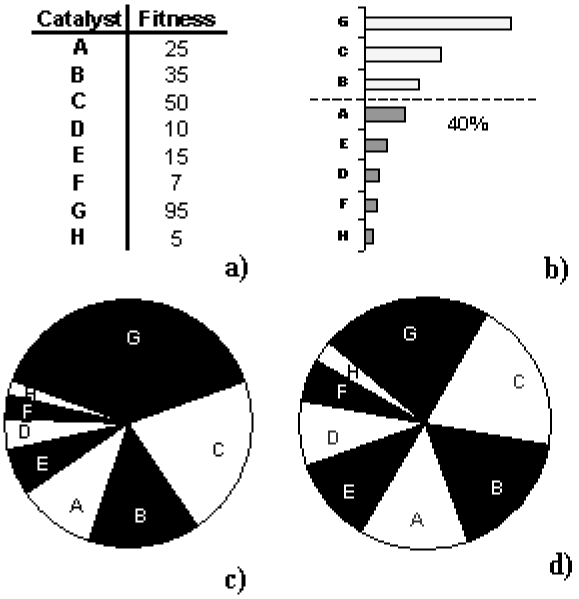


Figure 6.2 - Example of Selection type procedures: a) Observed fitness for a hypothetical 7 catalysts population, b) 40% Threshold selection c) Wheel selection probability, d) Ranking selection probability.

The *Crossover operator* mixes the genetic information of the selected individuals from the previous generation, creating new combinations with their genetic information. Multipoint and uniform crossover operator types are depicted in Figure 6.3.

The *Mutation operator* has an explorative role in GA. Due to the possible insertion of genetic information that was absent in the previous generations, new areas of the search space can be investigated, preventing the optimization procedure from being trapped into local optima. A representation of the mutation operator procedure is shown in Figure 6.4. Here also, the probabilities of strings to be modified by either crossover (CP) or mutation (MP) can be monitored by setting respective values. Detailed information on GA operators can be found elsewhere [4].

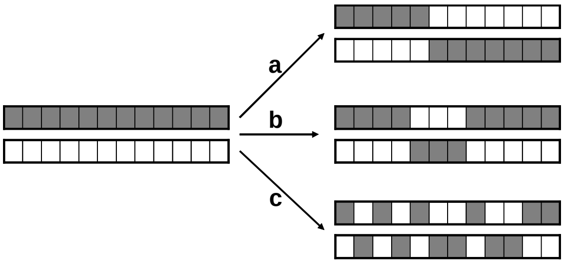


Figure 6.3 - Crossover Types: a) 1-Single Point Crossover, b) 2-Multipoint crossover, c) Uniform Crossover with 0.5 gene-flip probability

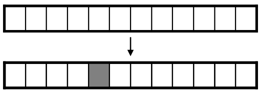


Figure 6.4 - Mutation Operator

6.2.1 - Catalyst representation

In GA the candidate solutions are usually coded as bit strings composed of binary numbers (0, 1) representing the controllable independent variables that may influence the response. In the present case the string was composed of 24 bits. Each of the 4 discrete variables (Temperature, Support, TM and NM types) was encoded in genes of 4 bits each. As indicated in Chapter 5, they can take 3 different modalities: Cu, Au and Pt for the main active metal (NM); V, Mo and Nb for the transition promoter metal (TM); Ce, Ti, Zr oxides for the Support; and T1, T2 and T3 for the tested temperature. The 2 continuous variables were also encoded as genes of 4 bits resulting in 16 steps of 0.06% and 0.25% for the NM (0.1-1.1%) and TM (1-5%) concentrations, respectively. Therefore a sub-space defined by a ternary at a given temperature encompasses 16^2

different catalysts compositions and the whole search space more than 20.000 experiments (3^4 x 16^2). An example for coding an experiment performed on a catalyst composition containing 1.4% Pt and 2.6 % Nb on TiO_2 and tested at 300 ºC is shown in Figure 6.5.

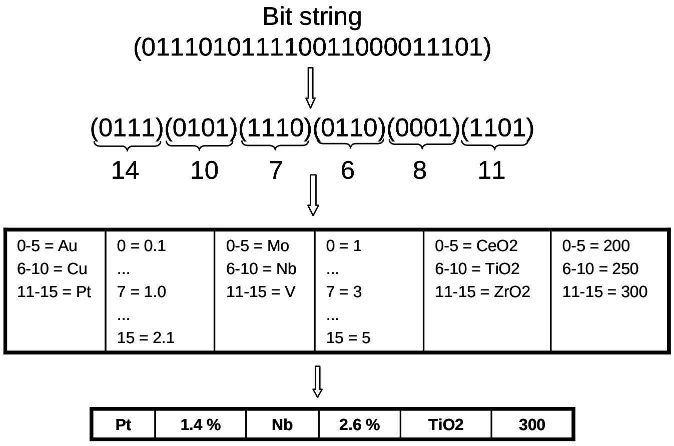


Figure 6.5 - String representation of 1.4 % Pt, 2.6 % Nb, TiO2, 300°C catalyst solution.

6.3 - Study of the effect of GA configurations on algorithm performances

Genetic Algorithms have both exploration and exploitation features. Best catalysts are selected and reproduced while browsing the rest of the search space. The right balance between the amount of diversity (exploration) and speed of convergence (exploitation) is crucial to a successful optimization. This balance is generally tuned by the choice of GA parameter settings during the GA implementation. This choice needs, however, to be performed a priori and can considerably influence the optimization efficiency.

The different GA parameter settings such as crossover, selection and mutation types and their corresponding percentage, the population size, and use of elitism operator were compared for their effect on the GA performance using the COOX and SELOX benchmarks. Information on the most influential parameters in the optimization of these benchmarks could in this way be obtained.

6.3.1 - Performance assessment

Keeping the same GA configuration and operator settings, the behaviour of an optimization procedure can still vary because of the initial population that is designed randomly, and also because of the random features introduced by the crossover and mutation operators. In order to get statistically significant results on the performance of the GA configurations, at least 40 replicate runs were monitored from which the average of the achieved mean and best fitness were computed. The quality of different GA configurations was accessed via a "*performance*" criterion defined as the percentage of successful runs after evaluation of a fixed number of catalyst solutions (192). A run is considered successful if, after evaluating 192 individuals, at least one catalyst is found displaying a fitness value in the considered optimality region: $D_{SELOX}>0.92$ and $D_{COOX}>0.99$ for the SELOX and COOX benchmarks, respectively. In turn, the defined quality assessment relates provides a measure of the robustness feature of the GAs and also the convergence speed since just a limited number of catalysts can be used (192). This criterion was considered more representative than the more commonly employed evaluation after a fixed number of generations since, in the latter case, the experimental effort necessary can vary greatly when configurations with different population sizes are compared. In this study the quality of each configuration was evaluated after a defined number of generations (Table 6.1) depending on the population size (Generation number = 192/Population size).

Table 6.1 - Number of generations at which the GA is evaluated in relation to the correspondent population size.

Population	Generations
8	24
16	12
24	8
48	4

6.3.2 - COOX and SELOX benchmark features

Although there exists no clear study specifying sources of difficulty in applying GA, some factors that have been suggested include: multi-modality (the existence of multiple misleading local or sub- optima), deception (when lower-order scheme information is misleading, causing the GA to get attracted to sub-optimal solutions), isolation (needle-in-the-haystack problem when no information exists to direct the optimization) and collateral noise (which hides the presence of a good sub-solution within a solution) [5]. The difficulty of the optimization of the COOX and SELOX response surfaces arises mainly by the presence of local optima with areas as large as or larger than the ones of the global optima. On the other hand, there are general trends such as the effect of the temperature, of supports (CeO_2>ZrO_2>TiO_2), of metal oxides (V>Nb>Mo) and of the active metals (Pt>>Cu>Au), which can be used as guidelines during the iterative loops of the algorithm to facilitate the achievement of the final optimum conditions (low deception). It is obvious that the assessment criteria which are chosen will depend on the shape of the benchmark and on the context [15]. Nevertheless, whatever the criteria and the benchmarks, the optimization of algorithm settings is usually performed by screening different configurations and parameters via trial aand error. The use of DoE methodology allows a more systematic approach to the study of how changing the algorithm parameters can influence its optimization performance. This strategy has been used for finding suitable GA configuration parameter settings for the optimization of molecular conformations [18] and will be used in this chapter to ascertain the best configuration settings to optimize the COOX and SELLOX benchmarks.

Due to the stochastic behaviour of GA, and the multiple optima present in the COOX and SELOX benchmarks, different optimal solutions were found at the end of the optimization cycles performed. The sub-spaces where usually the GAs optimization cycles converged are listed in Table 6.2, and the values indicate the probability of convergence to a particular sub-space or optimal region of the parameter space.

Table 6.2 - Composition of the sub-spaces were the optimal solutions were mostly found.

Optimum	COOX	%	SELOX	%
1	Pt/V/Zr/200 ° C	24	Pt/V/Zr/200 ° C	43
2	Pt/Nb/Zr/200 ° C	27	Pt/Nb/Zr/200 ° C	33
3	Pt/Nb/Ti/200 ° C	22	Pt/V/Ti/200 ° C	20
4	Pt/Mo/Ti/200 ° C	27	Cu/V/Ce/200 ° C	0
5	Pt/Mo/Zr/200 ° C	<1		

At the end of a GA optimization cycle one optimal solution is usually reached. As it can be noticed in table 6.2 some optimum solutions are achieved via the optimization cycles more frequently than others, namely the ones positioned at the sub-spaces Pt/Mo/Zr/200°C for COOX and Cu/V/Ce/200°C for SELOX. These probabilities match well with the size of the areas that inside the sub-spaces are at the optimum plateau values. Indeed, when looking at the best sub-spaces in the benchmarks reported in Table 6.2 it can be seen that, in most cases, the performance is at the global optimum value considered (D_{SELOX}>0.92 and D_{COOX}>0.99) whatever the metals concentration composition, resulting in a large plateau covering almost entirely the sub-space (Chapter 5, Figure 5.5 and 5.6). This is not the case for Pt/Mo/Zr/200°C for COOX and Cu/V/Ce/200°C for SELOX for which only a minor area (top of peaks) reaches the optimum. Therefore, it is more probable that GAs converge for large optimal areas rather than for small ones, which is in line with the obtained results.

6.3.3 - Effect of GA configuration on the optimization performances on COOX and SELOX

Different GA parameters for the Elitism, Crossover, Population size and Selection type were compared for their effects on the GA performance, using the COOX and SELOX benchmarks. The different modalities studied for each parameter are depicted in Table 6.3.

The combination of each different modality represents a given GA configuration. The total number of combinations results in 128 different algorithm configurations. Sixteen configurations were selected by a D-Optimal experimental design. By using this design the main effects of each operator modality on the performance of GA optimization could be studied with a reduced set of experiments.

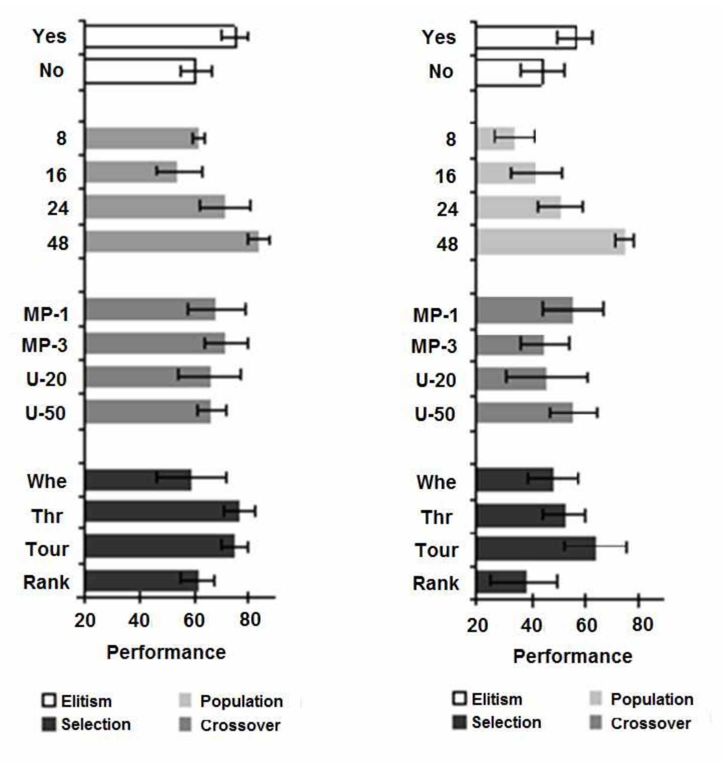
Table 6.3 - Genetic operators and respective modalities.

Elitism	Population	Selection type	Cross-over type
Yes	8	Wheel (Whe)	1- Point (MP-1)
No	16	Ranking (Rank)	3- Point (MP-3)
	24	Threshold (Thr)	Uniform 20% (U-20)
	48	Tournament (Tour)	Uniform 50% (U-50)

Constant parameters: Selection Pressure = 1, Crossover Probability = 70%, Mutation Probability = 1%, 15% Elitism, 3 individuals per Tournament, 40% Threshold.

The main effects of the GA settings calculated by regression from the 16 different GA configurations are shown in Figure 6.6. The bar chart represents the weighted average effects for each GA parameter with the corresponding standard error, while the performance is the percentage of successful runs as defined in section 6.3.1.

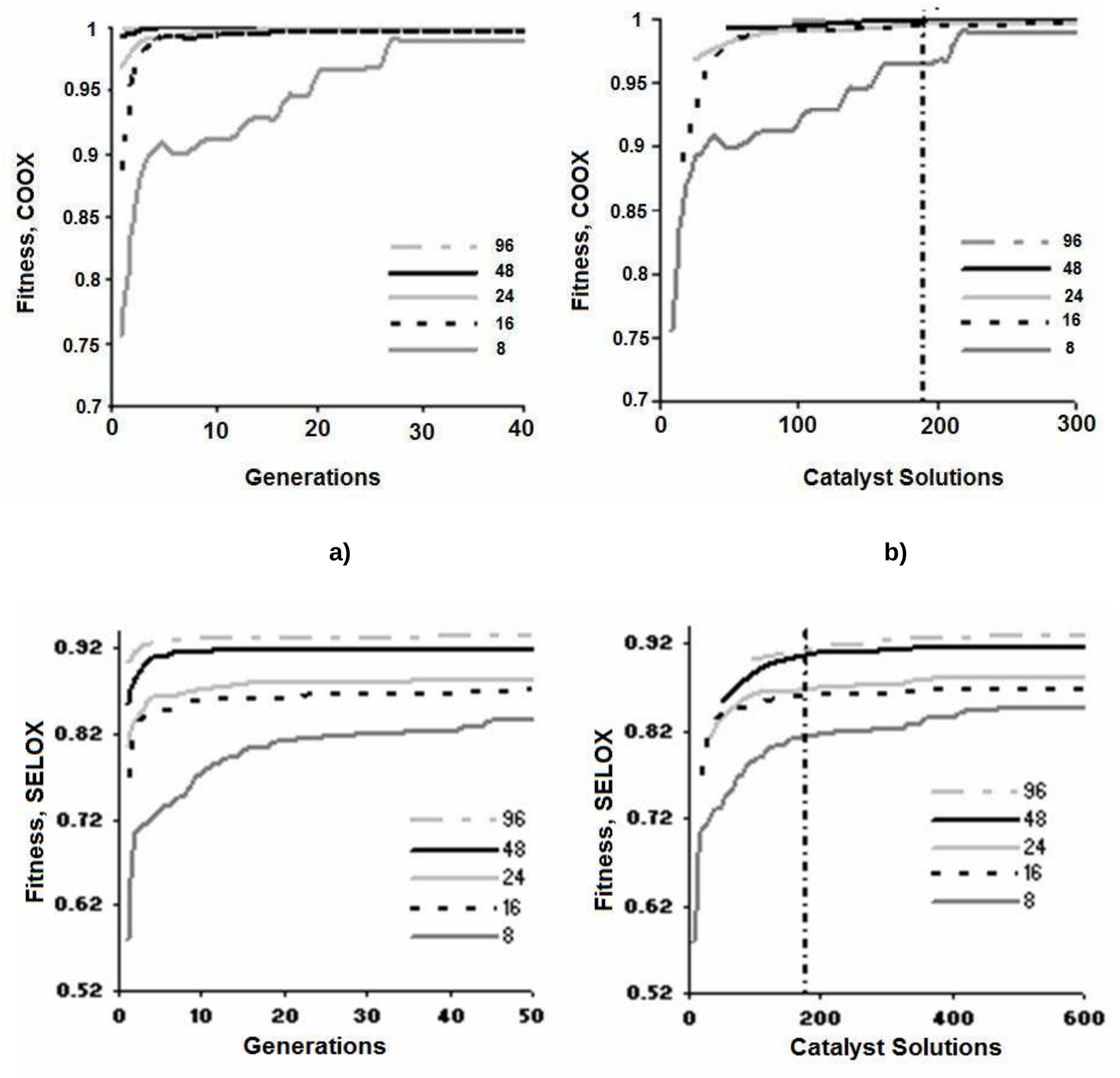
As can be seen in Figure 6.6, the GA performance was favoured independently of the benchmark when using the Elitism operator, large population size and Tournament selection. The population size parameter showed the largest effect for both benchmarks, although the number of final tested catalysts was the same. The larger the population was, the more robust the optimization process proved to be. In addition, a pronounced decrease of the standard error was observed when the population size was the highest (48) for both benchmarks. This means that, whatever the other parameters, a more reliable optimization procedure can in principle be obtained when using a large population size. The performance of GA depends also on the way the selection is carried out. The performance appears to be better when the selection included a large percentage of the very best catalysts. Indeed, when the “elitist” mode was used (systematic selection of the 15% top catalysts), the performance optimization performance was significantly better. On the other hand it was noticed that the different crossover operator parameters did not influence significantly the optimization performance in both benchmarks.



a) **b)**

Figure 6.6 - Mean operators performance for a) COOX and b) SELOX benchmarks. The lines represent standard error. Elitism (yes, no); Population size (8, 16, 24 and 48 individuals) Selection types: Wheel (Whe), Ranking (Rank), Threshold (Thr) and Tournament (Tour); Crossover types: 1 and 3 Multi-point crossover (MP-1, MP-3), 20% and 50% Uniform crossover (U-20, U-50).

From this study, it comes out that the best GA configuration consists in using the Elitism operator, tournament selection and 1-point crossover (MP-1). The typical optimization behaviour calculated over 40 runs of this best configuration is shown in Figure 6.7.



a) **b)**

c) **d)**

Figure 6.7 - Optimization profile for the selected configuration (Elitism, Tournament selection and 1-Point crossover), for several different population sizes. a) and b) COOX, c) and d) SELOX. The vertical dash-dot line indicates the moment of the optimization iterative loop at which the GA performance was evaluated.

The plots correspond to the mean of the best catalysts of a population in the course of the optimization and report the effect of the population size as a function of the generation number and as a function of the number of catalysts tested. It can be seen that for the COOX benchmark the optimization is easier and a lower population number is enough to provide efficient optimization than for the SELOX benchmark. For the

COOX benchmark, with a population higher than 16 individuals, convergence is reached before all 192 catalysts have been tested, whereas in the SELOX benchmark a population size above 24 catalysts is required. For both reactions a population of 8 individuals performs poorly.

As shown in Figures 6.6 and 6.7, the population size is the parameter with the highest impact in GA optimization performance. When the population is small, the diversity in the initial random population is also small and the probability to be trapped in a local maximum increases.

When using a large population size a higher diversity already exists in the initial random population and the best solution in the population is expected to be already close to the global solution (Figure 6.7). For smaller population sizes, if the initial population does not contain the genetic information for optimal solutions, we have to rely on the stochastic behaviour of the crossover and mutation operators to insert the desired information into the population during the new generation breeding steps. This requires a higher number of generations in order to achieve optimality (Figure 6.7). This confirms that there is a minimum threshold population size below which the GA optimization has difficulties to operate. It can also be seen (Figure 6.7b) that above a certain population size no further improvement is achieved in the optimization.

As demonstrated by comparing the two reactions, the convergence in the SELOX benchmark was found more difficult than in the COOX one, due to the presence of more local optima (Figure 6.7). For the former, a higher population size was required. The use of elitism prevents the possibility of losing good catalyst solutions during the selection stage, increasing the convergence speed. No significant difference in the effect of crossover type was observed.

6.3.4 - Optimization of operators for the SELOX benchmark

A further study of the GA operators was done with respect to the selective pressure (SP), mutation probability (MP) and crossover probability (CP) based on the full factorial design of the parameters presented in Table 6.4.

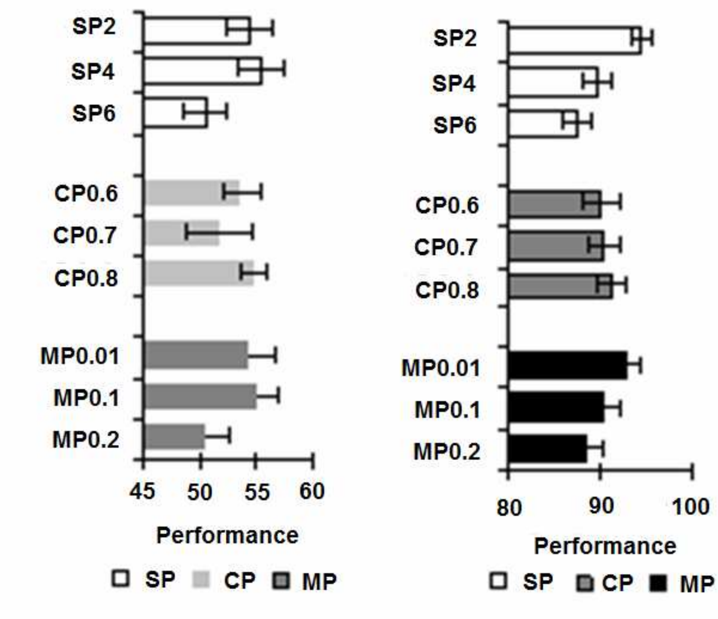
Table 6.4 - Studied levels of the Selective pressure (SP), mutation probability (MP) and crossover probability (CP).

Selective Pressure*	**Crossover Probability (%)**	**Mutation Probability (%)**
2	60	1
4	70	10
6	80	20

* Number of candidate solutions per tournament.

For the tournament selection, changing the number of individuals picked from a shuffled population from 2 to 6 increases the selective pressure and decreases the diversity. The effect of changing SP, and the probability of an individual to be subject to crossover (60-70-80 %) and mutation (1-10-20%) operators, for populations of 16 and 48 catalyst solutions on the SELOX benchmark is presented in Figure 6.8a and b, respectively. From this figure it can be seen that the range of crossover probabilities studied does not have a significant effect for population. The performance for the 16 individuals' population remains rather low and cannot be tuned within the tested number of individuals to achieve desired conversion values of more than 90%. In contrast, for the population of 48 individuals (Figure 6.8.b) the desired performance is obtained and a SP of 2 and MP 0.01 are the preferred settings.

The different nature of the diversity originated by the selective pressure and mutation probability operators explains the results obtained in Figure 6.8. In the first case the selective pressure (SP) will enable the already existent genetic information of the population to pass to the next generation while the diversity originated by the mutation probability (MP) can modify this genetic information by introducing genes, not previously present in the population, from the rest of the unexplored search space. A low selective pressure (SP-2) keeps a good level of diversity for future generations breeding, by decreasing in this way the probability of premature convergence to local optima. Clearly a low Selective pressure is preferred for both populations. With a large population containing already good genetic information (Figure 6.8b) the GA will perform better with low mutation values, while with a smaller population a higher mutation can be necessary to introduce good genetic diversity (Figure 6.8a). However, extreme values of mutation (MP 0.2) can be prejudicial in both cases since it can slow down the convergence rate by destroying already-found relevant information.



a) **b)**

Figure 6.8 - Mean effects for the SP, CP and MP. Constant parameters: Elitism 15%, Tournament selection and 1 Point crossover. a) Results for a 16 individual population size, b) Results for a 48 individual population size.

The above features show that for an efficient GA run a large variety in the initial sample of the search space is of utmost importance. It is nevertheless obvious that, from a practical point of view, a compromise has to be found by considering the size limitation arising from high-throughput equipment constraints.

The optimization profile for the population of 48 individuals with SP of 2 and MP 0.01 is shown in Figure 6.9. With this configuration a performance of 97% according to the established criteria in section 2.6 is attained. This means that in almost all the simulated runs an optimal catalyst solution is achieved in less than 4 generations or 192 catalyst candidate solutions evaluated.

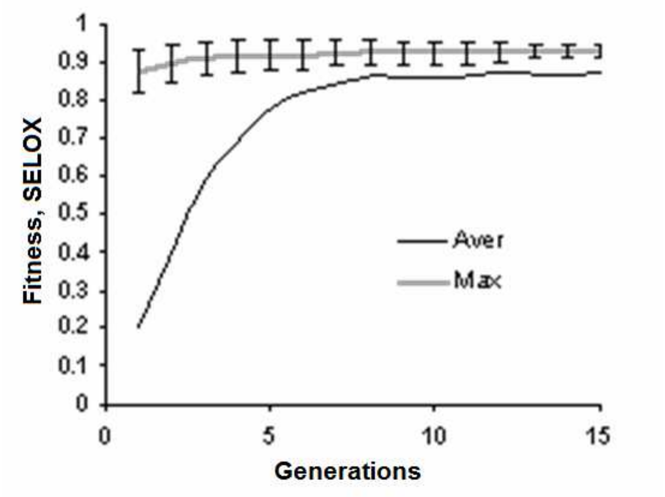


Figure 6.9 - Optimization profile for the settings: 15 % Elitism, 48 individuals population size, Binary Tournament, 1 Point crossover with 60% crossover probability and 0.01 Mutation on the SELOX benchmark.

6.5 - Conclusions

Due to the large influence of parameter setting in the optimization efficiency of GA and the usual difficulty or even impossibility to validate the optimization results via repeated optimization cycles (due to the large experimental effort necessary) the settlement of a proper algorithm parameter configuration in the beginning of an experimental GA optimization is highly important. These configurations are however dependent on the response surface to optimize and this knowledge is usually not known a priori. The construction of benchmarks can allow the optimization of GA settings through simulation of its optimization performance, giving an indication of which sort of configurations would be more adapted for certain catalyst optimization studies.

The use of Design Experiments showed to be an efficient strategy for obtaining a systematic procedure to investigate the effect of changing the Genetic Algorithms settings on its optimization performance. The main trends on the use of the selected parameters were obtained and optimal GA configuration settings for the optimization of the COOX and SELOX benchmarks were achieved.

The results of the GA parameter optimization highlight that a large variety in the initial sample of the search space is crucial. It is nevertheless obvious that, from a practical point of view, a compromise has to be found by considering the size limitation arising from high-throughput equipment constraints and efficiency of workflow.

The optimal GA configuration found for the optimization of the COOX and SELOX benchmark consisted of using as parameter settings: 15 % Elitism, 48 individuals population size, Binary Tournament, 1 Point crossover with 60% crossover probability and 1% Mutation. These settings can give an indication of the parameters to be adopted for investigating similar catalytic response surfaces. When HT equipment for catalyst screening will have become more conventional in academic laboratories, access to databases may facilitate the study of algorithm optimization processes in other catalytic related studies.

6.6 - References

[1] J. Holland, *Adaptation In Natural and Artificial Systems*, Ann Arbour, **1975**.

[2] L. Davis, *Handbook of Genetic Algorithms*, Van Nostrand Reinhold, **1991**.

[3] W. F. Maier, K. Stöwe and S. Sieg, *Angew. Chem. Int. Ed.,* **2007**, 46, 6016.

[4] D. Renner and A. Ekart, *Comp-Aided Des.,* **2003**, *35*, 709.

[5] D. L. Massart, B. G. M. Vandeginste, L. M. C. Buydens, S. De Jong, P. J. Lewi and J. Smeyers-Verbeke, *Handbook of Chemometrics and Qualimetrics*, Elsevier, **1997**, Vol. 20.

[6] D. Wolf, O. Buyevskaya, M. Baerns, U. Rodemerck and P. Claus, in *PCT Int. Appl.*, (Institut Für Angewandte Chemie, Berlin-Adlershof E.V., Germany), Wo 0015341, **2000**, 35.

[7] D. Wolf, O. V. Buyevskaya and M. Baerns, *Appl. Catal. A: General,* **2000**, 63.

[8] D. Wolf, in *Principles and methods for accelerated catalyst design and testing* (Eds.: E. G. Derouane, V. Parmon, F. Lemos, F. R. Ribeiro), Kluwer Academic Publishers, Dordrecht, Netherlands, **2002**, pp. 125.

[9] O. V. Buyevskaya, D. Wolf and M. Baerns, *Catal. Today,* **2000**, *62*, 91.

[10] O. V. Buyevskaya, A. Bruckner, E. V. Kondratenko, D. Wolf and M. Baerns, *Catal. Today,* **2001**, *67*, 369.

[11] U. Rodemerck, H. Kosslick, G. U. Wolf, S. Kolf and M. Baerns, *Chem. Ing. Tech.,* **2002**, *74*, 559.

[12] W. F. Maier, A. Frantzen, G. Frenzer, G. Kirsten, J. Saalfrank, J. Scheidtmann, B. Stutz, C. Thome, P.-A. W. Weiss and T. Wolter, *Polym. Mater. Sci. Eng.,* **2004**, *90,* 652.

[13] J. Serra, A. Corma, D. Farrusseng, L. Baumes, C. Mirodatos, C. Flego and C. Perego, *Catal. Today,* **2003**, *82*, 67.

[14] A. Corma, J. M. Serra and A. Chica, in *Principles and methods for accelerated catalyst design and testing* (Eds.: E. G. Derouane, V. Parmon, F. Lemos, F. R. Ribeiro), Kluwer Academic Publishers, Dordrecht, Netherlands, **2002**, pp. 153.

[15] R. Wehrens, E. Pretsch and L. M. C. Buydens, *J. Chem. Inf. Comput. Sci.,* **1998**, *38,* 151.

[16] U. Rodemerck, M. Baerns, M. Holena and D. Wolf, *Appl. Surf. Sci.*, **2004**, 168.

[17] K. Deb and S. Agrawal, in *Foundations of Genetic Algorithms 5*, (Eds.: W. Banzhaf, C. R. Reeves), **1998**, Morgan Kaufmann, United States of America, pp. 264.

[18] R. Wehrens, E. Pretsch and L. M. C. Buydens, *Anal. Chim. Acta,* **1999**, *388*, 265.

7

COMPARISON OF OPTIMIZATION METHODOLOGIES IN THE SELOX BENCHMARK

Abstract

The optimization efficiencies of several Global Optimization 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 optimization 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 optimization.

The contents of this chapter have been published in:
S. R. M. Pereira, F. Clerc, D. Farrusseng, J. C. van der Waal, T. Maschmeyer, Optimisation Methodologies and Algorithms for Research on Catalysis Employing High-Throughput Methods: Comparison Using the Selox Benchmark, *Comb. Chem. & High Through. Screen.*, 10, 2007, 149.

7.1 – 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 parallelisation 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 optimization [5]. A method starting to be commonly used within the combinatorial catalysis field is the Design of Experiments (DoE) explained in Chapter 2. Applications include catalyst formulation and preparation [6-9], catalytic kinetic modelling [10,11] reactor engineering [12,13], and the optimization 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 optimization 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 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 synthesised and tested. Catalyst screening and optimization is a tedious task even with the help of automated parallel equipments. As a consequence, the optimization procedure should be reliable and should minimise the risks of failures (e.g. optima not found). This represents a substantial increase in the effort required to assess 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 optimization [22]. 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 optimization algorithms using catalytic benchmarks. For instance, Genetic Algorithms search procedures have been studied in virtual catalytic benchmarks in the form of a Neural Network [23-26], 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.

In this chapter a general overview of the application of common chemometric optimization methods in the field of catalyst optimization is presented. Several global optimization algorithms are compared to each other using the same mathematical function, or benchmark, which derived from a dataset on the SELOX reaction (Chapter 5). Finally, Design of Experiment methodology is also studied using the same SELOX benchmark.

7.2 - 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 (Chapter 5) [27]. 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 main active metal (NM, [NM]%) acting as the main active phase; the type and amount of transition metal (TM, [TM]%) acting as modifier/promoter of the active phase, and the type of metal oxide acting as support (see Table 7.1). By modelling, for every combination of the remaining discrete parameters, the effects of varying [NM%] and [TM%] on the fitness function, 81 quadratic order functions were obtained - one for each of the conversion and selectivity responses (see Figure 7.1 and Appendix B). 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 [27]) The SELOX benchmark has the advantage of being composed of a series of mathematical functions and, therefore, it can be easily optimized 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 considered optimal areas (fitness > 0.92) are situated in the Pt-Zr-Nb and Pt-Ti/Zr-V at 200 ºC sub-regions of the response surface (see Figure 7.1).

Table 7.1 - Parameter Space for the SELOX Benchmark

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

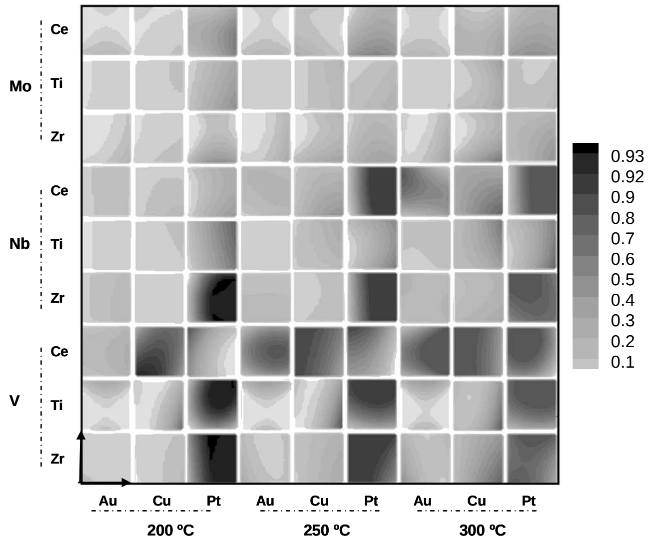


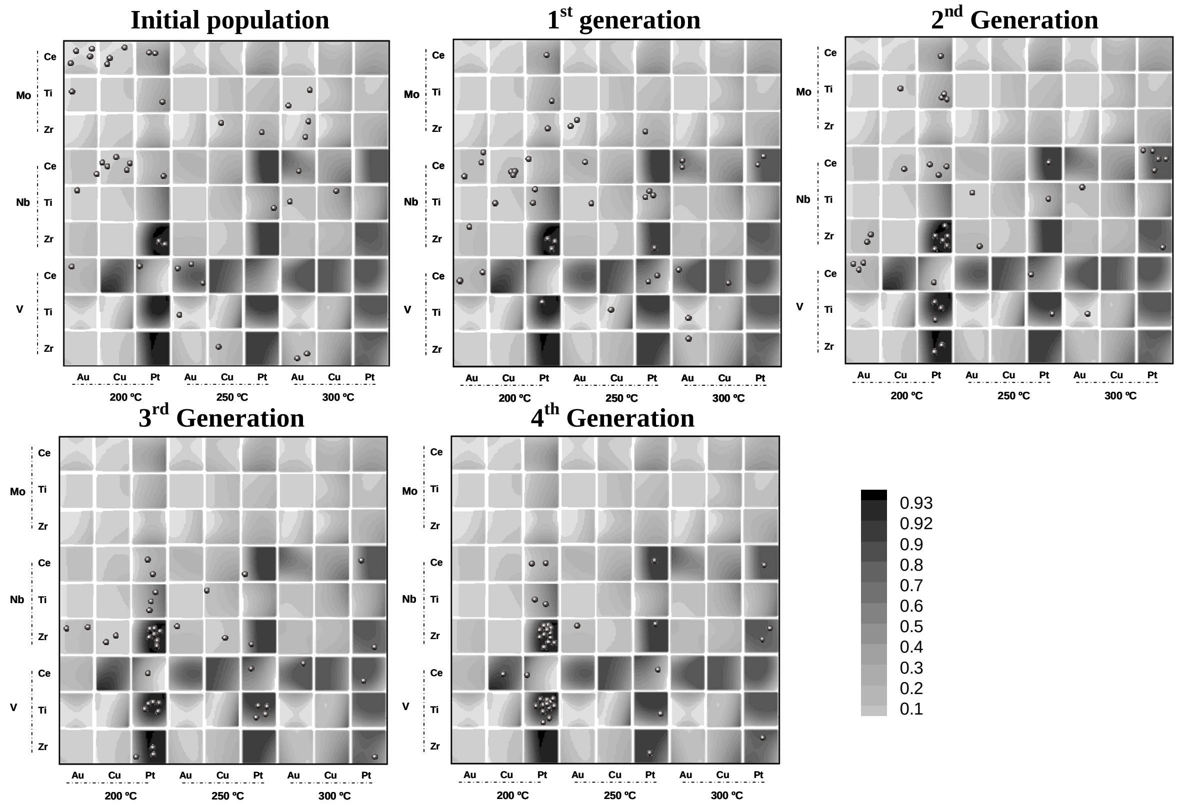
Figure 7. 1 - Contour Plot representation of the total desirability response surfaces of the SELOX benchmark.

7.3 - Global Optimization algorithms

A common methodology for the optimization of catalytic problems is the use of global optimization strategies. Unlike, for example, Simplex and Gradient Descent local optimization methodologies, which are deterministic algorithms, stochastic global optimization algorithms do not, as long as the necessary number of iterations is performed, get trapped in local optima. Among the global methodologies available, currently Evolutionary Algorithms (EA) represents the most popular method which has found a wide range of applications in chemistry [28]. After the introduction, by Wolf *et al.*, of this optimization strategy in heterogeneous catalyst formulation, the interest in this approach has risen considerably [24,25,30-34]. 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 optimization 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) [35-37], and Taboo Search (TS) [38], 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 [39], Kriging [40], Holographic [26,41] search, etc. In the following section, a brief set of explanations regarding search algorithms to be tested on the SELOX benchmark is presented.

7.3.1 - 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 [42,43]. 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 [44]. 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 [27]. Results obtained for the SELOX benchmark with a GA algorithm are shown in Figure 7.2 and the workflow chart for a GA and an ES are presented in Figure 7.3a.



Initial population

1st generation

2nd Generation

3rd Generation

4th Generation

Figure 7.2 - Scatter-plot of the initial random population and the subsequent 4 generations on the SELOX benchmark for the GA_1 algorithm settings (see Table 7. 2).

7.3.2 - 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 [45,46]: the state of the thermodynamic system is analogous to the candidate solution, *e.g.* the catalyst, and the energy of the state is 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 neighbouring 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 optimization (CT, Cooling Temperature scheme) controls the degree of system browsing and the algorithm's optimization speed. The workflow chart for a SA procedure is presented in Figure 7.3b. In this study the simulated annealing tested is a classic single-candidate algorithm. However, parallel SA optimizations that consider, for the same temperature, several neighbourhood candidate solutions have already been implemented [47].

7.3.3 - 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 [48]. 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 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 [49]. Taboo Search is a deterministic algorithm. The TS workflow chart is presented in Figure 7.3c.

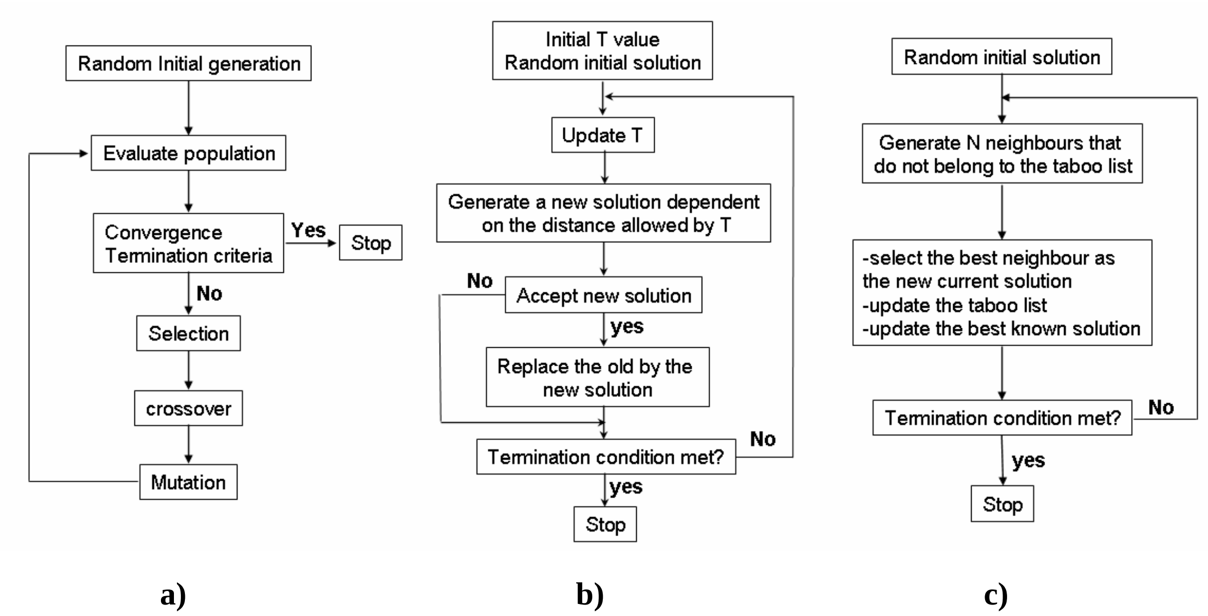


Figure 7.3 - Flow charts of the algorithm strategies. a) Genetic/Evolutionary Algorithm, b) Simulated Annealing (T = temperature), and c) Taboo Search algorithm.

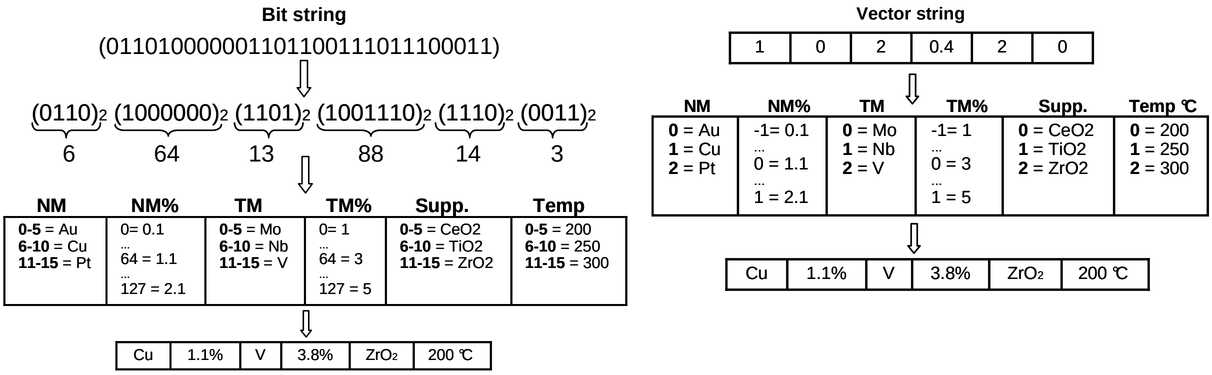
7.3.4 - Hybrid algorithms

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 [22,50,51] 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 estimate 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.

7.4 - Comparison of Global Optimization algorithms using the SELOX benchmark

7.4.1 - 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 Figure 7.4a. Vector coding representation of the variables was used for the Evolutionary Strategies as described in Figure 7.4b.



a) **b)**

Figure 7.4 - 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.

7.4.2 - Algorithm settings

Prior to the experimental optimization the algorithm settings need to be defined. This will influence the speed of optimization and browsing by the algorithm. The efficiency of the optimization is dependent on whether the settings adopted are appropriate for the problem to optimize [27]. 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 $SA_{3;}$ see Table 7.2 and Figure 7.5), which in this case defines the neighbourhood size change allowed at each iterative step, controlling the browsing/exploration speed of the optimization.

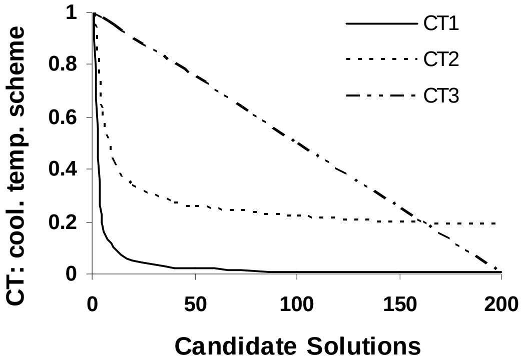


Figure 7.5 - Simulated annealing cooling temperature schemes (CT). (see Table 7.2)

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 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, so as to induce the simultaneous search of different optima [58]. The sharing property (Equation 7.1) uses a similarity operator, *sim*(*a*,*b*), (Equation 7.2), 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 neighbourhood, 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 (Equation 1 and 2).

$$f'(a) = \frac{f(a)}{\sum_{i=1}^{n} sim(a, b_i)}$$ **Equation 7.1**

$$sim(a,b) = \left(\frac{G}{d(a,b)} \right)^{\alpha}$$ **Equation 7.2**

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), G a neighbourhood parameter and *n* the number of candidate solutions from the current generation.

For the SA and TS, and any other 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.

7.4.3 - 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 (Appendix C) was used for performing all the algorithmic optimizations [60]. The results obtained are presented in Table 7.2.

7.4.4 - Results and discussion

The algorithms used for optimizing the SELOX benchmark response surface and the results obtained are presented in Table 7.2. From this table 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 optimization is performed. The poor performance of the SA search is likely to be related to having only a single candidate per iterative step (see Table 7.2). When starting from a bad candidate solution it can be more difficult to find the right track and the optimization 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 [47], was not within the scope of the present investigation but would be worthwhile to evaluate in future studies.

Table 7.2 - Optimization algorithms, their settings and optimization 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
Evolution. Algorithm	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.

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 optimization algorithms, as the name indicates, are usually good in finding the global optimal solution area. They are, however, not the best for fine local optimization 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 optimization procedure to overcome this limitation is to employ local optimization deterministic methods, like Simplex or Gradient Descent, after the global algorithm optimization.

Evolutionary Strategies (ES) performed in general better than the Genetic Algorithms (GA), indicating that the representation 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 optimization. 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 optimization [27]. The use of the Sharing property (Equation 7.1 and 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 optimization. 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.

7.5 - DoE strategy for the SELOX benchmark optimization

Design of Experiments (DoE) aims at maximising the amount of information obtained from experimentation while minimising the number of experiments (Chapter 2). DoE uses regression techniques to obtain the relationship between the response surface and the system factors in order to obtain a 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 optimization strategy, using the *Nemrodw 2000* software, is discussed below.

7.5.1 - 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 data-points are demanded. For the parameter space presented in Table 7.1 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 optimization 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 Figure 7.6 and the results obtained for the main factor effects study are depicted in Figure 7.**7**.

The estimate of the main effect of a factor is related to the average response values of the experiments performed at various levels of the factor. 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 Figure 7.7. 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 (in Chapter 5 the poor performance of the Au catalysts is explained in terms of the low metal dispersion resulting from the preparation method). 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 (Figure 7.6)).

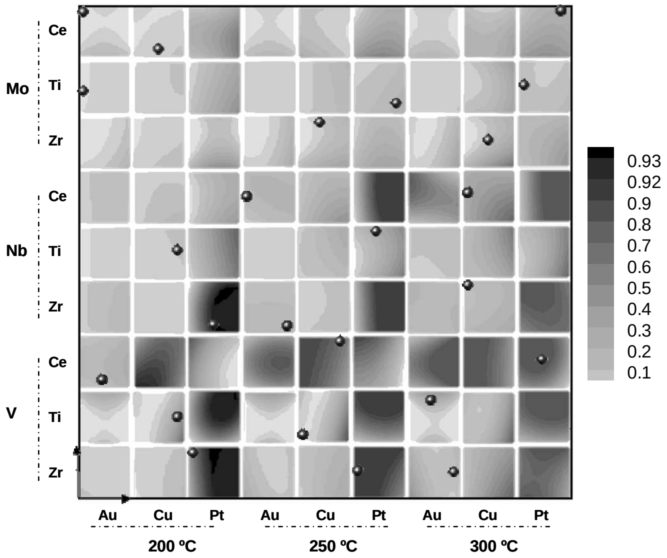


Figure 7.6 - Representation of one of the possible designs in the SELOX search space.

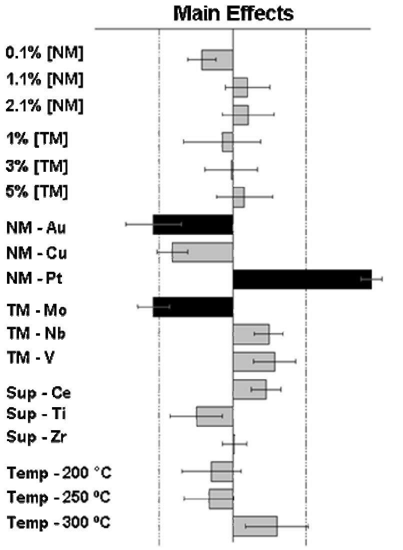


Figure 7.7 - Average main factor effects with 95% confidence interval and their variance. The bars directed to the right signify a positive relative effect and those to the left a negative one. The dotted lines represent the 95% confidence interval calculated from the estimated experimental variance. Effects higher than this confidence level are considered statistically significant and are represented in black.

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 determinant 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.

7.5.2 - Screening for interaction effects

In this second design the interaction effects of the discrete variables Support, TM and Temp were studied (see Table 3). By opting to perform fine-tuning optimization 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 (see Figure 7.8) could then be chosen comprising 18 experiments.

Table 7. 3 - Factors to be investigated in the second screening and their correspondent levels.

TM	Support	Temp (ºC)
V	CeO_2	200
Nb	ZrO_2	250
	TiO_2	300

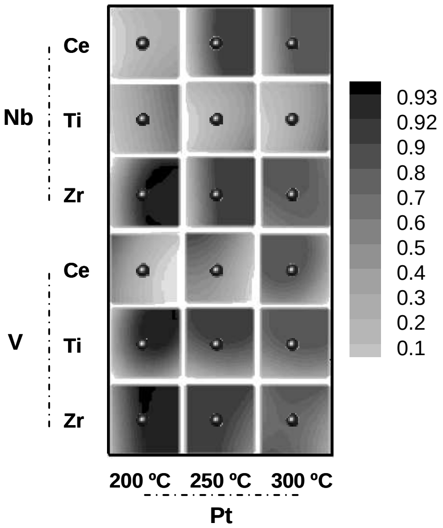
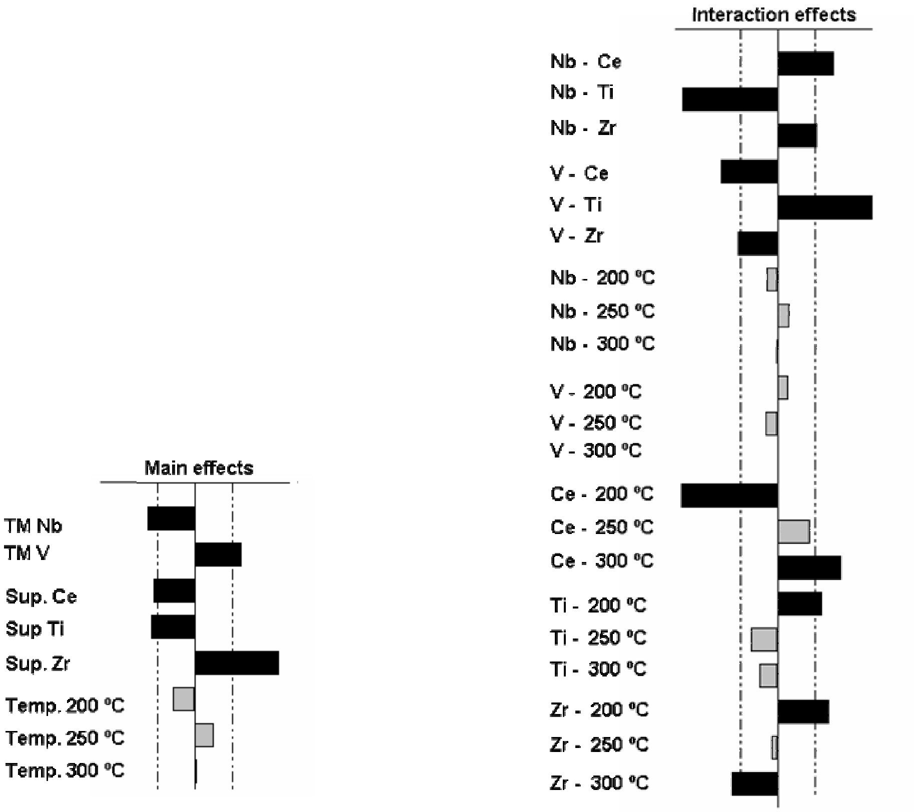


Figure 7.8 - Representation of the reduced search space after the main factors analysis, and the new experimental design points

The results obtained for the two-factor main and interaction effects are depicted in Figure 7.9. A detailed analysis of the main effects for this reduced search space shows 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.



a) **b)**

Figure 7.9 - Statistical effect analysis for the second screening. a) Main effects. b) Interaction effects.

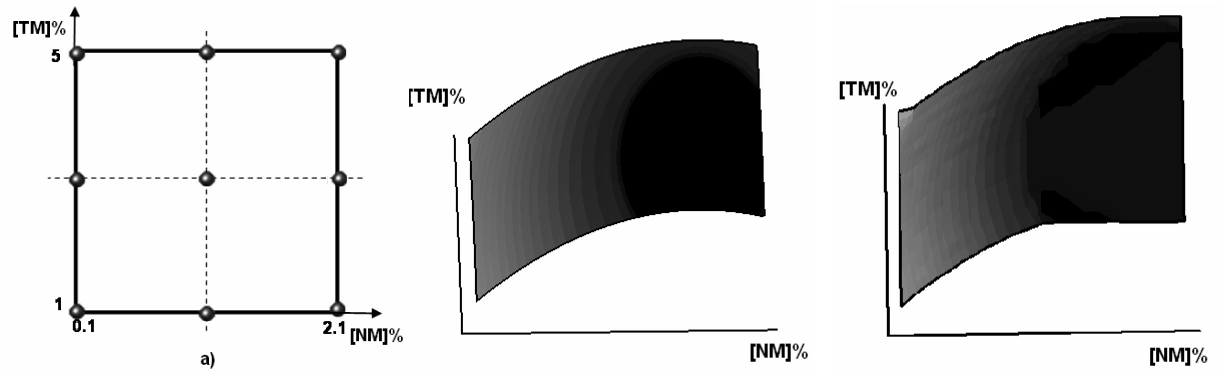
The interaction Ti-V is very strong, and there is a difference between the Nb-Zr and V-Zr interactions, where the Nb-Zr interaction has a relative positive effect and the V-ZR a negative. These trends can be confirmed by comparing them with the response surface in Figure 7.8. At this stage of the optimization 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.

7.5.3 - Optimization design

The final optimization step is exemplified with the selected V-Ti interaction at 200 ºC; in this final step the 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 modelling experiments are much larger than screening experiments and are only performed for a few factors. To model the curvature of the response surface a second-order model is considered to be sufficient. A Central Composite design [3], Figure 7.10a), was chosen to plan the experimental points to be performed. For the efficient modelling of the response surface seven 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 Figure 7.10b).

In this response surface (Figure 7.10b) the representative curvature of the benchmark surface (Figure 7.10c) 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.



a) **b)** **c)**

Figure 7.10 - Composition optimization 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 catalyst compositions.

7.6 - 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 optimization 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 optimization. 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 optimization procedure on existent benchmark studies.

If the parameter space to be investigated is very large, Design of Experiments may not constitute the most efficient strategy. The research can be divided into 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 optimization 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 optimization is the achievement of the optimal solution. For small search spaces many of the global optimization 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 optimization 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.

7.7 - References

[1] I. E. Maxwell, P. van den Brinck, R.S. Downing, A.H. Sijpkes, S. Gomez and T. Maschmeyer, *Top. Catal*., **2003**, 24, 125.

[2] P. P. Pescarmona and J. C. van de Waal, *Catal. Lett*., **1999**, 63, 1.

[3] D. L. Massart, B. G. M. Vandeginste, L. M. C. Buydens, S. De Jong, P. J. Lewi and J. Smeyers-Verbeke, *Handbook of Chemometrics and Qualimetrics: Part A and B*, Elsevier, Amsterdam, **1997**.

[4] R. G. Brereton, *Chemometrics : Data Analysis for the Laboratory and Chemical Plants*, John Wiley & Sons Ltd., Bristol, UK, **2003**.

[5] A. Holzwarth, P. Denton, H. Zanthoff and C. Mirodatos, *Catal. Today*, **2001**, 67, 309.

[6] E. A. Dawson and P. A. Barnes, *Appl. Catal. A*, **1992**, 90, 217.

[7] B. M. Weckhuysen, A. A. Verberckmoes, J. Debaere, K. Ooms, I. Langhans and R. A. Schoonheydt, *J. Mol. Catal. A*, **2000**, 151, 115.

[8] M. Tagliabue, L. C. Carluccio, D. Ghisletti and C. Perego, *Catal. Today*, **2003**, 81, 405.

[9] M. L. Bricker, J. W. A. Sachtler, R. D. Gillespie, C. P. McGonegal, H. Vega, D. Bem and J. S. Holmgren, *Appl. Surf. Sci.*, **2004**, 223, 109.

[10] M. M. Barsan and F. C. Thyrion, *Catal. Today***, 2003**, 81, 159.

[11] W. G. Hunter and R. Mezaki, *Can. J. Chem. Eng*., **1967**, 45, 247.

[12] A. Nordin, L. Eriksson and M. Ohman, *Fuel*, **1995**, 74, 128.

[13] J. Sjöblom, K. Papadakis, D. Creaser and C. U. I. Odenbrand, *Catal. Today*, **2005**, 100, 243.

[14] R. Carlson, *Data Handling in Science and Technology* , Vol. 8 : Design and Optimisation in Organic Synthesis, Elsevier, Umeå, Sweden, 1992.

[15] J. C. McWilliams, D. R. Sidler, Y. Sun and D. J. Mathre, *Applying statistical design of experiments and automation to the rapid optimisation of metal-catalysed processes in process development*, Merck, JALA, 2005, 10, 394.

[16] R. Carlson, *Chemom. Intell. Lab. Syst.*, **2004**, 73, 151.

[17] F. Stazi, G. Palmisano, M. Turconi and M. Santagostino, *Tetrahedron Lett.*, **2005**, 46, 1815.

[18] L. Veum, S. R. M. Pereira, J. C. van der Waal and U. Hanefeld, *Eur. J. Org. Chem*., **2006**, 7, 1664.

[19] A. R. Leach and V. J. Gillet, *An Introduction to Chemoinformatics*, Kluwer Academic Publishers, London, **2003**.

[20] C. Klanner, D. Farrusseng, L. Baumes, C. Mirodatos and Schüth, F. *QSAR Combi. Sci.*, **2003**, 22, 729.

[21] S. Sieg, B. Stutz, T. Schmidt, F. Hamprecht and W. F. Maier**,** *J. Mol. Mod.*, **2006**, 12, 611.

[22] F. Clerc, M. Lengliz, D. Farrusseng, C. Mirodatos, S. R. M. Pereira and R. Rakotomalala, *Rev. Sci. Inst.*, **2005**, 76(6).

[23] U. Rodermerck, M. Baerns, M. Holena and D. Wolf, *Appl. Surf. Sci*., **2004**, 223, 168.

[24] D. Wolf, O.V. Buyevskaya and M. Baerns, *Appl. Catal. A :Gen*, **2000**, 200, 63.

[25] Y. Watanabe, T. Umegaki, M. Hashimoto, K. Omata and M. Yamada, *Catal. Today*, **2004**, 89, 455.

[26] A. Tompos, J. Margitfalvi, E. Tfirst and L. Végvári, *Appl. Catal.* A, **2006**, 303, 72.

[27] S. R. M. Pereira, F. Clerc, D. Farrusseng, J. C. van der Waal, T. Maschmeyer and C. Mirodatos, *QSAR Comb. Sci.*, **2005**, 24, 45.

[28] H. Kordabadi and A. Jahanmiri, *Chem. Eng. J.*, **2005**, 108, 249.

[29] U. Rodermerk, D. Wolf, O.V. Buyeskaya, P. Claus, S. Senkan and M. Baerns, *Chem. Eng. J.*, **2001**, 82, 3.

[30] G. Grubert, S. Kolf, M. Baerns, I. Vauthey, D. Farrusseng, A.C. Veen, C. Mirodatos, E. R. Stobbe and P. D. Cobden, *Appl. Catal.* A: Gen., **2006**, 306, 17.

[31] J. S. Paul, R. Janssens, J. F. M. Denayer, G. V. Baron and P. J. Jacobs, *Comb. Chem*., **2005**, 7, 407.

[32] D. Wolf and M. Baerns, *Experimental Design for Combinatorial and High Throughput Development*, John Wiley & Sons, Hoboken, New Jersey, **2003**, p. 147.

[33] J.M. Serra, A. Corma, D. Farrusseng, L. Baumes, C. Mirodatos, C. Flego and C. Perego, *Catal. Today*, **2003**, 81, 425.

[34] Y. Yamada, A. Ueda, K. Nakagawa and T. Kobayashi, *Res. Chem. Intermed*., **2002**, 28, 5, 397-407.

[35] S. McLeod and L. F. Gladden, *J. Chem. Inf. Comput. Sci*., **2000**, 40, 981

[36] R. W. Maier and W. B. Whiting, *Comp. Chem. Eng.*, **1998**, 23, 47

[37] R. Yao, B. Yang, G. Cheng, X. Tao and Meng, F. *Chem. Eng. J.*, **2003**, 94, 113

[38] B. Lin and D. C. Miller, *Comp. and Chem. Eng*., **2004**, 28, 2287.

[39] L. Baumes, *J. Comb. Chem*, **2006**, 8, 304.

[40] J. Scheidtmann, D. Klaer, J.W. Saalfrank, T. Schmidt and W. F. Maier, *QSAR & Comb. Sci*., **2005**, 24, 203.

[41] L. Végvári, A. Tompos, S. Gobolos and J. Margitfalvi, J. *Catal. Today*, **2003**, 81, 517.

[42] C. B. Lucasius and G. Katemen, *Chem. Intell. Lab. Systems*, **1993**, 19, 1.

[43] C. B. Lucasius and G. Kateman, *Chemom. Intell. Lab. Systems.*, **1994**, 25, 99.

[44] D. B. Hibbert, *Chemom. Intell. Lab. Systems,* **1993**, 19, 277.

[45] S. Kirkpatrick, C. D. Gelatt and M. P. Vecchi, *Science*, **1983**, *220*, 671.

[46] V. Cerny, *J. Opt. Theory Appl*., **1985**, 45, 51.

[47] H. Sanvicente-Sánchez and J. Frausto-Solís, MPSA: *A methodology to parallelize simulated annealing and its application to the traveling salesman problem*, http://citeseer.ist.psu.edu.

[48] F. Glover and M. Laguna, *Tabu Search*, Kluwer Academic Publishers, Boston, **1997**.

[49] F. Glover and M. Laguna, *Modern Heuristic Techniques for Combinatorial Problems*, Blackwell Scientific Publishing, Oxford, England, **1993**.

[50] L. Baumes, PhD dissertation, Lyon, **2005**.

[51] F. Clerc, R. Rakotomalala and D. Farrusseng, *Learning fitness function in a combinatorial optimization process, in proceedings of ASMDA conference: Applied Stochastic Models and Data Analysis*, Brest, May **2005**.

[52] A. Corma, J . M. Serra, P. Serna, S. Valero, E. Argente and V. Botti, *J. Catal.,* **2005**, 229, 513.

[53] K. Huang, X. L. Zhan, F. Q. Chen and D. W. Lü, *Chem. Eng. Sci*., **2003**, 58, 81.

[54] T. Umegaki, Y. Watanabe, N. Nukui, K. Omata and M. Yamada, *Energy. Fuels*, **2003**, 17, 850.

[55] J. L. Bentley, *Comm. ACM*, **1975**, 18, 509.

[56] J. L Bentley and J. H. Friedman, *ACM Comput. Surv*. **1979**, 11, 397.

[57] E. Parzen, *Ann. Math. Stat.,* **1962**, *33*, 1065.

[58] D. E.Goldberg and J. Richardson, 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] P.-N. Tan, M. Steinbach and V. Kumar, *Introduction to Data Mining*, Pearson/Addison Wesley, Boston, **2005**.

[60] F. Clerc, *OptiCat - A Combinatorial Optimisation Software*. Free of charge download from http://eric.univ-lyon2.fr/~fclerc/

[61] H. Du and J. Lindsey, *Chemom. Intell. Lab. Systems*, **2002**, 62, 159

[62] M. E. Johnson and C. J. Nachtsheim, *Technometrics*, **1983**, 25, 271.

[63] P. Haaland, *Statistical Design and Analysis of Industrial Experiments,* Chapter 3, Marcel Dekker, New York, **1990**.

Appendix A - High-Throughput Experimentation Equipment

In the catalysis field a great deal of effort has been applied on the development and improvement of equipment that accelerates the speed of practical chemical experimentation. Examples are robotic liquid and solid handling equipment that pick and mix the reactants in different miniature vessels or reactors (e.g. Appendix A.1 and A.2), and platforms for catalyst synthesis preparations (Appendix A.2). Specialized reactor platforms are also available that can perform high-pressure parallel screening of catalyst reactions in the liquid phase (e.g. Appendix A.3) and gas-flow catalyst testing reactions (e.g. Appendix A.4).

The High throughput equipment employed in this thesis is presented and briefly described in this appendix.

Appendix A.1 - Liquid handling and parallel reactor 'Hamilton' platform

The Hamilton Dual Arm is a robotic liquid-handling platform. This robot dispenses and transfers liquid-phase chemicals using two types of devices: the single needle arm and the pipette D-arm. Set amounts of liquids can be transferred from one rack of vials to another via both arms, or dispensed to a rack of vials from an array of bottles containing different liquids stored in an adjacent cabinet via the needle arm. The needle arm is able to perforate septa and is adequate for experiments that require a closed environment. On the other hand, the pipette D-arm can dispense more viscous liquids and in more precise amounts than the needle arm. Between different liquid transfers the needle arm requires a washing cycle, while the D-arm picks up a new disposable plastic tip for each transfer, excluding contamination risks.

A special kind of rack has been designed for performing the terpene reactions presented in Chapter 3. This reactor rack was composed of 24 glass reactor HPLC vessels each with a working volume of 1.5 ml.

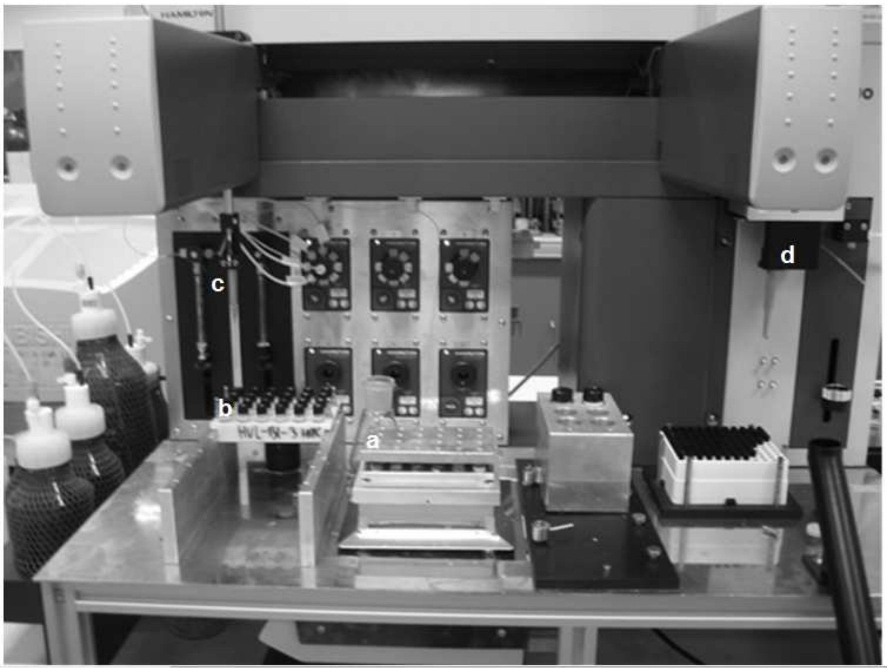


Figure A.1 - Hamilton Dual Arm liquid-handling and reactor rack platform. a) 6x4 glass reactor-rack, b) sample- rack, c) needle arm, and, d) pipette D-arm.

Appendix A.2 - Liquid handling and catalyst synthesis Zinsser-Sophas Analytics platform

The Zinsser- Sophas workbench employed in the research of Chapter 5 is a liquid handling platform specialised for catalyst synthesis procedures. This system has one arm which can be equipped with up to four independent pipetting or dispensing probes. Liquid transfer from vials or liquid dispensing from stored bottles can be performed. Several reactor and chemicals handling racks can be assembled in the system. For the catalyst syntheses in Chapter 5 a rack of 6x4 reactor vials has been used. The reactor racks can be positioned in a vortex mixer with a built-in heater (to +150 °C). All the operations performed by the workbench are controlled by a computer supplied with Zinsser Sophas software.

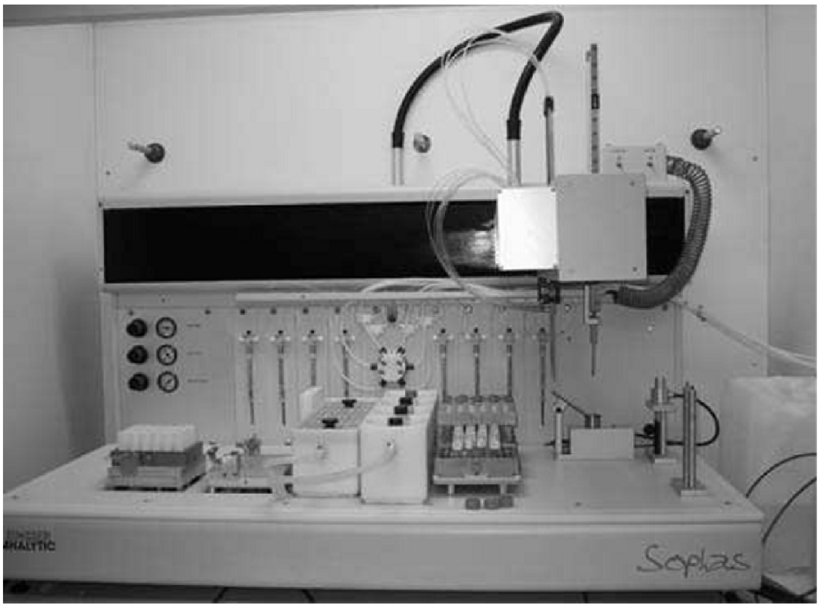
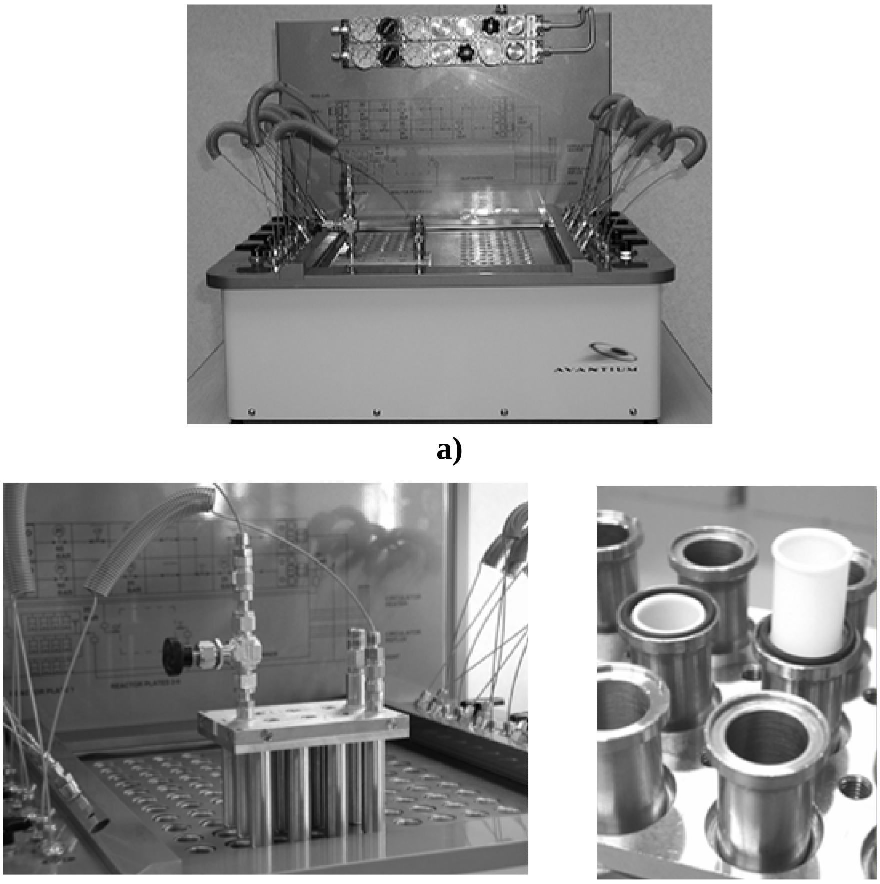


Figure A.2 - Sophas liquid handling and catalyst synthesis platform.

Appendix A.3 - High pressure ‘Quick Catalyst Screening’ parallel reactors platform

The Quick Catalyst Screening (QCS) platform (Figure A.3a), as employed for the catalytic hydrogenations of Chapter 4, is composed of up to 8 blocks of 12 (3x4) high-pressure stainless steel reactors (Figure A.3b). Pressure can be independently set for each reactor block up to 20 bar. The reactors have disposable Teflon inserts (Figure A.3c), a working volume of 1-2 ml and are mixed magnetically. The temperature can be varied from -40 to 180 °C.



a)

b) **c)**

Figure A.3 - Quick Catalyst Screening (QCS) Avantium BV platform. a) General view of the equipment, b) Detail of one high pressure 12-reactor block, and c) Detail of the reactors.

Appendix A.4 - Gas-phase AMTEC reactor system

The 16-fold parallel reactor system from Amtec has been employed for the testing of the synthesised catalyst library in Chapter 5. The system, which is equipped with 16 individual, continuous-flow, fixed-bed, stainless steel reactors, with a catalyst volume bed of 4 ml, ensures uniform temperatures, pressures and fluid flows in all reactors. The system allows a maximum flow rate of 200 ml/min per reactor, a temperature of 600 ºC and 13 bar of pressure. Micro-GC and Micro-ms analytical equipment were adapted to the system for semi-online analysis.

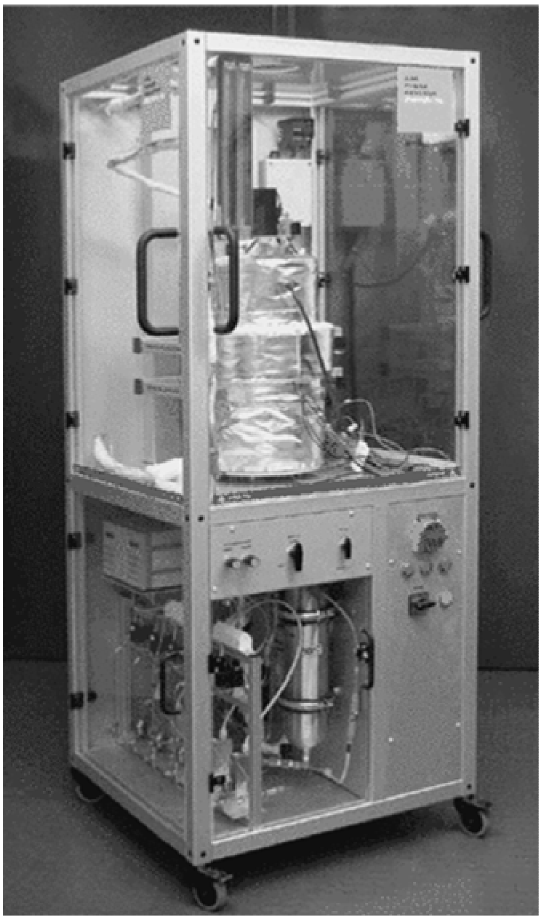
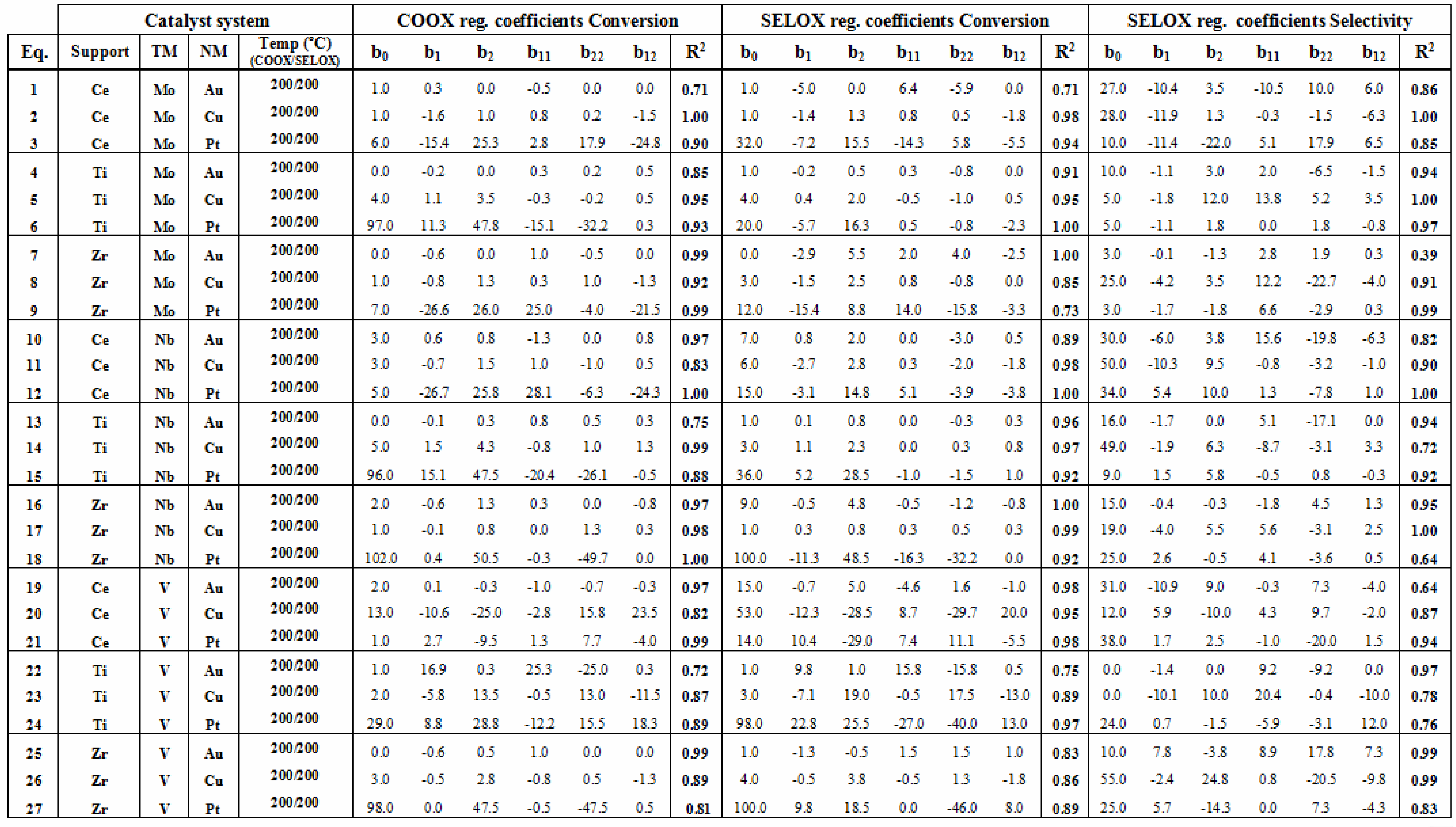


Figure A.4 - Switch-16 - Amtec gas-phase reactor system.

Appendix B - COOX and SELOX Benchmark Regression Functions

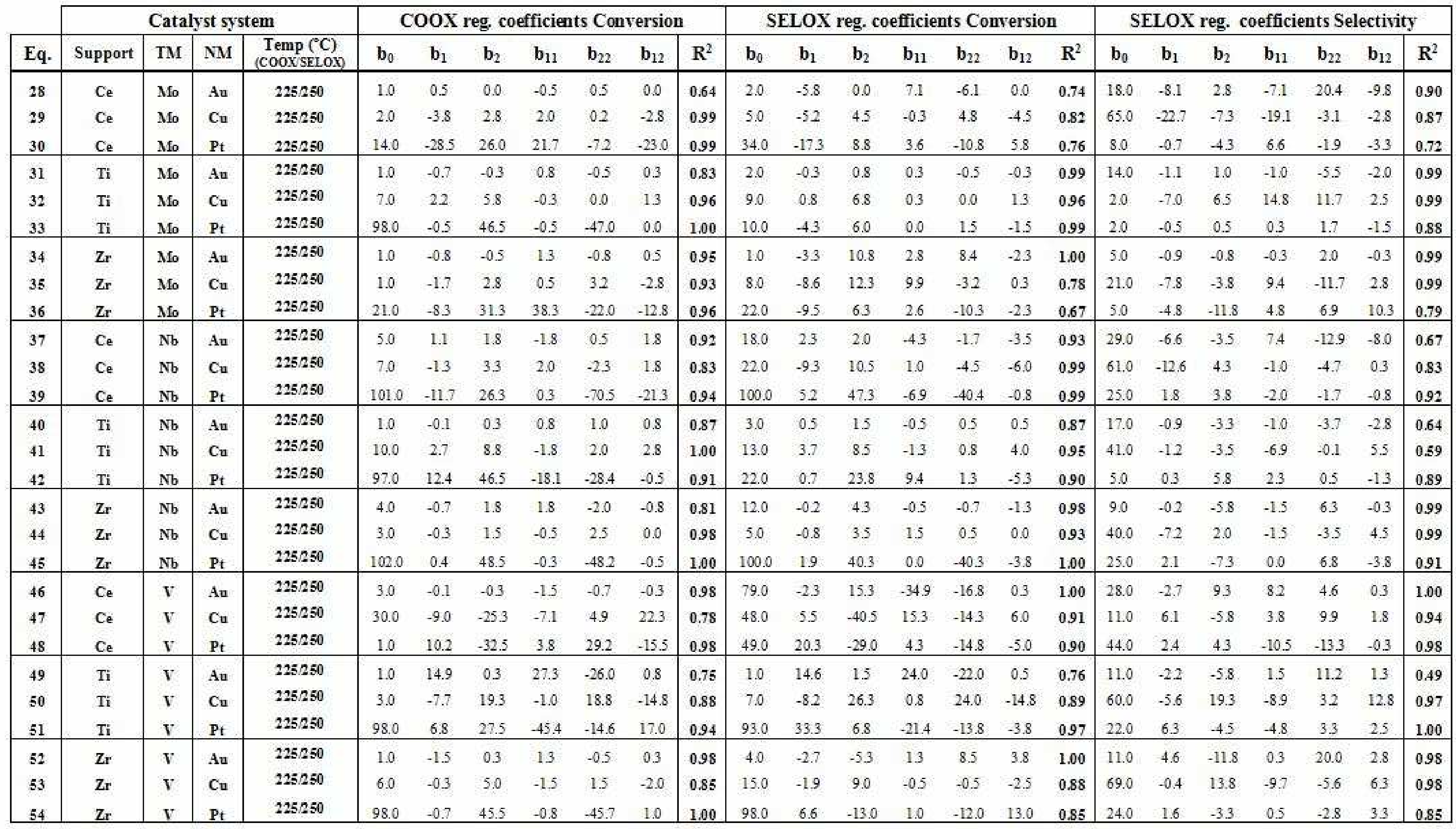
The regression coefficients (b_0, b_1, b_2, b_{12}, b_{11}, b_{22}) of the experimental data from the COOX and SELOX reactions (Chapter 5) obtained by the Nemrowd 2000 software are presented in table TB.1. A polynomial quadratic equation ($y = b_0 + b_1x_1 + b_2x^2 + b_{12}x_1x_2 + b_{11}x_1^2 + b_{22}x_2^2$) based on these coefficients could be derived for each metal ternary combination, at a certain reaction temperature, for the responses of CO conversion and CO_2 selectivity. The equations of table TB.1 constitute the analytical forms of the COOX and SELOX benchmarks before the application of the desirability criteria. A visual representation of the corresponding response surfaces can be observed in Chapter 5 (Figures 5.4 and 5.5).

Table B.1 – COOX and SELOX benchmark functions. Polynomial regression coefficients obtained for the conversion and selectivity responses: ($y = b_0 + b_1x_1 + b_2x^2 + b_{12}x_1x_2 + b_{11}x_1^2 + b_{22}x_2^2$) where x_1 is the TM amount and x_2 the NM amount in %.



	Catalyst system				COOX reg. coefficients Conversion							SELOX reg. coefficients Conversion							SELOX reg. coefficients Selectivity						
Eq.	**Support**	**TM**	**NM**	**Temp (°C) (COOX/SELOX)**	b_0	b_1	b_2	b_{11}	b_{22}	b_{12}	R^2	b_0	b_1	b_2	b_{11}	b_{22}	b_{12}	R^2	b_0	b_1	b_2	b_{11}	b_{22}	b_{12}	R^2
1	**Ce**	**Mo**	**Au**	**200/200**	1.0	0.3	0.0	-0.5	0.0	0.0	**0.71**	1.0	-5.0	0.0	6.4	-5.9	0.0	**0.71**	27.0	-10.4	3.5	-10.5	10.0	6.0	**0.86**
2	**Ce**	**Mo**	**Cu**	**200/200**	1.0	-1.6	1.0	0.8	0.2	-1.5	**1.00**	1.0	-1.4	1.3	0.8	0.5	-1.8	**0.98**	28.0	-11.9	1.3	-0.3	-1.5	-6.3	**1.00**
3	**Ce**	**Mo**	**Pt**	**200/200**	6.0	-15.4	25.3	2.8	17.9	-24.8	**0.90**	32.0	-7.2	15.5	-14.3	5.8	-5.5	**0.94**	10.0	-11.4	-22.0	5.1	17.9	6.5	**0.85**
4	**Ti**	**Mo**	**Au**	**200/200**	0.0	-0.2	0.0	0.3	0.2	0.5	**0.85**	1.0	-0.2	0.5	0.3	-0.8	0.0	**0.91**	10.0	-1.1	3.0	2.0	-6.5	-1.5	**0.94**
5	**Ti**	**Mo**	**Cu**	**200/200**	4.0	1.1	3.5	-0.3	-0.2	0.5	**0.95**	4.0	0.4	2.0	-0.5	-1.0	0.5	**0.95**	5.0	-1.8	12.0	13.8	5.2	3.5	**1.00**
6	**Ti**	**Mo**	**Pt**	**200/200**	97.0	11.3	47.8	-15.1	-32.2	0.3	**0.93**	20.0	-5.7	16.3	0.5	-0.8	-2.3	**1.00**	5.0	-1.1	1.8	0.0	1.8	-0.8	**0.97**
7	**Zr**	**Mo**	**Au**	**200/200**	0.0	-0.6	0.0	1.0	-0.5	0.0	**0.99**	0.0	-2.9	5.5	2.0	4.0	-2.5	**1.00**	3.0	-0.1	-1.3	2.8	1.9	0.3	**0.39**
8	**Zr**	**Mo**	**Cu**	**200/200**	1.0	-0.8	1.3	0.3	1.0	-1.3	**0.92**	3.0	-1.5	2.5	0.8	-0.8	0.0	**0.85**	25.0	-4.2	3.5	12.2	-22.7	-4.0	**0.91**
9	**Zr**	**Mo**	**Pt**	**200/200**	7.0	-26.6	26.0	25.0	-4.0	-21.5	**0.99**	12.0	-15.4	8.8	14.0	-15.8	-3.3	**0.73**	3.0	-1.7	-1.8	6.6	-2.9	0.3	**0.99**
10	**Ce**	**Nb**	**Au**	**200/200**	3.0	0.6	0.8	-1.3	0.0	0.8	**0.97**	7.0	0.8	2.0	0.0	-3.0	0.5	**0.89**	30.0	-6.0	3.8	15.6	-19.8	-6.3	**0.82**
11	**Ce**	**Nb**	**Cu**	**200/200**	3.0	-0.7	1.5	1.0	-1.0	0.5	**0.83**	6.0	-2.7	2.8	0.3	-2.0	-1.8	**0.98**	50.0	-10.3	9.5	-0.8	-3.2	-1.0	**0.90**
12	**Ce**	**Nb**	**Pt**	**200/200**	5.0	-26.7	25.8	28.1	-6.3	-24.3	**1.00**	15.0	-3.1	14.8	5.1	-3.9	-3.8	**1.00**	34.0	5.4	10.0	1.3	-7.8	1.0	**1.00**
13	**Ti**	**Nb**	**Au**	**200/200**	0.0	-0.1	0.3	0.8	0.5	0.3	**0.75**	1.0	0.1	0.8	0.0	-0.3	0.3	**0.96**	16.0	-1.7	0.0	5.1	-17.1	0.0	**0.94**
14	**Ti**	**Nb**	**Cu**	**200/200**	5.0	1.5	4.3	-0.8	1.0	1.3	**0.99**	3.0	1.1	2.3	0.0	0.3	0.8	**0.97**	49.0	-1.9	6.3	-8.7	-3.1	3.3	**0.72**
15	**Ti**	**Nb**	**Pt**	**200/200**	96.0	15.1	47.5	-20.4	-26.1	-0.5	**0.88**	36.0	5.2	28.5	-1.0	-1.5	1.0	**0.92**	9.0	1.5	5.8	-0.5	0.8	-0.3	**0.92**
16	**Zr**	**Nb**	**Au**	**200/200**	2.0	-0.6	1.3	0.3	0.0	-0.8	**0.97**	9.0	-0.5	4.8	-0.5	-1.2	-0.8	**1.00**	15.0	-0.4	-0.3	-1.8	4.5	1.3	**0.95**
17	**Zr**	**Nb**	**Cu**	**200/200**	1.0	-0.1	0.8	0.0	1.3	0.3	**0.98**	1.0	0.3	0.8	0.3	0.5	0.3	**0.99**	19.0	-4.0	5.5	5.6	-3.1	2.5	**1.00**
18	**Zr**	**Nb**	**Pt**	**200/200**	102.0	0.4	50.5	-0.3	-49.7	0.0	**1.00**	100.0	-11.3	48.5	-16.3	-32.2	0.0	**0.92**	25.0	2.6	-0.5	4.1	-3.6	0.5	**0.64**
19	**Ce**	**V**	**Au**	**200/200**	2.0	0.1	-0.3	-1.0	-0.7	-0.3	**0.97**	15.0	-0.7	5.0	-4.6	1.6	-1.0	**0.98**	31.0	-10.9	9.0	-0.3	7.3	-4.0	**0.64**
20	**Ce**	**V**	**Cu**	**200/200**	13.0	-10.6	-25.0	-2.8	15.8	23.5	**0.82**	53.0	-12.3	-28.5	8.7	-29.7	20.0	**0.95**	12.0	5.9	-10.0	4.3	9.7	-2.0	**0.87**
21	**Ce**	**V**	**Pt**	**200/200**	1.0	2.7	-9.5	1.3	7.7	-4.0	**0.99**	14.0	10.4	-29.0	7.4	11.1	-5.5	**0.98**	38.0	1.7	2.5	-1.0	-20.0	1.5	**0.94**
22	**Ti**	**V**	**Au**	**200/200**	1.0	16.9	0.3	25.3	-25.0	0.3	**0.72**	1.0	9.8	1.0	15.8	-15.8	0.5	**0.75**	0.0	-1.4	0.0	9.2	-9.2	0.0	**0.97**
23	**Ti**	**V**	**Cu**	**200/200**	2.0	-5.8	13.5	-0.5	13.0	-11.5	**0.87**	3.0	-7.1	19.0	-0.5	17.5	-13.0	**0.89**	0.0	-10.1	10.0	20.4	-0.4	-10.0	**0.78**
24	**Ti**	**V**	**Pt**	**200/200**	29.0	8.8	28.8	-12.2	15.5	18.3	**0.89**	98.0	22.8	25.5	-27.0	-40.0	13.0	**0.97**	24.0	0.7	-1.5	-5.9	-3.1	12.0	**0.76**
25	**Zr**	**V**	**Au**	**200/200**	0.0	-0.6	0.5	1.0	0.0	0.0	**0.99**	1.0	-1.3	-0.5	1.5	1.5	1.0	**0.83**	10.0	7.8	-3.8	8.9	17.8	7.3	**0.99**
26	**Zr**	**V**	**Cu**	**200/200**	3.0	-0.5	2.8	-0.8	0.5	-1.3	**0.89**	4.0	-0.5	3.8	-0.5	1.3	-1.8	**0.86**	55.0	-2.4	24.8	0.8	-20.5	-9.8	**0.99**
27	**Zr**	**V**	**Pt**	**200/200**	98.0	0.0	47.5	-0.5	-47.5	0.5	**0.81**	100.0	9.8	18.5	0.0	-46.0	8.0	**0.89**	25.0	5.7	-14.3	0.0	7.3	-4.3	**0.83**

Table B.1 – COOX and SELOX benchmark functions. Polynomial regression coefficients obtained for the conversion and selectivity responses: ($y = b_0 + b_1x_1 + b_2x^2 + b_{12}x_1x_2 + b_{11}x_1^2 + b_{22}x_2^2$) where x_1 is the TM amount and x_2 the NM amount in %. (Continuation)



	Catalyst system				COOX reg. coefficients Conversion							SELOX reg. coefficients Conversion							SELOX reg. coefficients Selectivity						
Eq.	**Support**	**TM**	**NM**	**Temp (°C) (COOX/SELOX)**	b_0	b_1	b_2	b_{11}	b_{22}	b_{12}	R^2	b_0	b_1	b_2	b_{11}	b_{22}	b_{12}	R^2	b_0	b_1	b_2	b_{11}	b_{22}	b_{12}	R^2
28	Ce	Mo	Au	225/250	1.0	0.5	0.0	-0.5	0.5	0.0	**0.64**	2.0	-5.8	0.0	7.1	-6.1	0.0	**0.74**	18.0	-8.1	2.8	-7.1	20.4	-9.8	**0.90**
29	Ce	Mo	Cu	225/250	2.0	-3.8	2.8	2.0	0.2	-2.8	**0.99**	5.0	-5.2	4.5	-0.3	4.8	-4.5	**0.82**	65.0	-22.7	-7.3	-19.1	-3.1	-2.8	**0.87**
30	Ce	Mo	Pt	225/250	14.0	-28.5	26.0	21.7	-7.2	-23.0	**0.99**	34.0	-17.3	8.8	3.6	-10.8	5.8	**0.76**	8.0	-0.7	-4.3	6.6	-1.9	-3.3	**0.72**
31	Ti	Mo	Au	225/250	1.0	-0.7	-0.3	0.8	-0.5	0.3	**0.83**	2.0	-0.3	0.8	0.3	-0.5	-0.3	**0.99**	14.0	-1.1	1.0	-1.0	-5.5	-2.0	**0.99**
32	Ti	Mo	Cu	225/250	7.0	2.2	5.8	-0.3	0.0	1.3	**0.96**	9.0	0.8	6.8	0.3	0.0	1.3	**0.96**	2.0	-7.0	6.5	14.8	11.7	2.5	**0.99**
33	Ti	Mo	Pt	225/250	98.0	-0.5	46.5	-0.5	-47.0	0.0	**1.00**	10.0	-4.3	6.0	0.0	1.5	-1.5	**0.99**	2.0	-0.5	0.5	0.3	1.7	-1.5	**0.88**
34	Zr	Mo	Au	225/250	1.0	-0.8	-0.5	1.3	-0.8	0.5	**0.95**	1.0	-3.3	10.8	2.8	8.4	-2.3	**1.00**	5.0	-0.9	-0.8	-0.3	2.0	-0.3	**0.99**
35	Zr	Mo	Cu	225/250	1.0	-1.7	2.8	0.5	3.2	-2.8	**0.93**	8.0	-8.6	12.3	9.9	-3.2	0.3	**0.78**	21.0	-7.8	-3.8	9.4	-11.7	2.8	**0.99**
36	Zr	Mo	Pt	225/250	21.0	-8.3	31.3	38.3	-22.0	-12.8	**0.96**	22.0	-9.5	6.3	2.6	-10.3	-2.3	**0.67**	5.0	-4.8	-11.8	4.8	6.9	10.3	**0.79**
37	Ce	Nb	Au	225/250	5.0	1.1	1.8	-1.8	0.5	1.8	**0.92**	18.0	2.3	2.0	-4.3	-1.7	-3.5	**0.93**	29.0	-6.6	-3.5	7.4	-12.9	-8.0	**0.67**
38	Ce	Nb	Cu	225/250	7.0	-1.3	3.3	2.0	-2.3	1.8	**0.83**	22.0	-9.3	10.5	1.0	-4.5	-6.0	**0.99**	61.0	-12.6	4.3	-1.0	-4.7	0.3	**0.83**
39	Ce	Nb	Pt	225/250	101.0	-11.7	26.3	0.3	-70.5	-21.3	**0.94**	100.0	5.2	47.3	-6.9	-40.4	-0.8	**0.99**	25.0	1.8	3.8	-2.0	-1.7	-0.8	**0.92**
40	Ti	Nb	Au	225/250	1.0	-0.1	0.3	0.8	1.0	0.8	**0.87**	3.0	0.5	1.5	-0.5	0.5	0.5	**0.87**	17.0	-0.9	-3.3	-1.0	-3.7	-2.8	**0.64**
41	Ti	Nb	Cu	225/250	10.0	2.7	8.8	-1.8	2.0	2.8	**1.00**	13.0	3.7	8.5	-1.3	0.8	4.0	**0.95**	41.0	-1.2	-3.5	-6.9	-0.1	5.5	**0.59**
42	Ti	Nb	Pt	225/250	97.0	12.4	46.5	-18.1	-28.4	-0.5	**0.91**	22.0	0.7	23.8	9.4	1.3	-5.3	**0.90**	5.0	0.3	5.8	2.3	0.5	-1.3	**0.89**
43	Zr	Nb	Au	225/250	4.0	-0.7	1.8	1.8	-2.0	-0.8	**0.81**	12.0	-0.2	4.3	-0.5	-0.7	-1.3	**0.98**	9.0	-0.2	-5.8	-1.5	6.3	-0.3	**0.99**
44	Zr	Nb	Cu	225/250	3.0	-0.3	1.5	-0.5	2.5	0.0	**0.98**	5.0	-0.8	3.5	1.5	0.5	0.0	**0.93**	40.0	-7.2	2.0	-1.5	-3.5	4.5	**0.99**
45	Zr	Nb	Pt	225/250	102.0	0.4	48.5	-0.3	-48.2	-0.5	**1.00**	100.0	1.9	40.3	0.0	-40.3	-3.8	**1.00**	25.0	2.1	-7.3	0.0	6.8	-3.8	**0.91**
46	Ce	V	Au	225/250	3.0	-0.1	-0.3	-1.5	-0.7	-0.3	**0.98**	79.0	-2.3	15.3	-34.9	-16.8	0.3	**1.00**	28.0	-2.7	9.3	8.2	4.6	0.3	**1.00**
47	Ce	V	Cu	225/250	30.0	-9.0	-25.3	-7.1	4.9	22.3	**0.78**	48.0	5.5	-40.5	15.3	-14.3	6.0	**0.91**	11.0	6.1	-5.8	3.8	9.9	1.8	**0.94**
48	Ce	V	Pt	225/250	1.0	10.2	-32.5	3.8	29.2	-15.5	**0.98**	49.0	20.3	-29.0	4.3	-14.8	-5.0	**0.90**	44.0	2.4	4.3	-10.5	-13.3	-0.3	**0.98**
49	Ti	V	Au	225/250	1.0	14.9	0.3	27.3	-26.0	0.8	**0.75**	1.0	14.6	1.5	24.0	-22.0	0.5	**0.76**	11.0	-2.2	-5.8	1.5	11.2	1.3	**0.49**
50	Ti	V	Cu	225/250	3.0	-7.7	19.3	-1.0	18.8	-14.8	**0.88**	7.0	-8.2	26.3	0.8	24.0	-14.8	**0.89**	60.0	-5.6	19.3	-8.9	3.2	12.8	**0.97**
51	Ti	V	Pt	225/250	98.0	6.8	27.5	-45.4	-14.6	17.0	**0.94**	93.0	33.3	6.8	-21.4	-13.8	-3.8	**0.97**	22.0	6.3	-4.5	-4.8	3.3	2.5	**1.00**
52	Zr	V	Au	225/250	1.0	-1.5	0.3	1.3	-0.5	0.3	**0.98**	4.0	-2.7	-5.3	1.3	8.5	3.8	**1.00**	11.0	4.6	-11.8	0.3	20.0	2.8	**0.98**
53	Zr	V	Cu	225/250	6.0	-0.3	5.0	-1.5	1.5	-2.0	**0.85**	15.0	-1.9	9.0	-0.5	-0.5	-2.5	**0.88**	69.0	-0.4	13.8	-9.7	-5.6	6.3	**0.98**
54	Zr	V	Pt	225/250	98.0	-0.7	45.5	-0.8	-45.7	1.0	**1.00**	98.0	6.6	-13.0	1.0	-12.0	13.0	**0.85**	24.0	1.6	-3.3	0.5	-2.8	3.3	**0.85**

Table B.1 – COOX and SELOX benchmark functions. Polynomial regression coefficients obtained for the conversion and selectivity responses: ($y = b_0 + b_1x_1 + b_2x^2 + b_{12}x_1x_2 + b_{11}x_1^2 + b_{22}x_2^2$) where x_1 is the TM amount and x_2 the NM amount in %. (Continuation)

	Catalyst system				COOX reg. coefficients Conversion							SELOX reg. coefficients Conversion							SELOX reg. coefficients Selectivity						
Eq.	Support	TM	NM	Temp (°C) (COOX/SELOX)	b_0	b_1	b_2	b_{11}	b_{22}	b_{12}	R^2	b_0	b_1	b_2	b_{11}	b_{22}	b_{12}	R^2	b_0	b_1	b_2	b_{11}	b_{22}	b_{12}	R^2
55	Ce	Mo	Au	250/300	1.0	0.9	-1.0	-0.3	1.3	-0.5	0.85	4.0	-6.3	-1.3	7.9	-6.7	1.3	0.83	16.0	2.5	0.3	4.1	-1.8	-4.3	0.63
56	Ce	Mo	Cu	250/300	3.0	-7.8	4.8	5.4	-0.6	-4.8	0.99	18.0	-10.5	4.3	-5.9	9.1	-4.3	0.73	49.0	-7.2	-2.5	-14.5	7.5	1.0	0.98
57	Ce	Mo	Pt	250/300	102.0	-21.9	29.3	-16.1	-51.2	-16.8	1.00	37.0	-15.5	5.3	1.8	-9.0	3.8	0.98	9.0	-2.2	-1.5	2.8	-2.3	-1.5	0.56
58	Ti	Mo	Au	250/300	2.0	-8.3	0.3	11.2	-10.0	0.8	0.69	3.0	-0.4	1.5	0.5	0.5	-0.5	0.92	12.0	-0.2	-1.3	-1.5	-1.7	-2.8	0.96
59	Ti	Mo	Cu	250/300	12.0	3.4	10.0	-0.5	0.5	2.5	0.99	11.0	-0.2	16.3	7.1	7.1	3.8	0.95	3.0	-7.8	-2.5	12.5	11.5	3.5	1.00
60	Ti	Mo	Pt	250/300	97.0	-1.1	43.8	0.0	-42.8	2.3	1.00	6.0	-2.3	1.0	0.8	4.7	-3.0	0.95	2.0	-0.5	0.3	0.0	0.8	-0.8	0.93
61	Zr	Mo	Au	250/300	2.0	-1.3	-0.5	2.6	0.4	2.0	0.90	2.0	-5.5	13.5	2.3	12.7	-6.5	0.99	10.0	0.7	-1.8	-2.8	1.6	-3.3	0.99
62	Zr	Mo	Cu	250/300	3.0	-3.5	6.8	0.0	8.3	-5.8	0.90	12.0	-16.9	20.8	17.3	-0.6	-7.8	0.96	13.0	-2.6	-8.5	8.9	-0.4	2.5	0.75
63	Zr	Mo	Pt	250/300	98.0	-10.9	23.0	-0.5	-22.0	23.0	0.87	26.0	-12.1	-6.5	1.5	-3.5	8.5	0.98	6.0	-4.3	-9.8	3.6	4.7	8.8	0.82
64	Ce	Nb	Au	250/300	14.0	2.2	2.3	-5.6	-2.1	2.8	0.90	65.0	9.7	-12.0	-27.0	3.5	-26.0	0.91	24.0	-1.8	-4.0	-3.8	-3.2	-5.0	0.79
65	Ce	Nb	Cu	250/300	16.0	4.2	21.3	3.3	8.9	17.3	0.80	48.0	-12.3	12.8	-2.0	-2.7	-11.8	0.95	48.0	-4.5	-3.3	-1.3	-2.5	-4.3	0.52
66	Ce	Nb	Pt	250/300	101.0	-1.5	45.0	0.5	-45.5	3.0	1.00	98.0	3.0	39.3	-2.0	-40.7	0.8	1.00	23.0	0.4	3.8	-0.3	-4.0	0.8	0.96
67	Ti	Nb	Au	250/300	3.0	0.0	0.3	1.5	1.7	0.8	0.73	7.0	1.4	2.8	-1.5	2.3	-0.3	0.81	15.0	-1.0	-4.8	-2.3	-1.0	-3.3	0.80
68	Ti	Nb	Cu	250/300	18.0	4.6	14.5	-2.6	2.1	4.5	0.99	31.0	6.3	22.0	-3.3	6.3	7.5	0.87	28.0	-2.5	-5.8	-2.6	-0.2	3.3	0.52
69	Ti	Nb	Pt	250/300	97.0	-5.7	34.0	-0.8	-33.7	10.0	0.98	29.0	-1.5	32.3	5.4	6.9	-6.3	0.92	7.0	-0.5	7.5	1.3	1.7	-1.5	0.93
70	Zr	Nb	Au	250/300	12.0	-1.0	3.0	0.3	-4.3	-1.0	0.71	18.0	1.0	0.5	-2.3	1.3	-3.5	0.70	6.0	0.4	-5.5	-1.0	5.5	-1.5	0.95
71	Zr	Nb	Cu	250/300	6.0	-0.9	3.8	-0.5	6.3	-0.8	0.96	16.0	-3.8	6.3	4.3	1.4	-1.3	0.91	34.0	-5.1	-4.3	-0.5	-1.7	1.8	1.00
72	Zr	Nb	Pt	250/300	102.0	-3.5	34.8	-0.3	-34.5	7.3	0.98	93.0	9.2	-3.5	-4.1	-11.9	-12.5	0.96	22.0	-0.3	-5.5	-1.0	2.0	1.5	0.86
73	Ce	V	Au	250/300	4.0	0.2	-1.0	-1.8	-0.2	0.0	0.98	99.0	-5.0	37.0	-41.6	4.6	3.0	1.00	24.0	-3.1	-4.3	1.3	5.0	3.3	0.98
74	Ce	V	Cu	250/300	98.0	12.5	-44.8	-19.1	-26.1	-0.3	0.88	49.0	0.8	-40.3	24.0	-13.2	2.3	1.00	11.0	2.9	-1.0	5.9	6.1	6.0	0.81
75	Ce	V	Pt	250/300	11.0	4.9	-45.5	2.8	38.2	0.0	0.98	100.0	22.4	-23.0	-20.7	-31.3	0.0	0.96	25.0	4.1	4.5	-3.3	-0.2	-0.5	0.99
76	Ti	V	Au	250/300	2.0	6.8	-0.3	39.0	-37.3	1.8	0.96	3.0	15.2	0.8	28.1	-26.8	0.3	0.78	26.0	3.7	-4.8	-6.1	-1.6	1.3	0.76
77	Ti	V	Cu	250/300	4.0	-9.8	25.5	0.3	24.7	-18.5	0.89	12.0	-7.6	25.3	4.3	26.9	-11.3	0.84	47.0	-1.2	-1.8	-3.6	-3.2	10.3	0.59
78	Ti	V	Pt	250/300	98.0	3.3	25.0	-42.1	-2.4	13.0	0.98	96.0	29.6	9.3	-21.2	-12.6	0.3	1.00	23.0	6.1	0.0	-5.1	-0.9	2.0	0.98
79	Zr	V	Au	250/300	2.0	-2.2	0.0	2.3	-0.3	1.0	0.99	9.0	-4.3	-12.3	1.0	15.2	10.8	0.99	6.0	3.8	-7.0	0.8	11.7	2.5	0.98
80	Zr	V	Cu	250/300	10.0	-0.9	9.3	-1.8	3.5	-3.8	0.83	38.0	-9.7	25.3	1.8	0.0	-8.8	0.86	60.0	-0.6	5.8	-9.9	-6.8	14.3	0.93
81	Zr	V	Pt	250/300	98.0	-4.2	36.3	-0.5	-36.7	8.3	0.98	84.0	9.2	-9.3	8.2	-17.4	15.3	0.81	20.0	2.1	-2.3	2.0	-3.8	3.8	0.82

Appendix C - **OptiCat Software**

The OptiCat (Optimization for Catalysis) software is a platform for building stochastic optimization algorithms. A diverse variety of algorithms can be implemented via drag and drop operations of the pre-established procedure building blocks. Some of these algorithms are, for instance: evolutionary strategy, genetic algorithm, Tabu search, simulated annealing and a series of hybrid algorithms with knowledge based implementations used in Chapter 7. This software has been developed due to the need of soft computing tools in combinatorial high-throughput catalysis and can be downloaded from the website: http://eric.univ-lyon2.fr/~fclerc/.

Most optimization algorithm are traditionally composed of four iterative phases: 1) initialisation, to generate a first set of potential solutions to the problem; 2) evaluation of each solution (results visualisation); 3) confrontation with a stopping criterion; 4) generation of a new set of potential solutions and return at stage 2. Figure C.1 gives a schematic representation of the algorithm implementation in Opticat.

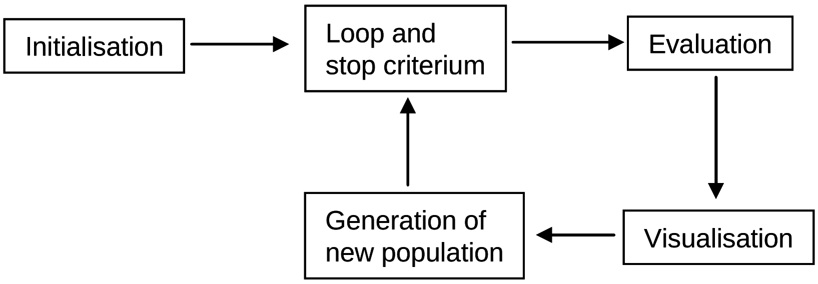


Figure C.1 - A classical optimization process

With judicious arrangement of the procedure building blocks existent in the Opticat software, diverse iterative optimization strategies like random search, simulated annealing, Tabu search, evolutionary algorithms and other algorithms can be implemented. Figure C.2 shows an example of an OptiCat screenshot of a Genetic algorithm construction.

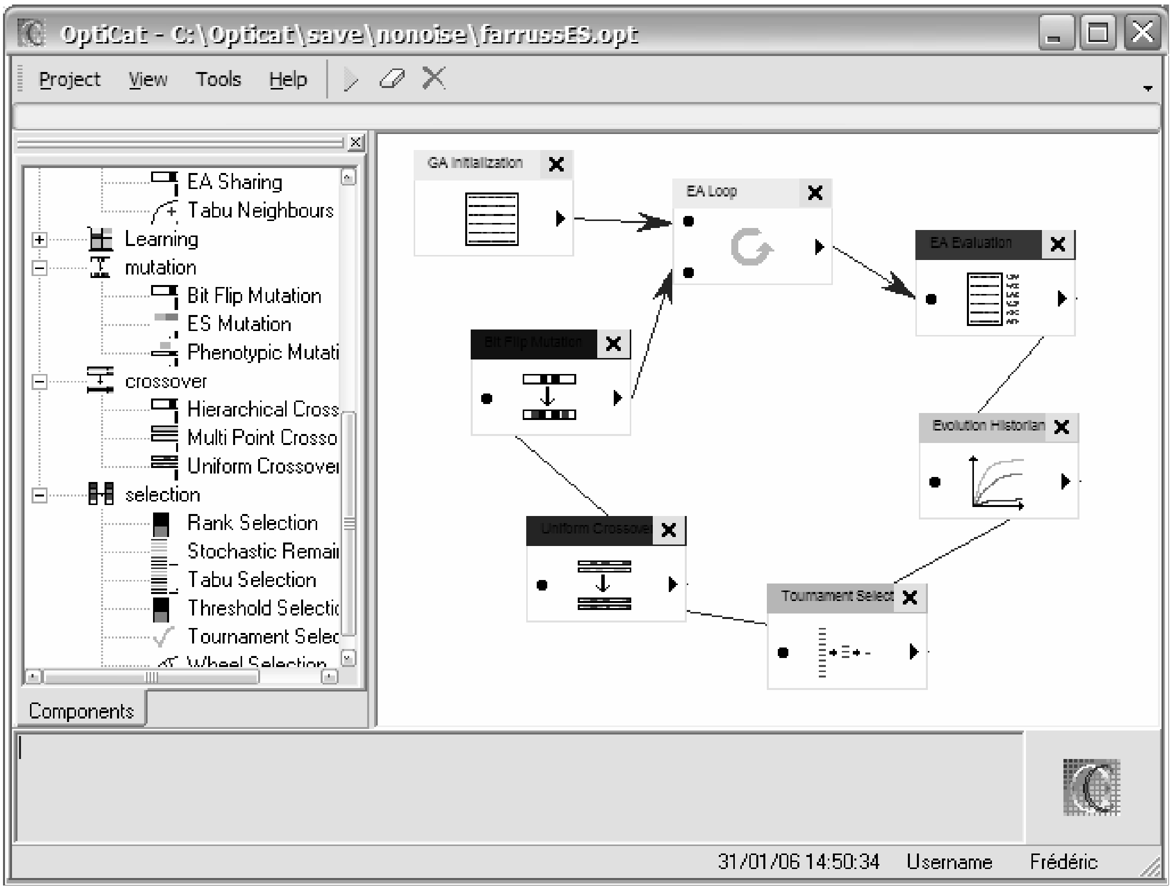


Figure C.2 - OptiCat screenshot of a Genetic algorithm.

Summary

The main topic of this thesis is the investigation of the synergies between High-Throughput Experimentation (HTE) and Chemometric Optimization methodologies in Catalysis research and of the use of such methodologies to maximize the advantages of using HTE methods.

In Chapter 1 an introduction to High-Throughput Experimentation (HTE) and Chemometric optimization methods and their combined use for catalysis research is presented. Considerations are given about the use of chemometric experimental planning methodologies such as Design of Experiments and Global Optimization algorithms.

An overview of the use of the Design of Experiments technique is presented in Chapter 2. This methodology is central to most of the thesis research, being applied through almost all its chapters.

Chapter 3 deals with the exploration of a terpene addition reaction parameter space, in this case a hydration reaction, using a multi-step Design of Experiments approach. From this investigation insights on the main reactivity trends were obtained. Namely that the highest selectivities were achieved with camphene substrate and the lowest with the two pinene substrates, due to the higher number of rearrangement pathways usually undergone with the latter. It is shown that the acid strength had a significant effect on the activity but not on the resulting product mixture; the solvent type and water concentration were shown to have smaller effects. DoE proved to be a fruitful strategy to explore and discover the underlying trends in such a challenging chemical parameter space.

In Chapter 4 is presented the optimization of the reaction conditions for the hydrogenation of cyanohydrin esters to N-acylated-ß-amino alcohols, using again a multi-step Design of Experiments approach. This strategy enabled, with only a small number of experiments, the location of the most promising region of the experimental space in which this reaction took place, namely the use of Ni and Rh catalysts, supported on either carbon or alumina, and with 2-propanol or dioxane as solvents. Via further optimization designs the best conditions for the reaction were obtained, yielding total conversion and yields of about 50% and 90 % for the benzylic and aliphatic cyanohydrin ester substrates, respectively. The application was demonstrated on a range of aliphatic and aromatic substrates. When used with enantiopure substrates no racemisation was observed, except to a small degree for the benzylic substrates. DoE together with HTE technologies proved efficient for the exploration and optimization of this chemical route towards the production of N-acylated-ß-amino alcohols.

In Chapter 5 the planning and modelling of catalytic benchmark response surfaces using Design of Experiments is performed. A library of 189 mixed-oxide catalysts was prepared and their activity tested and modelled for the CO oxidation reaction in both the absence (COOX) and presence (SELOX) of hydrogen. A DoE model (quadratic order) was used to map the parameter space. A visual representation of the resulting response surfaces was obtained together with the model analytical functions which constituted the benchmarks for the two conditions tested (COOX and SELOX). The multiple responses of interest and the process preferences (conversion of CO, selectivity towards CO_2 and temperature) were taken into account through the use of a desirability function. The visualisation of the response surface allowed obtaining a good understanding of the effects of the studied variables and the location of the optimality regions. The COOX and SELOX benchmarks were used in the following chapters to access the optimization performance of selected Global optimization algorithms.

In Chapter 6 the optimization procedure of employing the Global Optimization Genetic algorithm and the influence of the algorithm settings on its optimization efficiency is studied. Both the COOX and SELOX benchmarks obtained in Chapter 5 were used to validate the Genetic algorithm performance. A Design of Experiments approach was used to select the configurations to be tested and study the effects of the chosen parameter settings. The most important parameter found was the population size, where the largest population compatible with the desired HTE workflow (the synthesis and testing of one generation of catalysts per working day) was found to be preferential. An optimal GA configuration was found for the optimization of the SELOX and COOX benchmarks, consisting of the use of 48 individuals as population size, 15 % elitism, binary tournament as selection type, 1-point crossover with 60% crossover and 1% mutation probabilities, respectively.

In Chapter 7 the optimization performance of several algorithms tested on the SELOX benchmark is compared. The studied algorithms include Evolutionary Algorithms, Genetic Algorithms, Simulated Annealing, Taboo Search, and Hybrid Genetic Algorithms. The most successful in the SELOX benchmark optimization proved to be Evolutionary and Genetic Algorithms using the sharing procedure, and Genetic algorithms hybridized with some knowledge-extraction routine. A multi-step DoE optimization procedure was also exemplified using this benchmark. The main differences regarding the applicability of DoE and Global optimization strategies were highlighted, namely that better cause-effect knowledge can usually be obtained with the use of multi-step DoE but the risk of discarding synergetic variables at an early stage is increased. With Global optimization algorithms the achievement of the optimal solution can be theoretically guaranteed; furthermore, this strategy may present an advantage with regard to experimental effort, in relation to DoE, when a very large and complex parameter space needs to be investigated.

Altogether, it was shown that the use of HTE, as a powerful method for catalysis research, is further improved by the application of appropriate experimental planning techniques. Strategies involving Chemometric methods like Design of Experiments and Global Optimization techniques can help in avoiding exhaustive experimental investigations and enhance the power of exploring large parameter spaces and locate the optimum catalytic system conditions.

Samenvatting

Het hoofd onderwerp van dit proefschrift is het onderzoek naar de synergiën tussen *High-Throughput Experimentation* (HTE) en *Chemometric Optimization* methodologiën in katalyse onderzoek en het gebruik van dit soort methoden voor het maximaliseren van de voordelen van het gebruik van HTE methoden.

In hoofdstuk 1 wordt een introductie gegeven over *High-Throughput Experimentation* (HTE) en *Chemometric Optimization* methoden en hun gecombineerde gebruik in katalyse onderzoek. Verschillende methoden van chemometrische experimentele planning worden behandeld zoals Design of Experiments and *Global Optimization* algoritems.

Het gebruik van de *Design of Experiments* techniek wordt behandeld in hoofdstuk 2. Deze methodologie is een belangrijk thema in dit proefschrift aangezien zij in vrijwel alle hoofdstukken wordt gebruikt.

Hoofdstuk 3 gaat over het onderzoeken van een terpeen additie reactie parameter ruimte, gebruikmakend van een meerstaps *Design of Experiments* benadering. Hierdoor werd inzicht verkregen in de belangrijkste reactiviteits trends. De hoogste selectiviteit werd verkregen met een campheen substraat en de laagste selectiviteit met twee pineen substraten. Dit was verklaarbaar door het grotere aantal omleggingen dat pinenen kunnen ondergaan. Aangetoond werd dat de zuursterkte van de katalysator van invloed was op de activiteit, maar niet op de uiteindelijke verhouding van producten. Het type oplosmiddel en de water activiteit waren van beperktere invloed. DoE bleek een goede strategie te zijn voor het ontdekken van de onderliggende trends in deze gecompliceerde chemische parameter ruimte.

In hoofdstuk 4 wordt de optimalisatie van de reactie condities voor de hydrogenering van cyanohydrin esters naar *N*-geacyleerde *β*-amino alcoholen, gebruikmakend van de meerstaps *Design of Experiments* benadering beschreven. Deze strategie maakte het mogelijk om met een beperkt aantal experimenten de locatie van het meest veelbelovende deel van de parameter ruimte te vinden. Dit was met name het gebruik van Ni en Rh katalysatoren op kool of alumina, met isopropanol of dioxaan als oplosmiddel. Door verdere optimalisatie werden de beste condities voor de reactie gevonden. Deze gaven een opbrengst van ongeveer 50% voor benzylische en ongeveer 90% voor alifatische cyanohydrin esters. De reactie was succesvol voor een verscheidenheid aan alifatische en aromatische substraten. Er werd in de regel geen racemisatie gevonden als werd uitgegaan van enantiopure substraten behalve in geringe mate bij benzylische substraten. DoE in combinatie met HTE technologien bleek een efficientë methode voor het onderzoeken en optimaliseren van de chemische route naar *N*-geacyleerde *β*-amino alcoholen.

In hoofdstuk 5 wordt het plannen en modelleren van een *catalytic benchmark response surface* (grafische weergave van de resultaten), gebruikmakend van *Design of Experiments* beschreven. Een bibliotheek van 189 katalysatoren die bestonden uit oxide mengsels werd gesynthetiseerd, getest op activiteit en de resultaten gemodelleerd voor de CO oxidatie reactie. Dit gebeurde in de aanwezigheid van waterstof (SELOX) of zonder waterstof (COOX). Een DoE model (tweede orde) werd gebruikt om de parameter ruimte te beschrijven. Een visuele representatie van de *response surfaces* werd gegenereerd en de analytische functies van het model werden berekend. Deze gaven de *benchmarks* voor de twee geteste condities (COOX en SELOX). De meervoudige responses of interest en de gewenste proces condities (conversie van CO, selectiviteit naar CO_2 en temperatuur) werden meegenomen in de berekening door middel van een *Desirability function*. De visualisering van de *response surface* maakte het mogelijk een goed inzicht te krijgen in de effecten van de onderzochte variabelen en de plaats van de optimale gebieden. De COOX en SELOX *benchmarks* werden in de volgende hoofdstukken gebruikt om de optimalisatie efficiëntie van de geselecteerde *Global Optimization* algoritmes te onderzoeken.

In hoofdstuk 6 wordt de optimalisatie door middel van het gebruik van de *Global Optimization Genetic Algorithm* (GA) besproken en wordt de invloed van algoritme settings op de optimalisatie efficiency onderzocht. De benchmarks voor COOX en SELOX die in hoofdstuk 5 gevonden werden, werden gebruikt voor de validatie van de efficientie van het GA. Door middel van een DoE benadering werden GA configuraties geselecteerd die werden getest en waarvan de effecten van de gekozen parameter settings werden bestudeerd. De meest belangrijke parameter bleek de populatie omvang. De grootste populatie omvang die mogelijk was binnen de gewenste HTE workflow (synthese en testen van een generatie katalysatoren per werkdag), gaf het beste resultaat.

Als optimale GA configuratie werd gevonden voor de optimalisatie van de SELOX en COOX benchmarks: 48 individuen als populatie omvang, 15% Elitism, Binary Tournament als selectie type, 1-point crossover met 60% crossover en 1% mutatie waarschijnlijkheid.

In hoofdstuk 7 wordt de optimalisatie efficiëntie van verschillende algoritmen getest op de SELOX benchmark. De onderzochte algoritmen zijn: Evelotionary algoritmes, Genetic algoritmes, Simulated Annealing, Taboo Search en Hybrid Genetic algoritmes.

Meest succesvol in de optimalisatie van de SELOX benchmark bleken Evolutionary en Genetic algoritmes met gebruik van een Sharing procedure en Genetic algoritmes gehybridiseerd met een Knowledge Extraction routine. Gebruikmakend van deze benchmark werd ook een meerstaps DoE optimalisatie procedure toegepast. De belangrijkste verschillen in de toepasbaarheid van de DoE en *Global optimization* strategien worden belicht, met name het betere inzicht in cause-effect dat in de regel kan worden verkregen door middel van gebruik van meerstaps DoE. Er is echter een risico dat synergistische variabelen al in een vroeg stadium niet meer meegenomen worden. Bij gebruik van *Global optimization* algoritmes kan het bereiken van de optimale oplossing theoretisch worden gegarandeerd. Bovendien biedt deze strategie ook mogelijk het voordeel, in gevallen waarbij een grote en zeer complexe parameter ruimte moet worden onderzocht, om het aantal experimenten hanteerbaar te houden.

Samengevat werd er getoond dat het gebruik van HTE als methode in katalyse onderzoek verder wordt versterkt door het aanwenden van de juiste experimentele planning technieken. Strategieën die bestaan uit chemometriche methoden zoals DoE en *Global Optimization* technieken kunnen helpen bij het vermijden van zeer grote hoeveelheden experimenten en vergroten de kracht van het onderzoeken van grote parameter ruimten en het lokaliseren van optimale condities voor een katalytisch systeem.

Acknowledgements

A thesis is not an individual work; I would like therefore to share my gratitude with the many people that contributed to it:

Professor Thomas Maschmeyer for giving me the chance to initiate this project and for the effort spent trying to keep contact, even from distant Sydney. Professor Geert-Jan Witkamp and Professor Roger Sheldon for taking over the role of thesis advisor in the last stage of my PhD.

A special acknowledgement is given to my supervisor Dr. Jan Kees van der Waal. This thesis would not be possible without his continuous guidance, knowledge, good mood and friendship. Also a special thank to Dr. Roger Downing who supported this project in the last stage, always with encouraging inputs and not giving up contact. These persons were essential for the completion of this thesis.

Thanks to Dr. Claude Mirodatos for giving me the opportunity to do a part of my PhD with a Marie Curie scholarship at CNRS institute in Lyon. To Dr. David Farrusseng and Dr. Frederic Clerc for the very fruitful and good collaboration, on the Global Optimization Algorithms project. And to all the people present at CNRS that contributed for a very pleasant stay.

Thanks to Dr. Lars Veum for his particular approach to life and collaboration in the cyanohydrins research work. Thanks to Dr. Ulf Hanefeld for giving me, and HTE, the opportunity for collaboration. To my students Robert Antonijs and Arthur van Houten that worked with me in some of this thesis chapters. To Adrie Knol for the instruction and long hours passed at the GC-MS. And thanks to other members of the BOC-TOCK departments: Isabel Arends, Ulf Hanefeld, Fred van Rantwijk, Leen Maat, Koos Jansen, Joop Peters, Mieke Jacobs, Mieke van der Kooij, Remco, Anton and Kristina.

I wish to thank Avantium for its support to my thesis and for allowing me the use of its facilities. To the people who during my PhD were in various ways involved in my work: JK, René, Arie, Jos, Erik-Jan and Ernst. And to all the others I have met subsequently as colleagues, and who had a motivating influence on the finalisation of this thesis.

During my PhD years many people have passed through the BOC-TOCK departments, which added constant interest to the group: Allard, Alexandra, Andrea, Andreij, Angel,

Anne B., Annemieke, Arné, Bruno, Carla, César, Chretien, Cindy, Christian, Daniel, Dany, Daniela, Diane, Dirk, Fatma, Frank, Goran, Goska, Hilda, Jeroen, John, Karolin, Lars, Leszek, Lingqiu, Luca G., Luca F., Luuk, Marco, Martijn, Marina, Maikel, Menno, Mike, Michiel H, Michel V., Naseem, Paloma, Paolo, Paulo, Pedro, Sander, Sameh, Seda, Silvia G., Scoob, Stephan, Tim, Ton, Van Vliet, Xavier, Zhiping, Zbigniew and many others. From this group special acknowledgement for Mike and Luuk, my Erasmus supervisors, and later on colleagues and friends. Daniela for the surprising beginning of a nice friendship. And to both Michiel van Vliet and Hacking for all their good friendship and support.

To Alexandre for the long talks and exchange of architectural ideas. And to the colleagues and friends that I have met in conferences and courses during my PhD, for the useful sharing of experiences and ideas.

To Paul, Ceren, Josep, Ralph, Mark, Daniel, Mac, Marta, Vincent and Valentina for some nice times shared. To the “Getugatz” group for bringing a bit of Portugal to Delft.

The girls’ dinner club: Helene, Paloma, Rute, Marta, Anne B., Gerdien, Esther and Carla. Thanks for all the fun, girls!

Paolo for taking me to the world of strange movies, music and travels, and for a close friendship. Paloma for your strength and discrete and sharp humour. Carla for your good friendship. And to Rute for being my family away from home.

To Teresa for accompanying me as a friend even to more than 2000 km distance. And to all my friends in Portugal for keeping in touch.

To my family for all the support and motivation to endure my PhD experience.

And the most special thanks to Antonio for always being close to me.

Sílvia

Publications

- S. R. M. Pereira, R. Antonijs, J. C. van der Waal and T. Maschmeyer, High-throughput Methodology combined with DoE applied to the exploration of a complex chemical system: terpene addition reactions, **2008**, submitted to *Comb. Chem. & High Through. Screen*.
- S. R. M. Pereira, F. Clerc, J. C. van der Waal and T. Maschmeyer, Optimisation methodologies for High Throughout Research in Catalysis: Selox benchmark, *Comb. Chem. & High Through. Screen.*, **2007**, 10, 149-159.
- L. Veum, S. R. M. Pereira, J. C. van der Waal and U. Hanefeld, Catalytic hydrogenation of cyanohydrin esters as a novel approach to N-acylated β-amino alcohols: reaction optimisation employing a Design of Experiment approach, *Eur. J. Org. Chem*, **2006**, 7, 1664-1671.
- S. R. M. Pereira, F. Clerk, D. Farrusseng, J. C. van der Waal, C. Mirodatos , T. Mashmeyer, Effect of Genetic algorithm parameters on the optimisation of heterogeneous catalyst, *QSAR Comb. Sci.*, **2005**, 24, 45-57.
- F. Clerc, M. Lengliz, D. Farrusseng, C. Mirodatos, S. R. M. Pereira and R. Rakotomalala, Library design using genetic algorithms for catalyst discovery and optimisation, *Rev. Sci. Instrum.*, **2005**, 76 (6), 062208.
- L. M. Van Langen, M. H. A. Janssen, N. H. P. Oosthoek, S. R. M. Pereira, V. K. Svedas, F. van Rantwijk and R. A. Sheldon, Active site titration as a tool for the evaluation of immobilisation procedures of penicillin acylase, *Biotechnol Bioeng.*, **2002**, 79(2), 224-8.
- M. H. A. Janssen, L. M. van Langen, S. R. M. Pereira, F. van Rantwijk and R. A. Sheldon, Evaluation of the performance of immobilised penicillin G acylase using active-site titration, *Biotechnol. and Bioeng.*, **2002**, 78(4), 425-32.

Oral presentations and posters

- S. R. M. Pereira, J. C. van der Waal and T. Maschmeyer, Comparison of Algorithms optimization performance on the Selox benchmark, 9th Scandinavian Symposium on Chemometrics (SSC9), Reykjavik, Iceland, August 21-25, **2005**: poster.

- S. R. M. Pereira, L. Veum, J. C. van der Waal and U. Hanefeld, High-Medium-Low Throughput: An efficient optimisation methodology for new reactions (reduction of acylated cyanohydrins), Eurocombi3, Winchester, United Kingdom, 18-21 July, **2005**: oral presentation.

- S. R. M. Pereira, J. C. van der Waal and T. Maschmeyer, Optmization using Genetic Algorithms in catalysis research, NCCCVI, Noordwijkerhout, Netherlands,7-9 March, **2005**: poster.

- S. R. M. Pereira, J. C. van der Waal and T. Maschmeyer, Effect of GA parameter settings in the algorithms optimization, Chemometrics in Analytical Chemistry IX (CAC IX), Lisbon, Portugal, September 19-22, **2004**: poster.

- S. R. M. Pereira, J. C. van der Waal and T. Maschmeyer, Optimization of terpene hydration reactions, Netherlands Catalysis and Chemistry Conference IV (NCCC4), Noordwijkerhout, Netherlands, March 10-12, **2003**: poster.

- S. R. M. Pereira, J. C. van der Waal and T. Maschmeyer, Optimization in catalysis research: DOE and GA, French-German Summer School on Acceleration and Miniaturization in Catalysis (AMCat), Lyon, France, July 1-4, **2003**: poster.

- S. R. M. Pereira, J. C. van der Waal and T. Maschmeyer, Terpene hydration reactions study using Combinatorial catalysis, EuroCombiCat, European Workshop in Combinatorial Catalysis, Ischia, Italy, June 2-5, **2002**: poster.

- S. R. M. Pereira, J. C. van der Waal and T. Maschmeyer, Terpene hydration reactions study using Combinatorial catalysis, Combicat, Lisbon, Portugal, April 11-12, **2002**: poster.

- S. R. M. Pereira, J. C. van der Waal and T. Maschmeyer, Design of Experiments and HTE in terpene chemistry, Netherlands Catalysis and Chemistry Conference III (NCCC3), Noordwijkerhout, Netherlands, March 4-6, **2002**: poster.

Curriculum vitae

Silvia R. M. Pereira was born on the 10th of March 1976 in Bragança, Portugal. Following her science secondary school graduation in 1994 she started her university studies in Chemical Engineering at the Instituto Superior Técnico from the Technical University of Lisbon. Her practical graduation project entitled “Active site titration of Penicillin G acylase immobilised on Eupergit C and performance in cephalexin synthesis” was realized in the Biocatalysis and Organic Chemistry Department of the Delft University of Technology, The Netherlands, within the framework of the ERASMUS program. In 2001 she obtained her diploma in Chemical Engineering with specialization in Biotechnology. In the same year she started her PhD research in the Laboratory of Applied Organic Chemistry and Catalysis, initially under the supervision of Prof. T. Maschmeyer and later of Prof. R. A. Sheldon (Laboratory of Biocatalysis and Organic Chemistry) and Prof. G. J. Witkamp, and also of Dr. J. C. van der Waal, at the Delft University of Technology. From January to August 2004 she carried out a Marie Curie PhD interchange on the subject of “Application of Genetic algorithms in heterogeneous catalysis research” in the group of Prof. Claude Mirodatos at the CNRS - Institute de Recherches sur la Catalyse, in Lyon, France. On September 2006 to September 2007 she accomplished a post-graduate course on “Chemistry applied to cultural heritage” at the Universidade Nova de Lisboa, Portugal. From January 2008 onwards she has been working as Junior Project Leader at the Predictive Technologies Department of Avantium Technologies B.V. in Amsterdam. The results of her PhD research are described in this thesis.