STUDY OF CATALYTIC REACTION KINETICS USING STRONGLY TYPED GENETIC PROGRAMMING

by
Arman Basmacıoğlu
B. S. in Ch.E., Yıldız Technical University, 2006

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Chemical Engineering
Boğaziçi University
2009

to my family

ACKNOWLEDGEMENTS

Firstly, I would like to express my sincere thanks to my thesis supervisor, Assoc. Prof. Ramazan Yıldırım, who has given continuous encouragement, support and trust. I am grateful to him for his patience.

My sincere gratitude is due to my thesis committee members, Asst. Prof. Dr. Ahmet Kerim Avcı and Prof. Fikret Gürgen who have read and commented on my thesis.

Special thanks to Pınar Kanlıkılıçer, Ezgi Akkaya, Yasemen Güngörmez, Banu Yavuztürk and İlker Öztürk for giving me their friendship. Their friendship and support have given me strength.

Cordial thanks for my friends Murat Gostanyan, Burag Karayan and Zafer Natan for sharing all good and bad moments with me.

Finally, I wish to thank to my family for their support throughout my entire education. This thesis would not have been possible without them. They gave me everlasting support whenever I needed and a great source of love for me throughout the years; this thesis is primarily for them.

ABSTRACT

STUDY OF CATALYTIC REACTION KINETICS USING STRONGLY TYPED GENETIC PROGRAMMING

Kinetic modeling of some generic and real catalytic reactions systems were performed using strongly typed genetic programming. The reaction rate laws were created using the data generated for generic reactions and experimental data reported for the real reactions in the literature, and the structure of these equations were used to predict the possible reaction mechanisms as an aid to the traditional kinetic methods. It was found that the strongly typed genetic programming techniques are quite successful predicting the simple model arising from a simple mechanisms or complex mechanisms with simplifying assumptions. This is especially true for the surface reaction controlled systems which constitute about the 75% of the heterogeneous catalytic reactions. However, this method is not effective in the presence of complexities such as too many terms in the denominator of the rate equation, square term arising from the control of the surface reaction step in dual site reactions, square root terms in the dominator due to the dissociative nature of the reaction. Then it was concluded that the method may not be sufficient to understand the catalytic mechanisms by itself but it may help to reduce the work load required in the traditional Langmuir-Hinshelwood Hougen Watson approach that is commonly utilized for the kinetic studies of the catalytic reactions.

ÖZET

MODULER GENETİK PROGRAMLAMA İLE KATALİTİK REAKSİYON KİNETİĞİNİN ÇALIŞILMASI

Bazı kuramsal ve gerçek katalitik reaksiyon sistemlerinin kinetic modellemesi modüler genetic programlama kullanılarak yapılmıştır. Kuramsal reaksiyonlar için üretilen ve gerçek reaksiyonlar için literatürde yer alan veriler kullanılarak reaksiyon hız denklemleri türetilmiş ve bu denklemlerin matematiksel yapısından faydalanılarak olası reaksiyon mekanizmaları tahmin edilmeye çalışılmış, böylece genetik programlamanın geleneksel kinetik çalışmalara katkı yapıpı yapamaycağı test edilmiştir. Basit reaksiyon mekanizmalarından doğan ya da karmaşık mekanizmaların çeşitli kabullerle sadeleştirilmesi ile ortaya çıkan basit modellerin kestiriminde modüler genetic programlama tekniklerinin başarılı olduğu anlaşılmıştır. Bu durum özellikle heterojen katalitik reaksiyonların % 75'ini oluşturan hızın yüzey reaksiyonu tarafından kontrol edildiği sistemler için geçerlidir. Hız denkleminin paydasında fazla terim olması, yüzey reaksiyon kontrollü çift aktif bölgeli reaksiyonların hız denklemlerinin paydasının karesinin alınması gereği veya ayrışmalı reaksiyonların hız denklemlerinin paydalarında kareköklü ifadelerin bulunması gibi karmaşık durumlarda bu metodun kullanımı etkili olamamaktadır. Bu sebepten dolayı metodun tek başına katalitik reaksiyon mekanizmalarının anlaşılmasında yeterli olmadığı ancak genellikle katalitik reaksiyonların kinetic çalışmalarında kullanılan geleneksel Langmuir-Hinshelwood Hougen Watson yaklaşımının iş yükünün azaltılmasında yardımcı olabileceği sonucuna varılmıştır.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS/ABBREVATIONS

$C_{i.S}$	Coverage of specie i on active sites of the catalyst
C_v	Ratio of vacant sites
k_i	Rate constant of step i
k_i^f	Rate constant of forward reaction of step i
K_i	Equilibrium constant of reaction step i
P_i	Partial pressure of component i
r_i	Reaction rate according to component i
R^2	Regression coefficient

EA	Evolutionary Algorithm
GA	Genetic Algorithm
GP	Genetic Programming
LHHW	Langmuir–Hinshelwood Hougen Watson
STGP	Strongly Typed Genetic Programming
WGS	Water-Gas Shift Reaction

1. INTRODUCTION

The kinetic modeling of reactions has a great importance in the determination of the reaction mechanisms and the rates, which are essential to effectively design and operate reaction systems and processes. Hence the subject is always given a great emphasis by the researchers.

The detailed kinetic analysis for the optimization of the chemical reaction processes, to improve performance, to increase safety and to prevent by products, requires a time consuming study of complex reaction mechanisms. For this reason, instead of dealing with complex chemical reaction processes, the fast and effective methods are searched by developers and producers. The approximated kinetic models with the contribution of those methods can be used to establish fundamentals for chemical reaction process optimizations in further steps (Cao *et al.*, 1999).

In recent years, evolutionary algorithms (EAs) having attributes of self-adaptation, self learning, self-organization generality and intrinsic parallelism have been utilized successfully in many areas. If the application areas of economy, engineering, chemistry and scientific computation are taken as references, these algorithms can be thought to be successful in kinetic modeling of chemical reactions.

The definition of EA originally comprises three methodologies that are genetic algorithms (GAs), evolutionary programming (EP) and evolution strategies (ES). In addition to those, a new methodology called genetic programming (GP) was introduced by John Koza (Koza, 1994). GP is a method that creates programs for solution automatically by comprehending the nature of the problems directly from input-output correlation with the help of fitness function taken as driving force.

GP uses tree structured representation for the individuals of population in which the branch nodes consist of arithmetic operations and elementary functions that are defined in function set and leaf nodes include variables, constants and functions of no arguments coming from terminal set. The evaluation of fitter individuals is carried out with the contribution of fitness cases. The next generation is created by using genetic operators like crossover and mutation after the selection of fitter programs. This iteration continues until a predefined termination criterion is reached.

There are four steps in the working procedure of GP. The algorithm starts with the generation of an initial population consisting of random combinations of predefined functions and terminals of the problem. Following that step, the fitness of each program in the population is evaluated by execution, as a degree of how well it solves the problem. The third step is the creation of new population of computer programs by implementing the crossover and mutation operators or copying the best individual. Finally, the best individual that was found is presented as the result of genetic programming when any of defined termination criterions is reached.

Genetic programming searches a huge space consisting of unrestricted set of possible compositions of problem's functions and terminals. However, for certain problems solution candidates should have a certain structure. Strongly typed genetic programming (STGP) is a grammar that constrains the solution candidates to a specified structure desired in generated solutions with the contribution of modified initial population generation and genetic operator implementation procedure.

The traditional Langmuir-Hinshelwood Hougen Watson approach, which is based on proposing a mechanism and appropriate rate law equation and testing it using the experimental data, is a time consuming algorithm in order to discover the right reaction mechanisms and the rate law because there are many possibilities in the elementary reactions level. STGP can be helpful to test the approximate form of the rate law equation from the experimental data without any prior assumption of the mechanism. Then the mechanism may be based on this equation so that a more detailed analysis can be done in a shorter time.

In this thesis, the kinetics study of various reactions, of which the rate data, rate laws and the mechanisms reported in the literature, was performed using STGP program created by the modification of GPLAB, which is a genetic programming toolbox created for MATLAB by Sara Silva. The modifications were made in order to match the requirements of achieving successful results for modeling surface reaction controlled systems, which constitute at least about the 75% of the catalytic reactions (Fogler, 1992). The work started with some generic and well established reaction laws (with the generic data created) to test the power and the limits of the approach and extended to the real reactions reported in the literature. Then the rate law equations were compared to the ones proposed in literature and analyzed to predict the possible reaction mechanism.

Chapter 2 contains a literature survey on reaction kinetics followed by information about genetic programming, strongly typed genetic programming and GPLAB, the media that our proposed algorithm was implemented. Chapter 3 contains the computational work done while the attained results were presented in Chapter 4. Finally, the conclusions and recommendations are summarized in Chapter 5.

2. LITERATURE SURVEY

2.1. Reaction Kinetics

The characterization and the deduction of mechanism involved in the chemical processes are carried out by experiments in kinetic studies. Properties of the reaction such as the concentration and temperature dependency of the reaction rate, the nature of all components used in the system and their concentrations and time dependencies are important to be measured (Steinfeld, 1989). By the help of experimental data, the consistency of the suggested mechanisms can be realized by transforming it into a mathematical model.

The determination methodology of catalytic and heterogeneous mechanisms is named in the literature as Langmuir Hinshelwood Hougen Watson (LHHW) approach for the reason of the derivation from the ideas proposed by Hinshelwood based on Langmuir's principles for adsorption (Fogler, 1992). Due to the popularization of this approach by Hougen and Watson, their name was also included.

The first step of LHHW is the choice of mechanisms among molecular or atomic adsorption, and single- or dual-site reaction. The adsorbance of a molecule whether being nondissociative or dissociative is dependent on the surface. For dissociative adsorption, two adjacent vacant sites are needed while single vacant site is required by molecular adsorption. The representation of nondissociative and dissociative adsorption is respectively as follows:

$$\mathrm{AB} + \mathrm{S} \leftrightarrow \mathrm{AB.S} \tag{2.1}$$

$$\mathrm{AB} + 2\,\mathrm{S} \leftrightarrow \mathrm{A.S} + \mathrm{B.S} \tag{2.2}$$

The resulting rate of attachments for the adsorptions mentioned above is subsequently:

$$r_{AD} = k_A\left(P_{AB}\,C_v - \frac{C_{AB.S}}{K_A}\right) \tag{2.3}$$

$$r_{AD} = k_A\left(P_{AB}\,C_v^{\,2} - \frac{C_{A.S}\,C_{B.S}}{K_A}\right) \tag{2.4}$$

For the constitution of the reaction product, a reactant can react in three ways after being adsorbed onto the surface. These three ways are named as single site, dual site and Eley-Rideal. The only site on which the reactant is adsorbed participates the reaction in single site reactions while the adsorbed reactant interacts with another site in dual site reactions. In Eley-Rideal mechanisms, the reaction occurs between an adsorbed molecule and a molecule in the gas phase. The reactions for single site, dual site and Eley-Rideal mechanisms are as follows respectively:

$$A.S \leftrightarrow B.S \tag{2.5}$$

$$\left.\begin{aligned} &A.S + S \leftrightarrow B.S + S \\ &A.S + B.S \leftrightarrow C.S + D.S \end{aligned}\right\} \tag{2.6}$$

$$A.S + B(g) \leftrightarrow C.S \tag{2.7}$$

The concerning rate laws for three mechanisms are:

$$r_S = k_S\left(C_{A.S} - \frac{C_{B.S}}{K_S}\right) \tag{2.8}$$

$$\left.\begin{aligned} r_S &= k_S\left(C_{A.S}\,C_v - \frac{C_{B.S}\,C_v}{K_S}\right) \\ r_S &= k_S\left(C_{A.S}\,C_{B.S} - \frac{C_{C.S}\,C_{D.S}}{K_S}\right) \end{aligned}\right\} \tag{2.9}$$

$$r_S = k_S\left(C_{A.S}\,P_B - \frac{C_{C.S}}{K_S}\right) \tag{2.10}$$

What follows the surface reaction session is desorption of the products of surface reactions into the gas phase. For the desorption of a species (e.g., C):

$$C.S \leftrightarrow C + S \tag{2.11}$$

Consequently the rate of desorption is

$$r_D = k_D \left(C_{C.S} - \frac{P_C C_v}{K_D} \right) \tag{2.12}$$

Subsequently assumed to be reversible, all the rate laws for the individual steps are written. The coverage dependent terms are eliminated with the use of the steps that are not rate limiting after the assumption of rate limiting step. A site balance is formed and with the expressions for the concentration of the adsorbed species being substituted, the rate law is derived via the rate limiting step. The rate equation derived is confronted with the experimental data and the criterion for acceptance is the values of all the constants to be positive. Constants are calculated by linear regression. If the constants are not appropriate, then a rate limiting step or a different mechanism is chosen and this algorithm is implemented all over again. In the cases where two or more models agree, then the rate equation fitter than the others is preferred.

LHHW approach can be summarized as a six step algorithm with isomerization of n-pentene (N) to i-pentene (I) over alumina given as an example.

$$N \xleftrightarrow{Al_2O_3} I \tag{2.13}$$

1. Mechanism is selected as single site for isomerization of n-pentene.

$$\text{Adsorption: } N + S \leftrightarrow N.S \tag{2.14}$$

$$\text{Surface reaction: } N.S \leftrightarrow I.S \tag{2.15}$$

$$\text{Desorption: } I.S \leftrightarrow I + S \tag{2.16}$$

2. A rate limiting step is assumed. Surface reaction is chosen for our example. The rate law for the surface reaction step is

$$-r_{N'} = r_S = k_S \left(C_{N.S} - \frac{C_{I.S}}{K_S} \right) \tag{2.17}$$

3. The expression for the concentration of the adsorbed species is found by the assumption that the forward and the reverse rates are in equilibrium in the steps that are not limiting. For the example of n-pentene isomerization,

$$\frac{r_{AD}}{k_{AD}} \cong 0 : \rightarrow C_{N.S} = P_N K_N C_v \tag{2.18}$$

$$\frac{r_D}{k_D} \cong 0 : \rightarrow C_{I.S} = \frac{P_I C_v}{K_D} = K_I P_I C_v \tag{2.19}$$

4. The site balance is written as.

$$C_t = C_v + C_{N.S} + C_{I.S} \tag{2.20}$$

5. The rate law is derived by combining steps 2, 3, and 4.

$$-r_{N'} = r_S = \frac{\overbrace{k_S C_t K_N}^{k}(P_N - P_I / K_P)}{(1 + K_N P_N + K_I P_I)} \tag{2.21}$$

6. The rate law derived in step 5 is compared with the experimental data. If there is an agreement and all of the parameters are positive, and the model predicts the experimental data successfully, then the derived rate law has a great possibility of originating from the correct mechanism and rate determining step. In order to try different mechanisms, the same procedure should be followed. For example if the following mechanism is being tested,

$$N + S \leftrightarrow N.S \tag{2.22}$$

$$N.S + S \leftrightarrow I.S + S \tag{2.23}$$

$$I.S \leftrightarrow I + S \tag{2.24}$$

with the rate determining step being surface reaction, the rate law becomes

$$-r_N' = \frac{k\,(P_N - P_I / K_P)}{(1 + K_N P_N + K_I P_I)^2} \tag{2.25}$$

If two or more models agree, then the model with better fitness is preferred (Fogler, 1992).

2.2. Evolutionary Algorithms

The attention on operational procedures is increased due to the significance of the product quality, process performance and the environmental concerns in the chemical industries (McKay *et al.*, 1997). Chemical engineering production processes are always in need of high production knowledge reflected by mathematical modeling methods. The complex stoichiometry and thermodynamics are required to be studied for the kinetic analysis of chemical reactions in complex systems. Furthermore, kinetic analysis of heterogeneous reactions concern many factors requiring detailed knowledge of physics and chemistry like mass diffusion, interfacial area, diffusion coefficient and many other factors that affect the reaction rate (Cao *et al.*, 1999). Spending a lot of time on studying of detailed mechanism of complex reactions is not worthwhile for industrial applications due to the intricacies of the chemical reactions. Consequently, for the purpose of attaining approximate kinetic models and optimizing the processes for these complex reaction systems fast and effective methods are needed (Cao *et al.*, 1999). For that reason researchers have preferred to the utilization of evolutionary algorithms (EA's) which are implemented in a wide range of engineering and scientific computations (Goldberg, 1989). EA is a subset of evolutionary computation in which the main idea consists of a given population of individuals and environmental pressure causing a rise in the fitness of the population (Eiben and Smith, 2003). These adaptive methods are utilized for finding solution for not only for computational problems, but also for the applications in chemistry.

There are three well known types of EA that are named as genetic algorithms (GA), evolutionary programming (EP) and evolution strategies (ES) that differs in the utilization details and the bases of application fields. Under the leadership of John Koza, a new algorithm called genetic programming (GP) is added to the group. This new algorithm differs from others by its variable length and the representation of the solution being created as a computer program.

2.2.1. Genetic Algorithms

Genetic Algorithms drawn into consideration and improved by John Holland in 1960’s has been studied in detail, experimented and implemented in various applications of engineering. Theoretical knowledge of genetic algorithms is provided in Holland’s book named Adaptation in Natural and Artificial Systems (1975). Subsequently the mechanisms utilized by GA have been examined and clarified by Goldberg (Goldberg, 1989) and some other authors. With the contribution of improved ways of importing the mechanisms of natural adaptation into computer systems GA’s supply alternative methods in the solution of problems.

Optimization, machine learning and intelligent search are the main implementation areas of GA that has also been utilized by many researchers for different problem domains including engineering field; function optimization (Goldberg, 1989), the control of natural gas pipelines (Goldberg, 1989), the prediction of protein conformation (Schulze-Kremer, 1992), process control (Nordvik and Renders, 1991) and heat exchanger design (Androulakis and Venkatasubramanian, 1991).

The nomenclature from the natural genetics is used for the explanation of the algorithms in GA. The structure named as chromosome encrypts how the solution is to be constructed. Sometimes more than one chromosome may be required for the specification of the solution. Genotype and phenotype are named as the set of chromosomes and the resulting solution respectively. The implementation of GA starts with a population consisting of chromosomes of possible candidate solutions. Individuals of the population are displayed as strings composing of 0’s and 1’s. Subsequent to the start of the algorithm by random creation of individuals, evaluation is carried out over generations in which the fitness of each individual evaluated and the best individuals with high fitness are chosen to form a new population. The type of the selection process determines the way best individuals survive. The application of genetic operators, named as crossover and mutation, follows the selection process. Crossover handles a couple of individuals by swapping segments to each other and attains two new individuals. In addition to this, mutation provides uniqueness between the members of the population by modifying gene values of the individuals. Until the termination criterion is reached, newly created individuals will be used in the next iteration steps. The iteration process stops and the solution is found when the termination criteria are reached that is defined as a fitness value or maximum iteration number.

2.2.2. Genetic Algorithms in Chemical Engineering

Genetic algorithms have been commonly implemented in chemical engineering problems recently due to its attributes of being a fast and effective method for optimization and parameter estimation. Kinetic modeling and parameter estimation is studied in many cases with the use of genetic algorithms in recent years and the interest in these implementations is increasing.

The benefits of genetic algorithm for creating initial parameter estimations for kinetic models of catalytic processes were studied by Moros *et al.* (1996). The purpose in this study was to provide suitable initiation points for the applied combination of an integration process and a locally converging nonlinear parameter estimation algorithm. The implementation of GA was fulfilled in Turbo-Pascal programming language medium and the results were achieved reflects that implementation of GA saves time in computing and the solution of model parameters is reliable. Apart from that, the control parameters like mutation rate, the selection method and the population size play important role in the convergence rate of the algorithm and in the accuracy of the solution. The determined values for those parameters were 300 for population and 80 for mutation rate. In addition to this, the selection method was selected as fitness proportionate and ranking.

In the work of Rein *et al.* (2006), that also used genetic algorithms, the kinetics of polyurethane (PU) foam in smoldering combustion was tried to be find out. The kinetic parameters were calculated and the results were in a good match with the experimental data. The first aim was to develop a methodology for obtaining a global mechanism of thermo oxidative decomposition, followed by the objective of proposing a global mechanism of PU kinetics with five reactions based on the experimental data. Subsequent to that, the numerical reproduction of the experiments with proposed mechanisms done via a developed lumped model. In this study, the kinetic parameter values were chosen as the individuals of the population meanwhile the environment leading to the improvement of fitness values was consisting of the experimental results. It was concluded that the kinetic parameters evaluated and the five step mechanism were suitable for the estimation of thermo gravimetric data at various heating rates, and with the contribution of developed kinetics, the experimental observations of the species distribution in both direction could be possibly predicted.

In recent years, a popularization in GA applications can be examined in scientific and engineering disciplines comprising chemistry (Dawkins, 1999). In addition to this, the number of examples in kinetic modeling studied with the help of GA is in increase, and reasonable results have been attained. The great proportion of them takes the case as the prediction of initial estimates in diverse applications.

The kinetic modeling of Fischer-Tropsch synthesis (FTS) and Water Gas Shift (WGS) reactions that are used to obtain environmentally safer fuel in gas to liquid processes were studied by Boozarjomehry and Masoori (2007). For comparison of the models, the experimental data was taken from Van Der Laan *et al.* (2000). Langmuir-Hinshelwood-Hougen-Watson model based rate equations were utilized, and the optimum values of the unknown parameters were attained by the minimization of the overall error in the prediction of the rate of reactions with the help of GA. The tabulated average percent errors were spread in a wide range from 22.8 to 86.8. Nevertheless the big proportion of the outputs with errors below 27 proved the precision and the majority of the suggested kinetic models.

Other than this, the Binary and Decimal encoded GA have been compared to each other according to their performance. It was concluded that Binary GA could converge faster and had a higher possibility to become nearer to the global optimum being more independent from problem type.

2.2.3. Genetic Programming

The optimization of chemical processes, the prevention of secondary reactions, the maximization of the yield and the improvement of safety of operation are some of the major and time consuming issues in chemical engineering production process. The studies of the complex reactions that are given remarkable emphasis on are in need of a detailed knowledge in order to be decoded. Speed of agitation, interfacial area, diffusion coefficients, mass diffusion and further factors have great effect especially in the reaction rates of heterogeneous reactions (Cao *et al.*, 1999). Consequently, the efficient and practical methods are preferred by the engineers and manufacturers to overcome the amount of time for being wasted on dealing with those concerns. The tendency modeling and the approximations are the methods that sound most familiar to researchers for getting rid of that kind of problems (Filippi *et* al., 1986). Approximation methods only try the functions comprising polynomials, trigonometric series splines or some other functions to fit the experimental data by arranging the parameters without considering the structure of the system (Galvan *et al.*, 1996). In the case of the tendency model implementation, the mass and energy balances are taken into account in order to simplify the complex reactions to basic kinetic reactions, and the kinetic parameters are estimated with the contribution of regression, integral and differential method (Filippi *et al.*, 1986). From time to time, those methods may not be appropriate for chemical engineering applications, and the genetic programming, with its ability of learning nonlinear input-output relationship, can be used to overcome the drawbacks of the methods mentioned.

Genetic programming is an algorithm that generates solutions to the specified problems automatically via natural selection. John R. Koza is responsible from the proposal of GP and has great impact behind the increasing demand for this method in the field of computer based solving systems. The problem fields as symbolic regression, control, robotics, games and classification has been experimented for the implementation of the GP algorithm written by Koza in LISP computation language (Willis *et al.*, 1997). Koza made his name noticeable and earned his reputation with his publications that established basic algorithms and with his continuing studies in the improvement of the methodology.

Same as it was in GA, the working principle of GP is built on the biological evolution. The main difference between GA and GP are the utilization of tree structures and variable length chromosomes in GP while that of fixed length in GA. Also the solution is represented in the expression of what computer language used for the implementation of GP and chromosomes are coded in a relating way to the problem; on the other hand, GA create the solution represented as a string of numbers. Survival, reproduction, continuation of the most fit individuals in the population to adapt as a result of environmental pressures and utilization of genetic operators such as reproduction, mutation and crossover are also common both in GA and GP.

GP commences with random generation of the individuals consisting a population. The structural variety in the population is supplied by the genetic information of every individual that differentiates survivability of each of them at the same time. The variation in the population is provided due to the existence of genetic operators. Tree structure is utilized for the symbolization of the individuals that can provide many options including the circumstances of the computer language. Fitness, which is a measurement of relevance that the survival of individuals is bound to, is the criterion for determination of the best program in each intermediate and final generation. The error between the evaluated and observed data establishes the base of fitness functions (McKay *et al.*, 1997). Alterations in fitness functions are also possible and there are examples of it in the literature. The utilization of correlation coefficient between the predicted and actual outputs is also proposed in his work.

The definition of functional set that generally consists of mathematical functions such as addition, multiplication, square root logarithm or exponential has a great impact on the convergence to a potential solution to a given problem. The generated programs are combination of functions and terminals that are found from a search space whose members consisted same structure.

The implementation of GP initializes with the creation of population composed of random individuals. Afterwards, the fitness of each individual is evaluated on which, the member selection for reproduction is based. There are various types of selection procedures that exist in the literature. One of them is called tournament selection in which a random sampling of individuals in a fixed number from the parent population is executed to form a subset. The fittest member present in this subset is selected for reproduction. There is another method named as fitness proportionate selection. In this method, member of the population is sampled from the former population with a probability proportional to its fitness. The last selection method is elitist scheme in which the members of the population are ranked according to their fitness and the ones with the highest fitness value undergo reproduction. Subsequent to the selection process, the execution of genetic operators can be carried out. Those genetic operators are reproduction, crossover and mutation. Without making any change, reproduction operator copies an individual from the former population to the next generated one. Crossover operator takes two members of the population and combines them to create new offspring while mutation randomly changes the structure of the proposed model. These operations are performed and the newly generated population replaces the former ones and repeated until a specified termination criteria is met which can be a fitness value or a maximum generation number.

2.2.4. Genetic Programming in Chemical Engineering

Genetic programming has been implemented with great success in various scientific computation and engineering applications due to its properties of being fast and efficient (Mitchell, 1996). For the reason that development and estimation of complex reaction models is time consuming and requiring knowledge of reaction kinetics, GP can be preferred over the traditional methods (Cao *et al.*, 1999). The eligibility of GP is higher than that of GA because of GA's attribute of optimizing model parameters. On the other hand, GP can optimize the model structure and supply a better understanding of not only the parameters, but also the appropriateness of the model structure.

The kinetic modeling of thermolysis reaction of chloro-cyclohexane and the reaction between formaldehyde and carbomide in aqueous solution to give methylol urea were studied by Cao *et al.* (1999). The modeling of ordinary differential equations of the chemical reactions was done by a hybrid evolutionary modeling algorithm (HEMA) comprising two processes at the same time. The structure of the models were optimized by the GP algorithm while model optimization was fulfilled with GA inserted into GP. Tournament selection was chosen as the selection strategy while elitism scheme was also adopted, keeping the best individual of the population in the next generation. Fitness calculation was performed by modified Euler method with fixed step size 0,01 to do the integration during the modeling process. The function and terminal sets composed of (+, -, *, /, x^y, sin, cos, exp, ln) and (x_1, ..., x_n, t, c) respectively were used for evolutionary modeling where n is the number of equations in the system of ODE's and c is the set of random constants. The modeling results were concluded to be reflecting the kinetic characteristics of the thermolysis reaction. A population size of 50 per generation, a maximum tree depth of 3 and a maximum generation of 50 are taken as the parameter settings. The results were also containing an exponential function in the expression that was commented as a success as not being able to be imagined by human mind. The success of HEMA algorithm was proved by the average prediction error of the best models yielded as 0.16 and average fitting error be calculated as 0.029. The results were having an exact match with the predicted values based on the literature values. The effectiveness of HEMA is reflected from attainment of both the parameters and structures of generated models with great success.

Willis *et al.* (1997) used GP for the estimation of both the structure and the complexity of the models. A steady state model was created before the development of dynamic input output models with the algorithm. Vacuum distillation column and twin screw cooking extruder systems were studied. The main purpose in the study targeting vacuum distillation column was to propose a model that would be helpful in comprehension of the product composition. Parameter settings for the input-output data were done as follows; population size selected as 25 and maximum generation as 50. From the result it could be deduced that GP algorithm was able to discover the exact model automatically. 100 generations and population size of 30 were utilized as parameter settings for the study comprising twin screw cooking extruder. A conclusion of GP being capable of eliminating inappropriate model inputs automatically was reached.

In the work of McKay *et al.* (1997) steady state modeling of chemical process was studied by using GP. The system to be modeled was composed of two CSTR tank reactors in series. Population size and the generation number were preferred as 25 and 60 respectively while the selection method was chosen as fitness proportionate method. The probabilities of genetic operators were 0.2 and 0.8 for crossover and mutation in sequence. Also a function set of (+, -, /, *, x^y, $x^{1/2}$, x^2, log, exp) was utilized. The performance of the program was quantified with the root mean square error that was calculated as 0.06 as a reflection of attainments. It was deducted that GP could provide information about the characteristics of the process while neural network were not capable of that kind of application. Several numbers of implementations in the process system engineering indicated the efficiency and effectiveness of GP.

Dynamic system modeling was also experimented as an implementation field of genetic programming. The process output values can be predicted with known input values with the help of system identification known as developing process models from the plant data. Hinchcliffe and Willis (2003) studied dynamic system modeling using a multi-basis function GP algorithm for the aim of discrete-time model development. The main difference of the algorithm used in this study was its ability to use former values of model terms although they are not specified explicitly in the function and terminal sets. It was concluded that in spite of the more computational effort need, the models generated by GP had a good estimation performance.

2.2.5. Strongly Typed Genetic Programming

For the feature of being a powerful methodology for the automatic generation of computer programs, the genetic programming is implemented in a wide range of application fields (Koza, 1992). The formation of the individuals from former population to new one is carried out with the traditional selection and recombination operators of genetic algorithm (GA) (Haynes *et al.* 1995).

For various set of problems, the search space for an implementation of genetic programming composes of the unlimited set of possible compositions of the problem's functions and terminals (Koza, 1992). Restriction on the terminals and non-terminals defined by the user is called "closure", which is a term used to describe the obligation of non-terminals accepting arguments of a single data type and returning values of the same data type. This means that all non-terminals return values that can be used as arguments for any other non-terminal. The closure is claimed to have a great effect on the limitation of genetic programming by Montana (Montana, 1994).

Strongly typed genetic programming (so called constrained syntactic structure by Koza) is a composition of rules that defines the functions or terminals that are permitted to appear as a specified function in the program tree. Multiple function sets and multiple terminal sets exist in the algorithm typically whenever a constrained syntactic structure is utilized. The utilization locations of function sets or terminal sets are described with the rules of construction (Koza, 1992).

All of the members in the initial population are generated appropriate to the constrained syntactic structure at the start of strongly typed genetic programming. In addition to this, all of the genetic operations implemented during the run are configured in order to yield offspring suitable to the requirements of the constrained syntactic structure. As a result, all individuals existing during the run will have the desired grammatical structure (Koza, 1992). In this work, all arrangements were done targeting to match the models of surface reaction controlled systems at least, which constitute about the 75 % of the catalytic reactions (Fogler, 1992).

2.2.6. GPLAB

GPLAB is a versatile, generalist and easily extendible genetic toolbox for MATLAB that generates a program giving the best generated model with the contribution of the provided data. In addition to provided data, modeling parameters play important role in the models.

2.2.6.1. Operational Structure: The operational structure of GPLAB consists of three main modules, named as GENPOP, GENERATION and SET VARS. SET VARS either initializes the parameters with their default values or updates them with the options set by the user. At the start of the execution of the program SET VARS is called by GENPOP for the parameter initialization. GENPOP generates the initial population with one of the defined initialization methods, calculates its fitness and finishes by passing the implementation to GENERATION. GENERATION generates a new group of individuals in the result of applying the genetic operators to the previous population. The individuals are chosen from a pool of parents that is formed by the sampling method that the algorithm is determined to use. As long as the population is not filled, the genetic operators produce new individuals. Subsequent to the calculation of fitness, the new individuals entering the new generation are selected appropriate to the elitism level parameter in the survival module. The implementation of the GENERATION module continues until the stop condition is fulfilled or when the maximum generation number is reached.

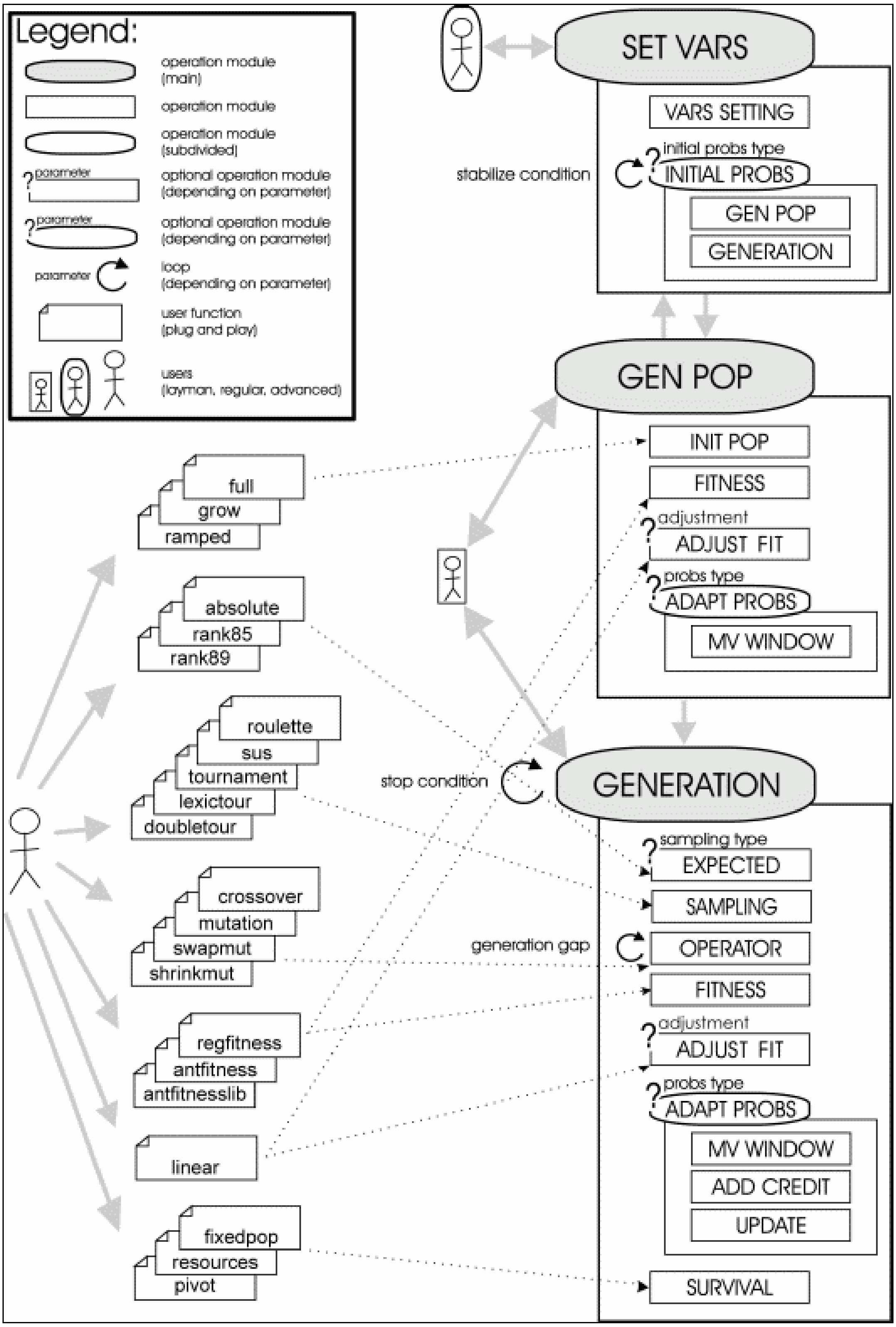

Figure 2.1. Operational structure of GPLAB (Silva, 2007)

The important parameters effective in the outputs of GPLAB are as follows:

2.2.6.2. Tree Initialization: The generation of the initial population in the beginning of GPLAB implementation is performed by the random selection of functions and terminals from respective predefined sets. Three types of initial population creation method exist in GPLAB. First of them is called ‘fullinit’ that utilizes full method in which new tree receives non terminal nodes until the initial maximum tree depth is reached. The second method is called ‘growinit’ utilizing grow method. Each new node is chosen in a random manner between terminals and non terminals, apart from nodes at the initial tree depth level, which obliged to be terminals. The last method is called ‘rampedinit’ which utilizes ramped half and half method. For each depth level considered between 2 and initial maximum tree depth level, half of the individuals are generated by the help of the full method and the other half using the grow method (Silva, 2007).

2.2.6.3. Functions and Terminals: Functions and terminals are required in GPLAB like any other genetic programming algorithm for the population creation. The available toolbox functions used in plug and play fashion are listed and concerning information is also given in Table 2.1. Functions that verify closure in addition to some protected and logical functions and the if-then-else statement can be utilized by the toolbox.

Table 2.1. Protected and logical functions for use with GPLAB

Protected Function	MATLAB Function	Input Arguments	Output Arguments
Division	mydivide	A,b	A (if b=0) a/b (otherwise)
Square root	mysqrt	a	0 (if a<=0) Sqrt(a) (otherwise)
Power	mypower	A,b	a^b (if a^b is a valid non-complex number) 0 (otherwise)
Natural Logarithm	mylog	a	0 (if a=0) log(abs(a)) (otherwise)
Base 2 Logarithm	mylog2	a	0 (if a=0) log2(abs(a)) (otherwise)
Base 10 Logarithm	mylog10	a	0 (if a=0) log10(abs(a)) (otherwise)
If-then-else Statement	myif	a,b,c	eval(c) (if eval (a)=0) eval (b) otherwise
Negation of AND	Nand	A,b	not(and(a,b))

Numerical constants which are influential in generating models determined as values drawn from the standard uniform distribution on the open interval (0, 1) by a random number generator. In the beginning of the run, the variables are generated automatically that are required for the evaluation of the fitness cases.

2.2.6.4. Selection for Reproduction: Parent individuals are needed to produce their children by genetic operators are selected according to one of the four sampling methods.

In the method called ‘roulette’, each individual owns a portion corresponding to their expected number of children and roulette with random pointers is spun. The second method called as ‘sus’ also relies on roulette in spite of its pointers being placed with equal portions.

The method of ‘tournament’ selects parents by random drawing a couple of individuals from the population following with the selection of the best member of the population.

The last method is called ‘lexictour’ selection and utilizes parsimony pressure. The main difference between ‘tournament’ and ‘lexictour’ selection is the secondary objective defined for the tree size in the ‘lexictour’ selection. If two of the chosen individuals at random have the same raw fitness values, then sizes are compared and the individual with the smaller size is selected (Luke and Panait, 2002).

2.2.6.5. Expected Number of Children: Depending on the sampling procedures, the parents are selected based on their estimated number of children or according to the comparison between each other due to their fitness. With the help of the selected method, the calculation of the expected number of children is performed for every single individual in the population if it is needed by the selection of the reproduction. There are three available methods in GPLAB: the expected number of children for each individual is proportional to its absolute fitness value in the method named as ‘Absolute’. In the second method called ‘Rank85’, the expected number of children for each individual is based on its rank in the population. In the last method called ‘Rank89’, the expected number of children for each individual is based on its rank in the population and how far it is from the maximum allowed generation.

2.2.6.6. Survival: Subsequent to generation of the new individuals for the new population, GPLAB start the implementation of the survival module where, a number of individuals are selected to form the new population from the compilation of the new children and the members of the previous population. There are four levels of elitism levels eligible for the use in GPLAB.

'replace' is not an elitist option for survival where the children replace the parent population completely.

Without regarded to being a child or parent, the best individual from both parents and children is taken for the new population in 'keepbest' method. The remainder of the population is filled with the child individuals only.

In 'halfelitism' method, half of the new population is forced to consist of the best individuals selected from a group comprising both parents and children. The remaining half will be formed by the best children.

The last survival method is 'totalelitism' that selects best individuals from both children and parents to form the new population (Silva, 2007).

2.2.6.7. Stop Conditions: The last profile setting in GPLAB indicating that evaluation of modeling came to an end is the termination criteria or stop conditions in other words. As far as reaching a stop condition or maximum generation criteria, GPLAB doesn't stop running. Termination criteria are determined by setting the parameter hits. One hit is a tuple [f d] where f is the percentage of fitness cases that must obey the stop condition and d is the definition of the stop condition itself that defines limitations for the obtained best individual in the range of minus and plus d% of the expected result. The default value of hits is [100 0] meaning that the program stops if the best individual produces exact output values in all fitness cases (Silva, 2007).

3. COMPUTATIONAL DETAILS

3.1. Conceptual Model

For the purpose of finding out the rate law equation of a reaction, many possible assumptions of reaction mechanisms and the rate determining steps have to be made in the Langmuir-Hinshelwood method, which are followed by the regression analysis to find the rate law equation constants from the experimental data. The utilization of this method have the drawback that the procedure should be repeated entirely if any change in the assumptions or in the elementary reaction of the mechanisms is made. If there is no prior knowledge about the mechanism, this approach is time consuming and there is no guarantee that it will be successful.

The genetic programming can be helpful to overcome this draw back by predicting the approximate rate law from the experimental data without making any assumption about the reaction mechanism and the rate limiting steps. Then the rate law obtained this way can be used to predict the mechanism since the structure of the equation (i.e. term in the numerator and denominator) will point out the significant steps in the mechanism. Even if the exact form of the rate law not be obtained, the approximate form of the equation will also help to start with a good guess of the mechanism for the Langmuir-Hinshelwood type analysis.

However genetic programming may also generate the equations that perfectly fit the experimental data but does not have any meaning as a rate law. Fortunately this potential problem can be prevented by using “strongly typed generic programming” which can restrict the form of the rate law to the common structure consisting the kinetic terms and driving force in numerator and the adsorption term in the denominator. This approach was fallowed in this study and numerator and the denominator of the rate law were optimized separately using the different sets of functions. In order to overcome the drawbacks of classical approach in determination of reaction rate law equations and rate law mechanisms, the strongly typed genetic programming was used to obtain the model equations directly from the experimental data. In this manner, contrary to the classical approach, the reaction mechanisms were deduced through kinetic parameters and the structure of the obtained rate law equations.

3.2. STGP Program Details

In general, GP scans a very large search space with the contribution of defined functions and terminals. The results expected to be achieved consist of the division of two expressions in the following form.

$$\text{Rate} = \frac{(\text{Kinetic Term})\,(\text{Driving Force})}{(\text{Adsorption Term})} \tag{3.1}$$

For the purpose of attaining desired results with ease, type of a modularity proposed in this work was provided which would shrink the search space.

The expressions, attempted to evaluate, can contain multiplication and square root functions (sometimes subtraction as part of the drawing force) in the numerator involving some constant (reaction and equilibrium constants) with the reactant concentrations while denominator can include functions of addition, square root, square and multiplication of the same variables and constants (there is also addition of 1 n most cases). A simple form of the rate law equation is illustrated in Equation 3.2.

$$r_a = \frac{k_1 P_A}{1 + K_1 P_B} \tag{3.2}$$

The individuals of initial population are created in desired structure as the initial step of modularization of the algorithm. The top node of every single tree structured individual is chosen as the division function. Sub trees at the left of the top nodes contain the functions that can be in the numerator of the expressions while sub trees remaining at the right side of the top nodes include the functions that can exist in the denominator. The schematic representation of the approach is given in Figure 3.1.

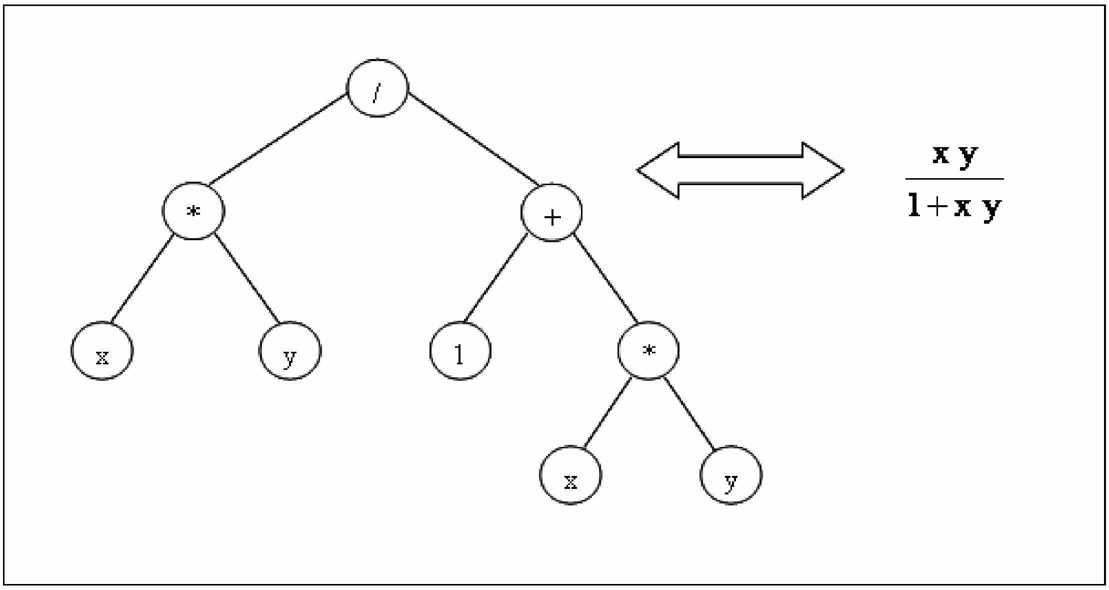

Figure 3.1. Tree structured representation of a solution candidate

The second step of the modularization is the modified implementation of crossover and mutation operators on the individuals in the population. First of all, a node from the first of two parent individuals is chosen randomly except the top node. Whether the chosen node is on the left or right side of the first node, the node from the second parent is chosen from the same side of its own top node as it is in the first parent individual. Same as the crossover operator, a node is selected randomly while the mutation operator is implemented on the individuals. Due to the location of the node whether it is on the left or right side of the top node, the function set of the numerator or the denominator is used in the generated new sub tree that is attached to the main tree under the chosen node. By the help of those adjustments in the genetic operators, the utilization of functions out of the function sets defined for the numerators and denominators is prevented. The implementation of crossover operator can be seen in Figure 3.2. for the denominator of the rate law expressions.

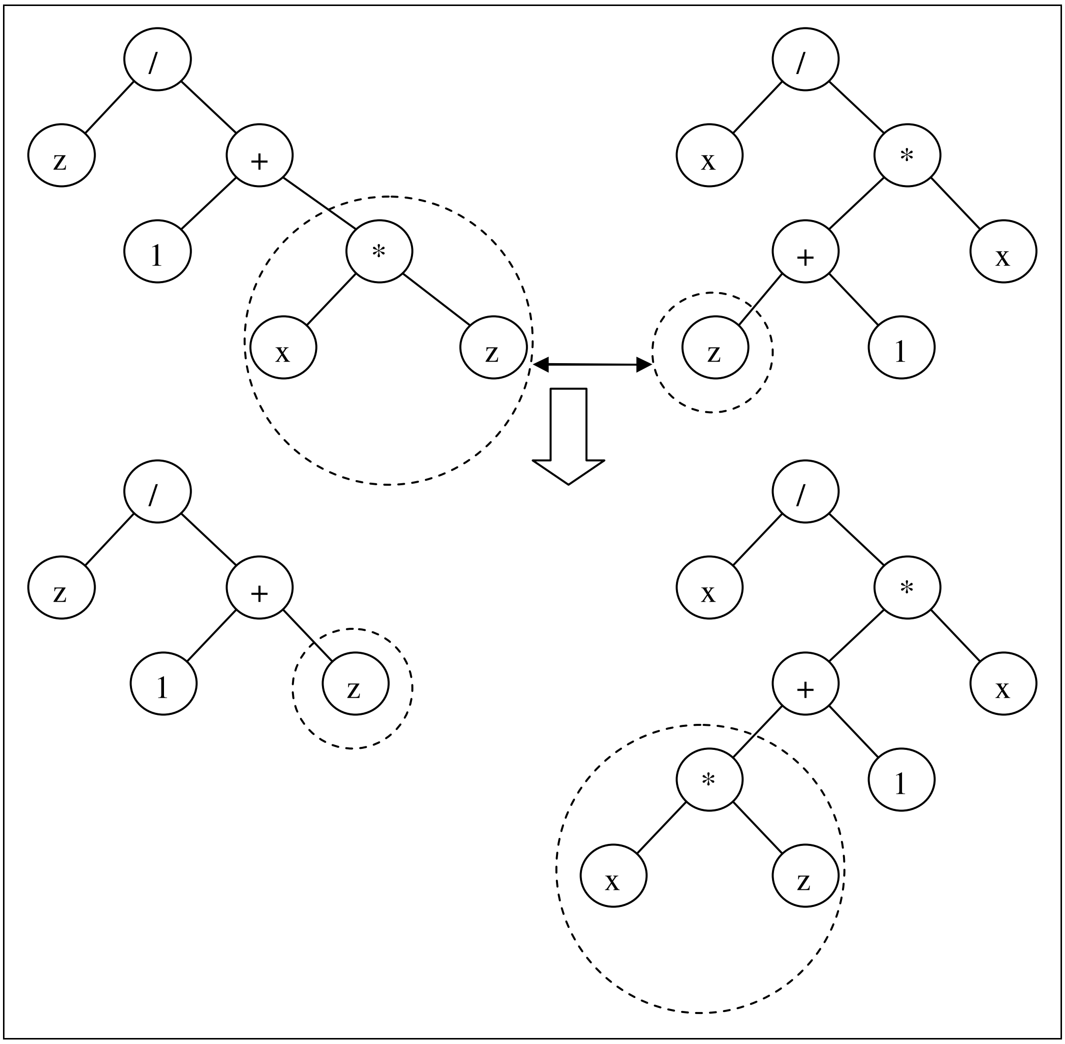

Figure 3.2. The implementation of crossover operator on denominators of expressions

Genetic programming makes possible the automatic evaluation of computer programs by taking advantage of principles from biological systems. For the aim of obtaining an appropriate model describing the main characteristics of the rate equation by genetic programming, GPLAB (Version 3) was used. In this section, the concept of GPLAB and arrangements were done for modeling of the reactions given in this work was examined in detail.

GPLAB is a versatile, generalist and easily extendible genetic toolbox for MATLAB that generates a program giving the best generated model with the contribution of the provided data. In addition to provided data, modeling parameters play important role in the models.

Numerical constants which are influential in generating models determined as values drawn from the standard uniform distribution on the open interval (0, 1) by a random number generator. Furthermore, ‘regfitness’ was always preferred as the fitness function being the sum of absolute differences between the measured output value and the value returned by the individual on all fitness cases.

Due to the nature of genetic programming that generates a model without a guarantee to be the shortest possible solution, the program size reduction plays an important role in the algorithm. As a matter of fact, extremely large programs can be obtained and bloat control problem can occur. For the intention of having a better understanding of the reaction mechanism and avoiding complexity, size limitation of the results can be helpful. Thus, lexicographic parsimony pressure which is a multi objective method for optimizing both tree size and fitness is utilized. The main difference between ‘tournament’ and ‘lexictour’ selection is the secondary objective defined for the tree size in the ‘lexictour’ selection. If two of the chosen individuals at random have the same raw fitness values, then sizes are compared and the individual with the smaller size is selected.

In addition to lexicographic parsimony pressure, dynamic maximum tree depth which presents a dynamic limit to the maximum depth is used to deal with the increase in the size of the tree structured individuals. In this method, dynamic and strict limits are implemented in collaboration that dynamic limit is initially set as a low value that is not higher than the maximum depth of the initial random trees. Subsequent to the application of the genetic operators, any new individuals not suitable for the limit is declined and substituted with one of its parents if it is not the individual with the best fitness. On other occasions the new individual, whose depth is matched by the raise of the dynamic limit, is accepted into the population. The dynamic limit is never lowered, and never exceeds the static limit in any condition. The initial dynamic level was set as 6 while the strict level was chosen as 15.

Initial population settings play an important role in the models. The high variety in the initial population is always beneficial in order to obtain convincing results. Three types of initial population creation method exist in GPLAB. First of them is called ‘fullinit’ that utilizes full method in which new tree receives non terminal nodes until the initial maximum tree depth is reached. The second method is called ‘growinit’ utilizing grow method. Each new node is chosen in a random manner between terminals and non terminals, apart from nodes at the initial tree depth level, which obliged to be terminals. The parameter for the initial population type was selected as the last method that is called ‘rampedinit’ which utilizes ramped half and half method. For each depth level considered between 2 and initial maximum tree depth level, half of the individuals are generated by the help of the full method and the other using the grow method.

Depending on the sampling procedures, parents are selected based on their estimated number of children or according to the comparison between each other due to their fitness. With the help of the selected method, the calculation of the expected number of children is performed for every single individual in the population if it is needed by the selection of the reproduction. In this work the method called ‘rank85’ in which the procedure for expected number of children is implemented via rank of the individuals in the population.

The probability of genetic operators is one of the key parameters that should be set by the user. The decisions made on those control parameters have crucial importance that affect the evaluation speed of the population and the introduction of new sub trees that contribute on diversity. As a result of adjusting probabilities, proportions of population close to probabilities are forced to undergo the specified operators. In this work, probability of reproduction is set to 5% while that of crossover and mutation is set to 65% and 35% of the remainder of probability in order to maintain the variety in the population.

Subsequent to generation of the new individuals for the new population, GPLAB start the implementation of the survival module where, a number of individuals are selected to form the new population from the compilation of the new children and the members of the previous population. An elitism level named as replace is used in our work in which the individuals created by the genetic operators replace the parent population completely, even if they are worse than their parents.

The last profile setting in GPLAB indicating that evaluation of modeling came to an end is the termination criteria or stop conditions in other words. As far as reaching a stop condition or maximum generation criteria, GPLAB doesn't stop running. Termination criteria are determined by setting the parameter hits. One hit is a tuple [f d] where f is the percentage of fitness cases that must obey the stop condition and d is the definition of the stop condition itself that defines limitations for the obtained best individual in the range of minus and plus d% of the expected result. The default value of hits which was also used in our work is [100 0] meaning that the program stops if the best individual produces exact output values in all fitness cases. The parameters and setting used in STGP is given in Table 3.1.

Table 3.1. Parameters used in STGP

Population Size	1000
Maximum Number of Generation	50
Initial Population Type	rampedinit
Selection of Reproduction	lexictour
Survival	replace
Probability of Reproduction	0.05
Probability of Crossover	0.65
Probability of Mutation	0.35
Initial Dynamic Level	6
Strict Level	15
Expected Number of Children	Rank 85

The function selection plays an important role in generating solutions. The derivation of rate law expressions can be made more efficient if the expected nature of the model is known. However, without any knowledge or clue about the nature of the reaction, the reaction mechanisms are tried to be estimated by generated rate law equations obtained directly from experimental data with the contribution of strongly typed genetic programming. For this reason, a function set selection procedure is implemented for evaluating models using the data taken from literature and the experimental data from our laboratory. This procedure consists of two steps:

- The first trials are done with a function set comprising addition and multiplication functions. Addition is used only in the denominator of the expressions while multiplication is used in both numerator and denominator.

- If there is a suspect of dissociation in the reaction due to theoretical knowledge or the result of the trial with the first function set is not convincing, then square roots of the variables are added to the terminal set as additional terminals. The function set remains the same as it was in the first trial.

The success of the second trial can be quantified by comparing the resulting model with the model derived by the first trial. The evaluated model by the second trial can be concluded to be the rate law equation if its R^2 value is better than the first one and the achieved model is acceptable by looking its functional structure. Nothing can be done special for dual site reaction mechanisms apart from the settings in the first trial. The rate law equations with the square of whole denominator are hard task for the algorithm proposed in this work because of the number of parameters in the denominator in the open expression. The increase of terms in the denominator makes the evaluation of models complicated in terms of structural convenience. A dual site reaction mechanism with surface reaction being the rate determining step can be suspected while there are terms with squares of variables multiplied by no negligible constants.

4. RESULTS AND DISCUSSION

In this work, the kinetic modeling reactions, of which the experimental rate data were reported in the literature or obtained in our laboratory (for CO oxidation) was performed using strongly typed genetic programming and confronted with the rate laws proposed. Before the utilization of the real data for the real reactions, some simple generic reactions obeying the well established reaction rate mechanisms and rate laws were used in order to have a basic understanding in derivation of rate laws by our modified GPLAB. The pressure versus rate data for these reactions generated through arbitrarily set rate laws with random error were used in the generic reactions part as training data. Then the method was tested for the real reactions with the real experimental data and proposed rate law, which can be used for comparison. The rate equations were derived directly from the data without making any mechanism assumption. On the contrary, the reaction mechanisms and the appropriate assumptions were tried to be predicted from the rate law obtained. During the evaluation of rate laws, the function sets were changed if it was necessary.

4.1. Generic Reactions

GP is able to derive appropriate models when discriminates the main functional groups that compose the main features of the rate equation. In order to have a better understanding of GP in reaction kinetics, various generic reactions were labored for determining weaknesses of our algorithm.

The first mechanism that was tried was a unimolecular isomerization reaction. The individual steps of the reaction are given in the equations 4.1, 4.2. and 4.3. Firstly, the reactant is adsorbed on an active site and the surface reaction occurs. Afterwards, the product of the reaction desorbs.

<u>Mechanism 1</u>

$$\mathrm{A + S \leftrightarrow A.S} \tag{4.1}$$

$$\mathrm{A.S \leftrightarrow B.S} \tag{4.2}$$

$$\mathrm{B.S \leftrightarrow B + S} \tag{4.3}$$

The corresponding rate laws with the rate determining steps and the models are given in Table 4.1.

Table 4.1. List of models tried for Mechanism 1

Model No	Rate Determining Step	Rate Laws
1.1	Desorption	$r=\dfrac{\frac{k_1 k_3 k_5}{k_2 k_4} P_A S_o}{1+\frac{k_1}{k_2}\left(1+\frac{k_3}{k_4}\right)P_A}$
1.2	Adsorption	$r=\dfrac{k_1 S_o P_A}{1+\left(\frac{k_6}{k_5}\right)\left(1+\frac{k_4}{k_3}\right)P_B}$
1.3	Surface Reaction	$r=\dfrac{\left(\frac{k_1 k_3}{k_2}\right)P_A S_o}{1+P_A\left(\frac{k_1}{k_2}\right)+P_B\left(\frac{k_6}{k_5}\right)}$
1.4*	Surface Reaction	$r=\dfrac{\left(\frac{k_1 k_3}{k_2}\right)P_A S_o}{1+P_A\left(\frac{k_1}{k_2}\right)+P_B\left(\frac{k_6}{k_5}\right)}$
1.5	Adsorption	$r=\dfrac{k_1 S_o P_A - \left(\frac{k_6 k_4 k_2}{k_5 k_3}\right)P_B S_o}{1+\left(\frac{k_6}{k_5}\right)\left(1+\frac{k_4}{k_3}\right)P_B}$
1.6	Surface reaction	$r=\dfrac{\left(\frac{k_1 k_3}{k_2}\right)P_A S_o - \left(\frac{k_6 k_4}{k_5}\right)P_B S_o}{1+P_A\left(\frac{k_1}{k_2}\right)+P_B\left(\frac{k_6}{k_5}\right)}$
1.7	Desorption	$r=\dfrac{\frac{k_1 k_3 k_5}{k_2 k_4} P_A S_o - k_6 P_B S_o}{1+\frac{k_1}{k_2}\left(1+\frac{k_3}{k_4}\right)P_A}$

*The same structure of model 1.3 was used with different numerical values to obtain two different model in Table 4.2.

These models were used to generate rate data by assuming some numerical values of the rate constants and then calculating the reaction rates corresponding to the assumed reactant partial pressures in a reasonable range. Some random errors representing the experimental error were also added to the rates. The reaction rate of models 1.1, 1.2, 1.3 and 1.4 were assumed to be examined at the beginning of the corresponding reactions. Therefore, the predict models of 1.1, 1.2, 1.3 and 1.4 do not contain terms with minus in numerator. Then the rate versus partial pressures data were used to predict the form and the constants of the rate law by utilizing the strongly typed genetic programming with the details and setting described in Chapter 3. In addition to the function set described in Chapter 3, the minus function was also used to predict the models 1.5, 1.6 and 1.7. The rate laws with the assumed rate constants and predicted by the STGP were given in Table 4.2.

Table 4.2. Comparison of created and STGP predicted model in Mechanism 1

Model No	Model created with assumed rate constants	Model predicted by STGP
1.1	$r = \frac{0.01\,P_A}{1 + 1.5\,P_A}$	$r = \frac{0.009978\,P_A}{1 + 1.5\,P_A}$
1.2	$r = \frac{0.2\,P_A}{1 + 1.2\,P_B}$	$r = \frac{0.197\,P_A}{1 + 1.14\,P_B}$
1.3	$r = \frac{0.1\,P_A}{1 + 1.5\,P_A + 0.2\,P_B}$	$r = \frac{0.094\,P_A}{1 + 1.4822\,P_A}$
1.4	$r = \frac{0.02\,P_A}{1 + 1.6\,P_A + 0.8\,P_B}$	$r = \frac{0.0211\,P_A}{1 + 1.734\,P_A + 0.95\,P_B}$
1.5	$r = \frac{0.01\,P_A - 0.0019\,P_B}{1 + 1.71\,P_B}$	No meaningful prediction
1.6	$r = \frac{0.125\,P_A - 0.00945\,P_B}{1 + 1.25\,P_A + 0.945\,P_B}$	No meaningful prediction
1.7	$r = \frac{0.126\,P_A - 0.0025\,P_B}{1 + 2.463\,P_A}$	$r = \frac{0.104\,P_A}{1 + 1.46\,P_A + 0.151\,P_B + 0.167\,P_A^2\,P_B + 0.189\,P_A^3\,P_B}$

The model 1.1, 1.2 and 1.3 were come naturally in the first trial of the program showing a perfect match with the rate law used to generate the data allowing the prediction of the mechanism easily since the mechanisms and the assumptions resulting these rate expressions are readily available in the kinetic literature. The model 1.3 on the other hand is not recovered completely apparently due to the relatively small value of (0.2 P_B) and it was indistinguishable from model 1.1. Model 1.5 and 1.6 were totally unacceptable due to their structural inappropriateness. Model 1.7 is not also completely estimated due to the small value of (0.0025 P_B). Also excess but negligible terms of (0.151 P_A), (0.167 P_A^2 P_B), and (0.189 P_A^3 P_B) were included in the denominator of the expression. The predicted versus generated rate values of four model are given in Figure 4.1 with the high R^2 values indicating that our approach is working for the models in the estimation of which minus operator was not used.

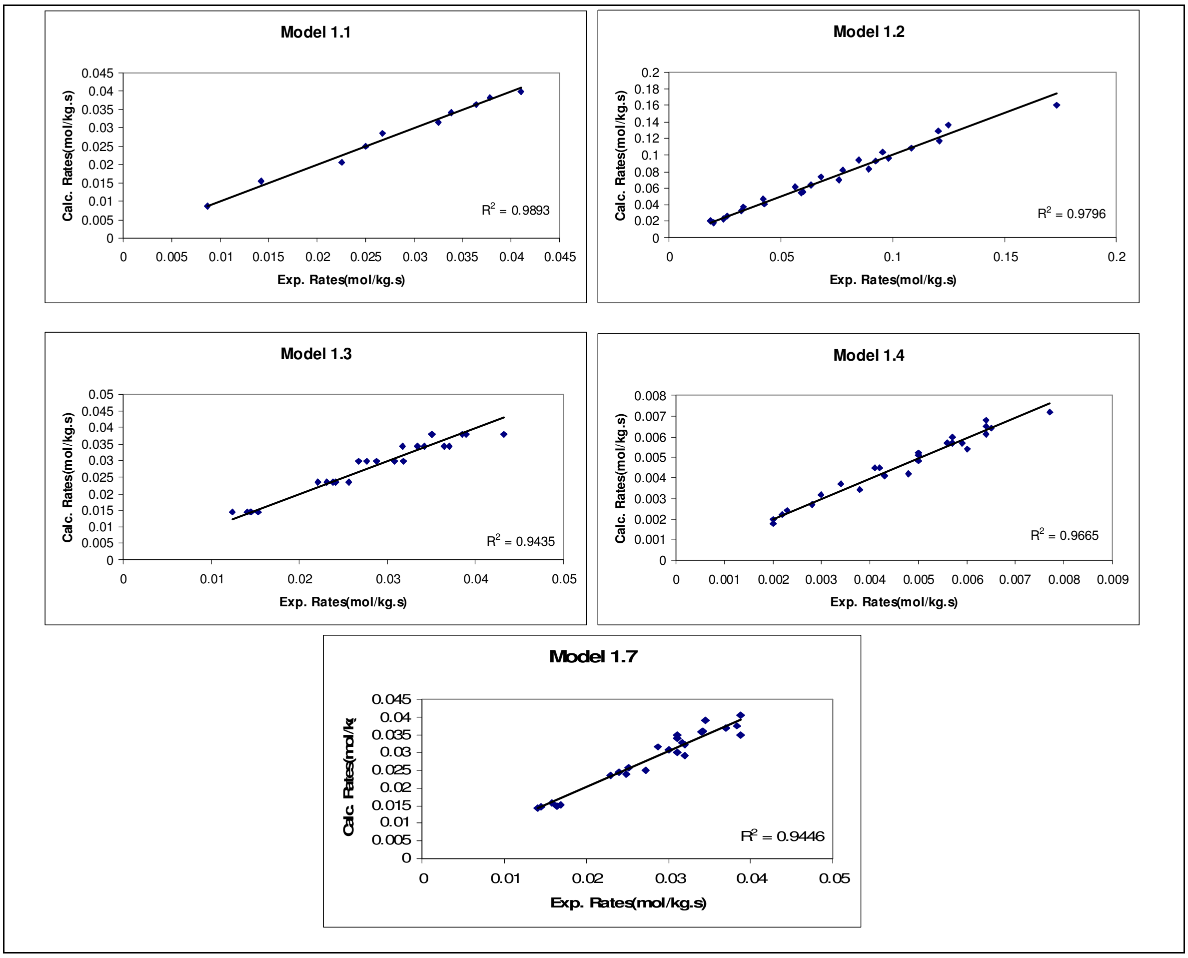

Figure 4.1. Predicted vs. generated rates for Mechanism 1

The second mechanism for trial was one of the surface-catalyzed bimolecular reactions in which, both species are preadsorbed and the reaction is carried out between adsorbed species A and B at the adjacent sites on the surface. The mechanism is written as follows:

Mechanism 2

$$A + S \leftrightarrow A.S \tag{4.4}$$

$$B + S \leftrightarrow B.S \tag{4.5}$$

$$A.S + B.S \leftrightarrow C.S + S \tag{4.6}$$

$$C.S \leftrightarrow C + S \tag{4.7}$$

The rate laws with the rate determining steps and proposed models for this mechanism are given in Table 4.3.

Table 4.3. List of models tried for Mechanism 2

Model No	Rate Determining Step	Rate Laws
2.1	Desorption	$r = \dfrac{\dfrac{k_2 k_4 k_5}{k_1 k_3} P_A P_B S_o}{1 + \dfrac{k_2}{k_1} P_A + \dfrac{k_4}{k_3} P_B + \dfrac{k_2 k_4 k_5}{k_1 k_3 k_6} P_A P_B}$
2.2	Surface reaction	$r = \dfrac{\dfrac{k_2 k_4 k_5 k_7}{k_1 k_3 k_6} P_A P_B S_o}{\left(1 + \dfrac{k_2}{k_1} P_A + \dfrac{k_4}{k_3} P_B + \dfrac{k_7}{k_8} P_C\right)^2}$
2.3	Adsorption	$r = \dfrac{k_1 P_A S_o - \dfrac{k_2 k_3 k_6 k_7}{k_4 k_5 k_8} \dfrac{P_C}{P_B} S_o}{1 + \dfrac{k_4}{k_3} P_B + \dfrac{k_7}{k_8} P_C + \dfrac{k_3 k_6 k_7 P_C}{k_4 k_5 k_8 P_B}}$
2.4	Desorption	$r = \dfrac{\dfrac{k_2 k_4 k_5}{k_1 k_3} P_A P_B S_o - k_8 P_C S_o}{1 + \dfrac{k_2}{k_1} P_A + \dfrac{k_4}{k_3} P_B + \dfrac{k_2 k_4 k_5}{k_1 k_3 k_6} P_A P_B}$

Similar to the first mechanism, some appropriate rate constants were assumed and the partial pressures versus rate data were generated for STGP. Models 2.1 and 2.2 were assumed to be examined at the beginning of the corresponding reactions and were generated with the function set that was consisting of addition which could be utilized only in denominator of the expressions and multiplication that could be used in both numerator and denominator. Minus function was also used in models 2.3 and 2.4 for the numerator of the expressions while division function was used only for model 2.3 because of the necessity. The results are compared in Table 4.4.

Table 4.4. Comparison of created and STGP predicted model in Mechanism 2

Model No	Model created with assumed rate constants	Model predicted by STGP
2.1	$r = \dfrac{0.4\,P_A\,P_B}{1 + 1.2\,P_A + 0.8\,P_B + 0.4\,P_A\,P_B}$	$r = \dfrac{0.3761\,P_A\,P_B}{1 + 0.999\,P_A + 0.999\,P_B}$
2.2	$r = \dfrac{0.01\,P_A\,P_B}{(1 + 1.6\,P_A + 0.9\,P_B + 0.6\,P_C)^2}$	$r = \dfrac{0.0024\,P_A\,P_B}{1.62\,P_A + P_B + 0.276\,P_C^2 + P_A\,P_B}$
2.3	$r = \dfrac{0.1\,P_A - 0.00124\dfrac{P_C}{P_B}}{1 + 0.786\,P_B + 1.033\,P_C + 1.24\dfrac{P_C}{P_B}}$	No meaningful prediction
2.4	$r = \dfrac{0.07\,P_A\,P_B - 0.001\,P_C}{1 + 0.637\,P_A + 0.92\,P_B + 0.7\,P_A\,P_B}$	$r = \dfrac{0.142\,P_A\,P_B}{1 + 2.6\,P_B + 2.78\,P_A\,P_B + \dfrac{0.208}{P_B}}$

The predicted model 2.1 is close to the model tested except that the term (0.4 P_AP_B) seem to be neglected by the STGP due to its small values compare to the other terms. The predicted rate model 2.2 is quite different from the its corresponding model created with the assumed constants by the structure of its denominator. Even the denominator of the desired model is expanded there is a few similarity with the predicted model.structurally, model 2.3 is not acceptable. In the estimation of model 2.4, except for the neglection of the term (0.637 P_A) by the program itself, a resemblance can be seen between the desired and predicted models that is insufficient. The generated versus predicted reaction rate was again in close agreement as presented in Figure 4.2 although there are some weakness in predicting the shape of the rate law.

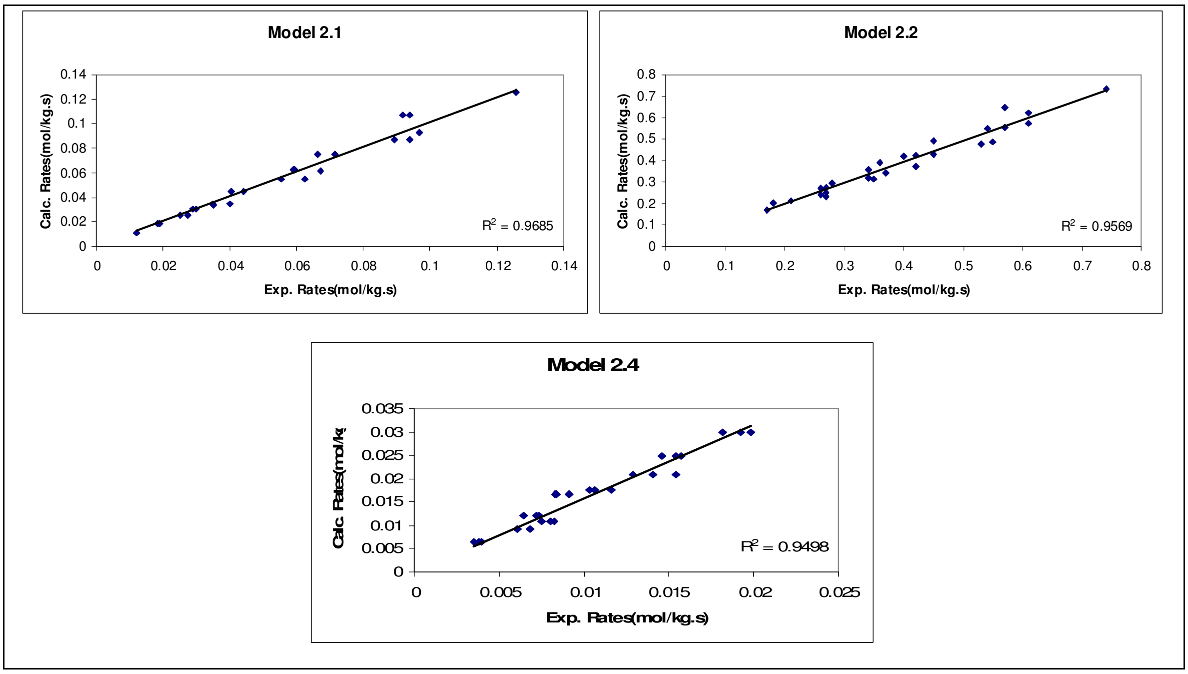

Figure 4.2. Predicted vs. generated rates for Mechanism 2

The third mechanism that is worked on was a decomposition reaction. A preadsorbed specie gets into the reaction with an adjacent site and two product species are produced on both sites as a result.

Mechanism 3

$$A + S \leftrightarrow A.S \tag{4.8}$$

$$A.S + S \leftrightarrow C.S + D.S \tag{4.9}$$

$$C.S \leftrightarrow C + S \tag{4.10}$$

$$D.S \leftrightarrow D + S \tag{4.11}$$

The related rate expression and the destined model are as follows with the rate determining step are shown in Table 4.5.

Table 4.5. List of models tried for Mechanism 3

Model No	Rate Determining Step	Rate Law
3.1.	Adsorption	$r = \dfrac{k_1 P_A S_0}{1 + \left(\dfrac{k_4 k_8}{k_3 k_7}\right) P_C P_D + \dfrac{k_6}{k_5} P_C + \dfrac{k_8}{k_7} P_D}$

The rate expression of reaction that is assumed to be examined at the beginning of the corresponding reaction generated by assigning numerical values for rate constants and the corresponding expression predicted by the STGP is presented in Table 4.6.

Table 4.6. Comparison of created and STGP predicted model in Mechanism 3

Model No	Model created with assumed rate constants	Model predicted by STGP
3.1	$r = \frac{0.02 P_A}{1 + 1.5 P_D P_C + P_D + 0.7 P_C}$	$r = \frac{0.012 P_A}{1 + 0.905 P_D P_C + 0.472 P_D + 0.412 P_C}$

Although there is some similarity between two expressions, the match is not sufficient this time probably because of too many parameters in the denominator that more than one combination of these parameters can produce the same numerical value for the reaction rates. This is also evident from the good agreement between the constructed versus predicted reaction rates in Figure 4.3. for this mechanism.

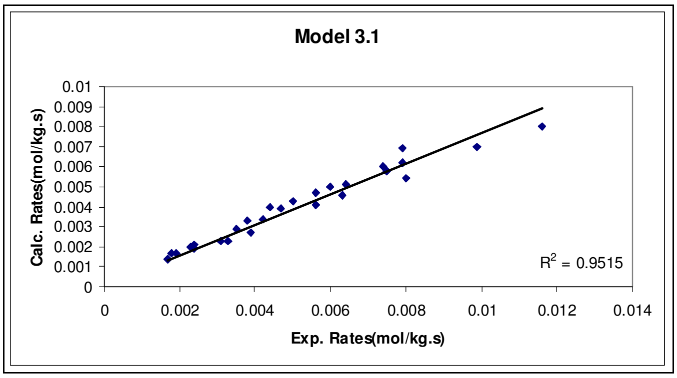

Figure 4.3. Predicted vs. generated rates for Mechanism 3

The fourth mechanism chosen was bimolecular reaction with dissociation of one reactant. Firstly, the reactant is adsorbed dissociatively and then the surface reaction is occurred between the adsorbed specie and the reactant in the gas phase.

Mechanism 4

$$A_2 + 2S \leftrightarrow 2A.S \tag{4.12}$$

$$A.S + B \leftrightarrow C + S \tag{4.13}$$

Consequently, the rate law for this mechanism above is given in Table 4.7 if the surface reaction was the rate controlling step. The reaction is assumed to be observed at the beginning.

Table 4.7. List of models tried for Mechanism 4

Model No	Rate Determining Step	Generated Rate
4.1	Surface Reaction	$r = \dfrac{k_1 P_A^{1/2} P_B S_0^{2}}{1 + \left(\dfrac{k_1}{k_2}\right)^{1/2} P_A^{1/2}}$

This model is the most difficult of the ones tested so far because it is hard for our algorithm to find out the terms with the square root of variables. The utilization of square root function causes structural inappropriateness that is not comprised by Langmuir-Hinshelwood and Rideal-Eley mechanisms. For that reason, square roots of variables were added to the terminal set of the program in order to overcome the mentioned drawback. Contrary to this solution, the increase in the variety of terms in the terminal set makes the convergence of the program more complicated.

Table 4.8. Comparison of created and STGP predicted model in Mechanism 4

Model No	Model created with assumed rate constants	Model predicted by STGP
4.1	$r = \dfrac{0.03 P_A^{1/2} P_B}{1 + 0.1 P_A^{1/2}}$	$r = \dfrac{0.0392 P_A^{1/2} P_B}{1 + 0.272 P_A^{1/2} + 0.19 (P_A P_B)^{1/2}}$

Again two expressions have some similarities suggesting that the modeling with STGP generates some meaningful results as also evident from Figure 4.4. However the additional term in the denominator creates some difficulties to decide about appropriate mechanism, which would be unknown in the real problems.

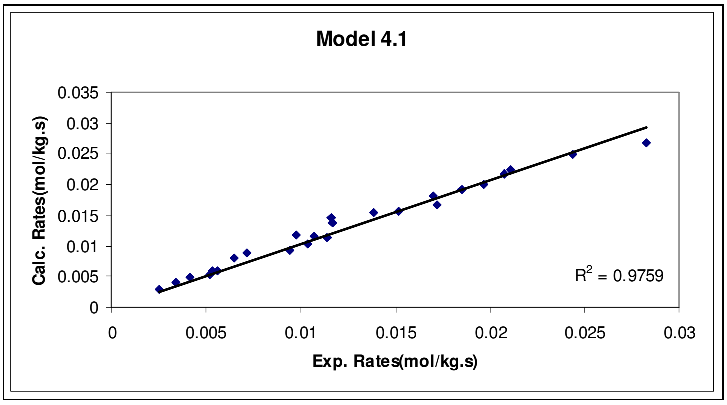

Figure 4.4. Predicted vs. generated rates for Mechanism 4

As being different from the previous mechanisms, the next observed reaction with the individual steps written in equations 4.14., 4.15., 4.16. and 4.17. were carried out with reactants preadsorbed on different sites.

Mechanism 5

$$A + S_1 \leftrightarrow A.S_1 \tag{4.14}$$

$$B + S_2 \leftrightarrow B.S_2 \tag{4.15}$$

$$A.S_1 + B.S_2 \leftrightarrow C.S_1 + S_2 \tag{4.16}$$

$$C.S_1 \leftrightarrow C + S_1 \tag{4.17}$$

The rate laws with the rate determining steps and proposed models for this mechanism are given in Table 4.9.

Table 4.9. List of models tried for Mechanism 5

Model No	Rate Determining Step	Generated Rate
5.1	Adsorption of A	$r = \dfrac{k_1 P_A S_1{}^0 - k_2 \dfrac{k_6 k_8}{k_7 k_5} P_C S_1{}^0 \left(1 + \dfrac{k_3}{k_4 P_B}\right)}{1 + \dfrac{k_8}{k_7} P_C \left[1 + \dfrac{k_8 k_6}{k_7 k_5}\left(1 + \dfrac{k_3}{k_4 P_B}\right)\right]}$
5.2	Adsorption of B	$r = \dfrac{k_3 P_B S_2{}^0 - \dfrac{k_2 k_4 k_6 k_8}{k_1 k_5 k_7} \dfrac{P_C}{P_A} S_2{}^0}{1 + \dfrac{k_6 k_8 k_2}{k_1 k_5 k_7} \dfrac{P_C}{P_A}}$
5.3	Desorption	$r = \dfrac{\dfrac{k_1 k_3 k_5 k_7}{k_2 k_4 k_6} P_A P_B S_1{}^0 - k_8 P_C S_1{}^0}{1 + \dfrac{k_1}{k_2} P_A + \dfrac{k_1 k_3 k_5}{k_2 k_4 k_6} P_A P_B}$

Models 5.1, 5.2 and 5.3 were tried to be predicted with the addition of minus function into the function set of the program for the numerator of the rate law expressions. Division function was also utilized as well for the estimation of models 5.1 and 5.2 which was necessary in this case.

Table 4.10. Comparison of created and STGP predicted model in Mechanism 5

Model No	Model created with assumed rate constants	Model predicted by STGP
5.1	$r = \dfrac{0.03 P_A - 0.00183 P_C \left(1 + \dfrac{1.12}{P_C}\right)}{1 + 0.957 P_C \left[1 + 0.913\left(1 + \dfrac{1.12}{P_B}\right)\right]}$	$r = \dfrac{0.0036 P_A}{\dfrac{P_C}{0.6 + P_B P_A + P_B P_C}}$
5.2	$r = \dfrac{0.02 P_B - 0.000823 \dfrac{P_C}{P_A}}{1 + 0.823 \dfrac{P_C}{P_A}}$	$r = \dfrac{0.0214 P_B}{1 + \dfrac{P_C}{P_A} + 0.086 \dfrac{P_C}{P_A{}^4 P_B}}$
5.3	$r = \dfrac{0.0326 P_A P_B - 0.002 P_C}{1 + 1.34 P_A + 1.632 P_A P_B}$	No meaningful prediction

Except for model 5.2, model 5.1 and 5.3 are totally unacceptable in terms of mathematical structure. In the evaluation of model 5.2, the program neglected the term (0.0000823 P_C / P_A) which has a small numerical value. However because of the additional term that STGP generated in denominator, the prediction of model 5.2 can not be verified to be successful. The related constructed versus predicted reaction rates are given in Figure 4.5. for this mechanism.

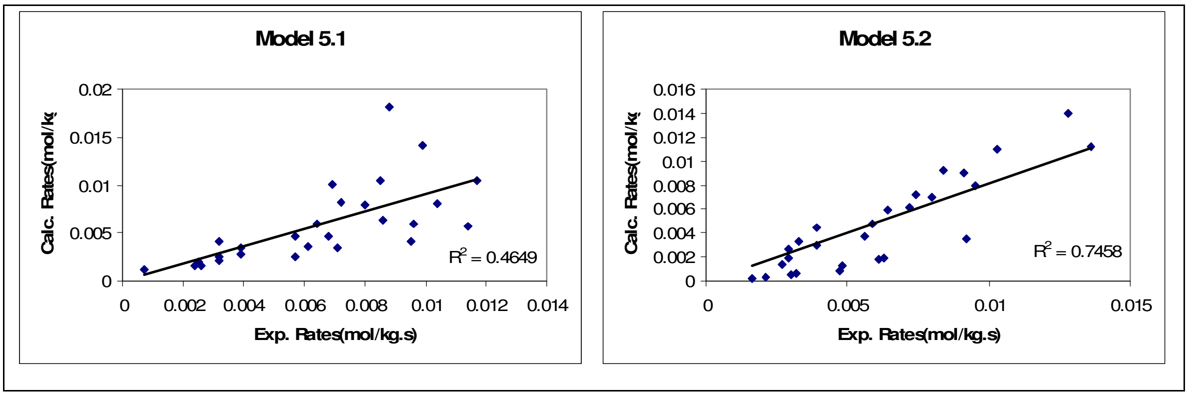

Figure 4.5. Predicted vs. generated rates for Mechanism 5

The subsequent mechanism from which, the rate law expressions are evaluated to test our algorithm is one of the Eley-Rideal mechanisms. The preadsorbed reactant is taken into the surface reaction with the second reactant in the gas phase. Finally the product desorbs from the active site of the catalyst.

Mechanism 6

$$A + S \leftrightarrow A.S \tag{4.18}$$

$$A.S + B \leftrightarrow C.S \tag{4.19}$$

$$C.S \leftrightarrow C + S \tag{4.20}$$

The related rate law expressions with rate determining steps and models that are focused on to be evaluated are given in Table 4.11.

Table 4.11. List of models tried for Mechanism 6

Model No	Rate Determining Step	Generated Rate
6.1	Adsorption	$r = \dfrac{S_o\, k_1\, P_A - \dfrac{k_2\, k_4\, k_6}{k_3\, k_5}\left(\dfrac{P_C}{P_A}\right) S_o}{1 + \dfrac{k_6}{k_5} P_C}$
6.2	Surface Reaction	$r = \dfrac{\dfrac{k_1\, k_3}{k_2} P_A\, P_B\, S_o - \dfrac{k_4\, k_6}{k_5} P_C\, S_o}{1 + \dfrac{k_1}{k_2} P_A + \dfrac{k_6}{k_5} P_C}$

For both of the models; a function set including minus for only numerator, addition for only denominator and multiplication for both numerator and denominator was used in the prediction. The estimated rate law equations for corresponding models are given in Table 4.12.

Table 4.12. Comparison of created and STGP predicted model in Mechanism 6

Model No	Model created with assumed rate constants	Model predicted by STGP
6.1	$r = \dfrac{0.1\, P_A - 0.0033 \dfrac{P_C}{P_A}}{1 + 1.158\, P_C}$	$r = \dfrac{0.1781\, P_A^{\,2}}{1.863\, P_A + P_B + P_A^{\,2}\, P_C}$
6.2	$r = \dfrac{0.96\, P_A\, P_B - 0.07\, P_C}{1 + 1.2\, P_A + 0.889\, P_C}$	$r = \dfrac{0.636\, P_A\, P_B}{1 + 0.354\, P_C + 0.547\, P_A + 0.151\, P_C^{\,2} + 0.28\, P_C^{\,5} + 0.262\, P_C^{\,6}}$

Even though the R^2 values of both models are acceptable, the predicted models do not resemble to the models created with assumed constants at all. The additional last three terms of denominator of the predicted model 6.2 can be assumed as negligible. The program also neglected to evaluate the term (0.07 P_C) that has a low numerical value. The related constructed versus predicted reaction rates are given in Figure 4.6. for mechanism 6.

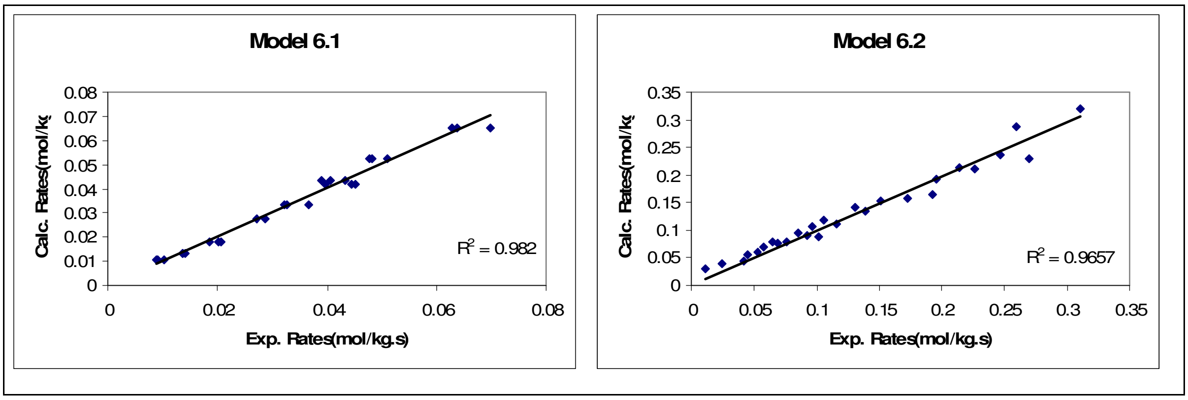

Figure 4.6. Predicted vs. generated rates for Mechanism 6

The last generic reaction mechanism that was worked on has the following individual mechanism steps. In the adsorption step, the product is attained on the active site and the product specie desorbs subsequently. The rate law equations with rate determining steps are shown in Table 4.13.

Mechanism 7

$$A + S \leftrightarrow B.S \tag{4.21}$$

$$B.S \leftrightarrow B + S \tag{4.22}$$

Table 4.13. List of models tried for Mechanism 7

Model No	Rate Determining Step	Generated Rate
7.1	Adsorption	$r = \dfrac{k_1 S_0 P_A - S_o \left(\dfrac{k_2 k_4}{k_3}\right) P_B}{1 + \dfrac{k_4}{k_3} P_B}$
7.2	Surface Reaction	$r = \dfrac{\dfrac{k_3 k_1}{k_2} P_A S_o - k_4 P_B S_o}{1 + \dfrac{k_1}{k_2} P_A}$

The function set used for mechanism 7 is the same as used for mechanism 6.

Table 4.14. Comparison of created and STGP predicted model in Mechanism 7

Model No	Model created with assumed rate constants	Model predicted by STGP
7.1	$r = \frac{0.3\, P_A - 0.039\, P_B}{1 + 0.78\, P_B}$	$r = \frac{P_A^{\,2}}{3.03\, P_A + 1.72\, P_B + 0.052\, P_A^{\,2} + P_A\, P_B + P_A^{\,2}\, P_B}$
7.2	$r = \frac{0.49\, P_A - 0.08\, P_B}{1 + 1.23\, P_A}$	$r = \frac{0.32\, P_A - 0.02\, P_B}{1 + 0.305\, P_A + 0.024\, P_A^{\,8} + 0.07\, P_A^{\,9} + 0.07\, P_A^{\,5} + 0.32\, P_A^{\,6}}$

When both the numerator and denominator are divided by 3.03 P_A, the additional terms can be neglected the predicted rate law of model 7.1. The term (0.039 P_B) is neglected by the program due to its small numerical value. Model 7.2 shows a good structural resemblance to the model created with assumed rate constants despite the negligible terms whose power are greater than 5. Both of generated models matches with the data generated quite well as it can be seen in Figure 4.7.

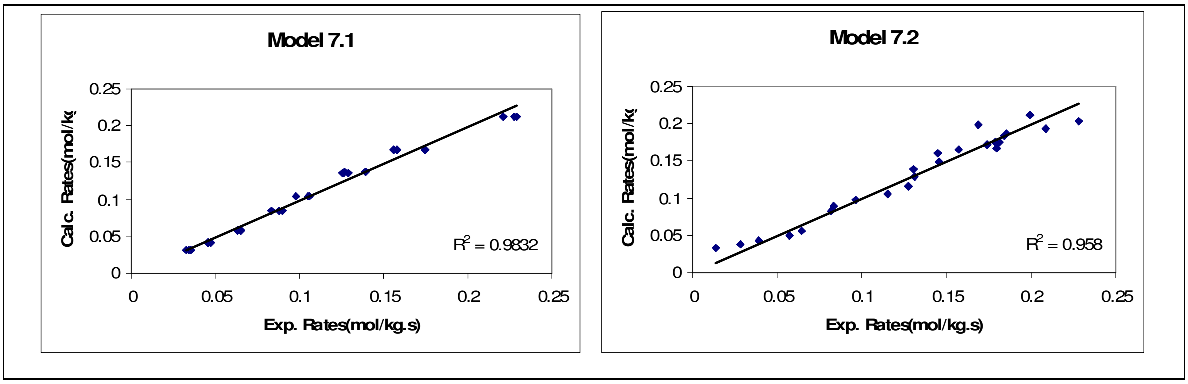

Figure 4.7. Predicted vs. generated rates for Mechanism 7

The main cause of convergence problems in those models was the nature of minus operator that could have great effect on the attainment of negative values. Also the utilization of minus function in this part of generic reactions section resulted in sub trees that didn't have any effect on outputs of generated model. The operation procedure of the proposed algorithm in this work was explained in detail in computational details section. While the minus function is not used, the genetic operators can easily manipulate the driving force group of individuals in the population whose node numbers are not increased yet at the beginning of the runs. The algorithm that already estimated the right structure in the numerator of mathematical expressions continues to run generation after generation and the tree structured individuals grow in size in the direction of denominators in order to converge to the desired model. The minus function hardens the structural estimation of numerator and also makes the estimation of denominator structure more difficult at the same time by increasing the node number in numerator of expressions that decreases the possibility of genetic operator implementation on the nodes in the denominators.

4.2. Dehydrogenation of Methylcyclohexane

As being one of the unimolecular reactions, the dehydrogenation of methylcyclohexane to produce toluene is a great test for the reliability of our strongly typed genetic programming algorithm in deriving model equations. The experimental data for the reaction was taken from Sinfelt *et al.* (1960). The reaction is also given in Fogler (1992) as an exercise in which the most suitable rate law is asked to be found among suggested four rate laws.

The reaction is carried out over a 0.3% Pt/Al_2O_3 catalyst in a differential catalytic reactor for the aim of coking prevention in the presence of hydrogen. The reaction is:

$$C_7\,H_{14}\,(M) \rightarrow C_7\,H_8\,(T) + 3\,H_2 \tag{4.23}$$

and the rate law of the reaction that is proposed as:

$$r_T = \frac{12.3\,P_M}{1 + 9\,P_M} \tag{4.24}$$

The STGP program was run using various functions and additional terminals, and the models were presented in Table 4.15. Although model M.1 and M.3 has almost a perfect match with the originally proposed model in their structures and constants, the model M.2 quite different indicating that the choice of the functions and terminals are critical. The experimental and predicted rates however, are again in a quite agreement with each other even for model M.2 (Figure 4.8).

Table 4.15. List of models derived for dehydrogenation of methylcyclohexane

Model No	Models	F* (num)	F* (den)	Additional Terminal
M.1	$r_T = \dfrac{12.36\,P_M}{1 + 9.361\,P_M}$	*	*,+	-
M.2	$r_T = \dfrac{3\,P_M^{1/2}}{1 + P_M^{1/2} + 0.429\,P_M}$	*	*,+	$x^{1/2}$
M.3	$r_T = \dfrac{12.7\,P_M}{1 + 9.58\,P_M}$	$*, x^{1/2}, x^2$	$*, +, x^{1/2}, x^2$	-

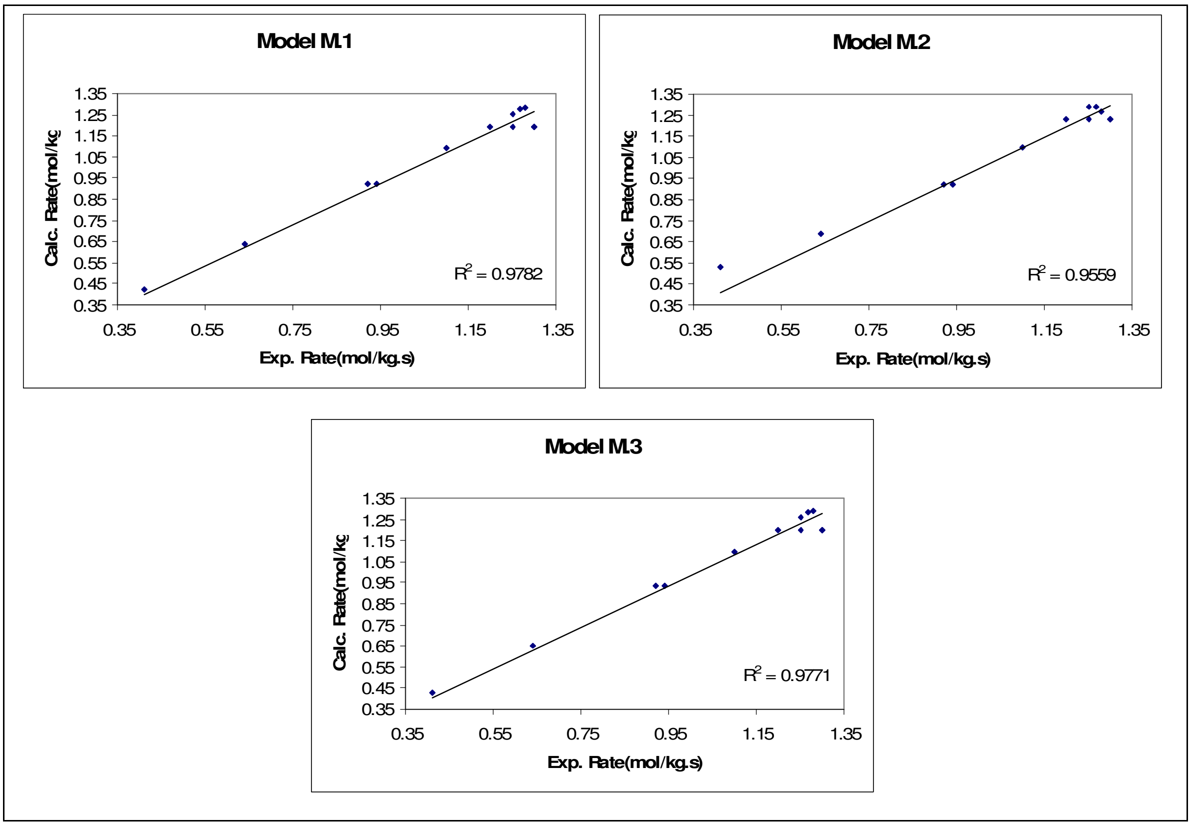

Figure 4.8. Calculated rates vs. experimental rates for dehydrogenation of methylcyclohexane

4.3. Formation of Methane from Carbon Monoxide and Hydrogen

Pursley (1952) worked on the formation of methane from carbon monoxide and hydrogen with the contribution of a nickel catalyst. As a result of having two products, bimolecular reactions such as methane formation have more complex rate expressions. The reaction is:

$$3\,H_2 + CO \rightarrow CH_4 + 2\,H_2O \qquad (4.25)$$

The temperature that the reaction was carried out in a differential reactor was 500°F. At the entrance of the reactor, the partial pressures of H_2 and CO were determined. The proposed rate law is in the form of equation 4.26. with the prediction of rate limiting step being the reaction between atomic hydrogen adsorbed on the nickel surface and CO in the gas phase.

$$r_{CH_4} = \frac{0.0183\,P_{H_2}^{\;1/2}\,P_{CO}}{1 + 1.5\,P_{H_2}} \qquad (4.26)$$

The models obtained are summarized in Table 4.16. with the functions and additional terminals that were used. As it was necessary, the square roots of variables were added in the modeling of this reaction as a terminal in two cases. Square root was not included in the function set for the reason of application on variables more than one time. All the models have good R^2 values as seen in Figure 4.9. However, model MET.1 is questionable since it falsely suggests dual site mechanisms while the reaction is believed to follow single site Eley-Riedel mechanism. On the other hand, model MET.2 and model MET.3 are the more similar ones by the means of equation structure. Model MET.3 matches with the proposed rate law in the literature perfectly at the same time having the highest fitness value. Model MET.2 can be also accepted as a correct rate law equation since the terms with low numerical constants can be neglected leading to model MET.3. There is no big difference between the R^2 values of the models while they do not show any inappropriateness by means of kinetic reaction rate law structure. Hence, all of the models can be assumed as acceptable. In those kinds of situations, models with better R^2 values are more preferable.

Table 4.16. Models derived for formation of methane

Model No	Models	F* (num)	F* (den)	A* T*
MET.1	$r_{CH_4} = \frac{0.102\, P_{H_2}\, P_{CO}}{1 + 10.68\, P_{H_2} + 2.312\, P_{H_2}^{\,2}}$	*	*,+	-
MET.2	$r_{CH_4} = \frac{0.01667\, P_{H_2}^{\,1/2}\, P_{CO}}{1 + 1.253\, P_{H_2} + 0.0042 \frac{P_{H_2}}{P_{CO}} + 0.0046 \frac{P_{H_2}^{\,2}}{P_{CO}} + 0.0012 \frac{P_{H_2}^{\,3}}{P_{CO}} + 0.0044\, P_{H_2}^{\,2}}$	*	*,+	$x^{1/2}$ (num)
MET.3	$r_{CH_4} = \frac{0.0188\, P_{H_2}^{\,1/2}\, P_{CO}}{1 + 1.564\, P_{H_2}}$	*	*,+	$x^{1/2}$

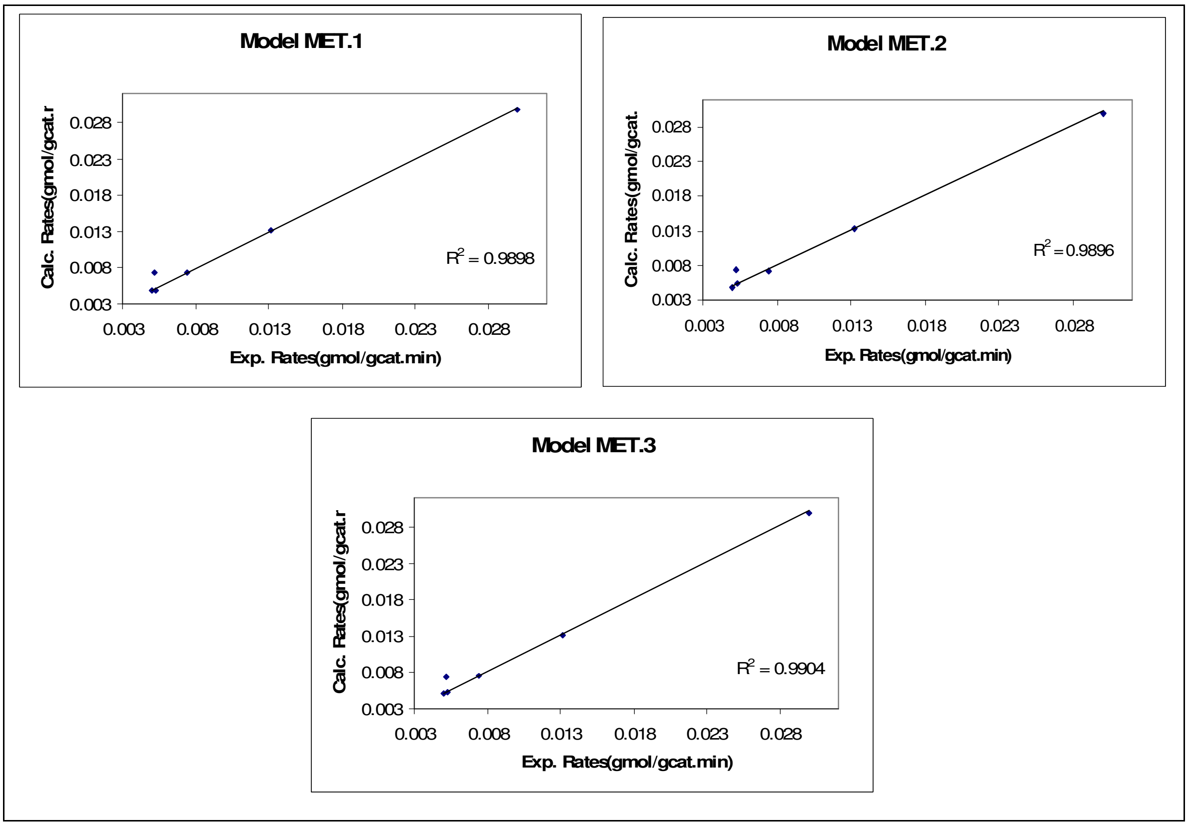

Figure 4.9. Calculated rates vs. experimental rates for formation of methane

4.4. Hydrogenation Reaction of Ethylene

As a bimolecular reaction other than formation of methane from carbon monoxide and hydrogen, the hydrogenation reaction of ethylene is also selected for testing the success of strongly typed genetic programming algorithm that is proposed in our work. The reaction was modeled by using experimental data described in Fogler (1992).

$$C_2H_4\,(E) + H_2\,(H) \rightarrow C_2H_6\,(EA) \tag{4.27}$$

The reaction is carried out over a cobalt molybdenum catalyst. To find the best rate law, four different rate law equations that were given in 4.28 to 4.31 were suggested in Fogler (1992) meaning to be a great task for strongly typed genetic programming algorithm in determining the most appropriate model. Equation 4.28 was reported as the best model.

$$r_{EA} = \frac{3.19\,P_H\,P_E}{1 + 2.1\,P_E} \tag{4.28}$$

$$r_{EA} = \frac{3.35\,P_E\,P_H}{1 + 0.043\,P_{EA} + 2.2\,P_E} \tag{4.29}$$

$$r_{EA} = \frac{2\,P_E\,P_H}{(1 + 0.36\,P_E)^2} \tag{4.30}$$

$$r_{EA} = 0.89\,P_E^{0.26}\,P_H^{1.06} \tag{4.31}$$

The calculated results with different functions are given in Table 4.17. The first two of three models are in excellent agreement with the rate law equation given in equation 4.28 since the last term and the denominator of Model E.2 can be dropped out compared to other two. Again if we introduce square root terms as function of terminal, the STGP is converging to some other rate law expressions that are not reflecting the reality. On the other hand, the R^2 values for first two models are clearly higher than the third one suggesting the right rate law although it is not hundred percent conclusive (Figure 4.10). This was an important test for our algorithm that chose the best model between the four equations that were given above.

Table 4.17. Models derived for hydrogenation reaction of ethylene

Model No	Model	F* (num)	F* (den)
E.1	$r_{EA} = \frac{3.0043\,P_H\,P_E}{1+1.89\,P_E}$	*	*,+
E.2	$r_{EA} = \frac{3.3.45\,P_H\,P_E}{1+2.187\,P_E+0.0268\,P_H}$	*	*,+
E.3	$r_{EA} = \frac{0.9742\,P_H\,P_E^{1/2}}{(0.0181\,P_H\,P_E+0.0012\,P_{EA}^{4}\,P_E^{2}\,P_H^{2})^{1/4}}$	*,$x^{1/2}$	*,+,$x^{1/2}$

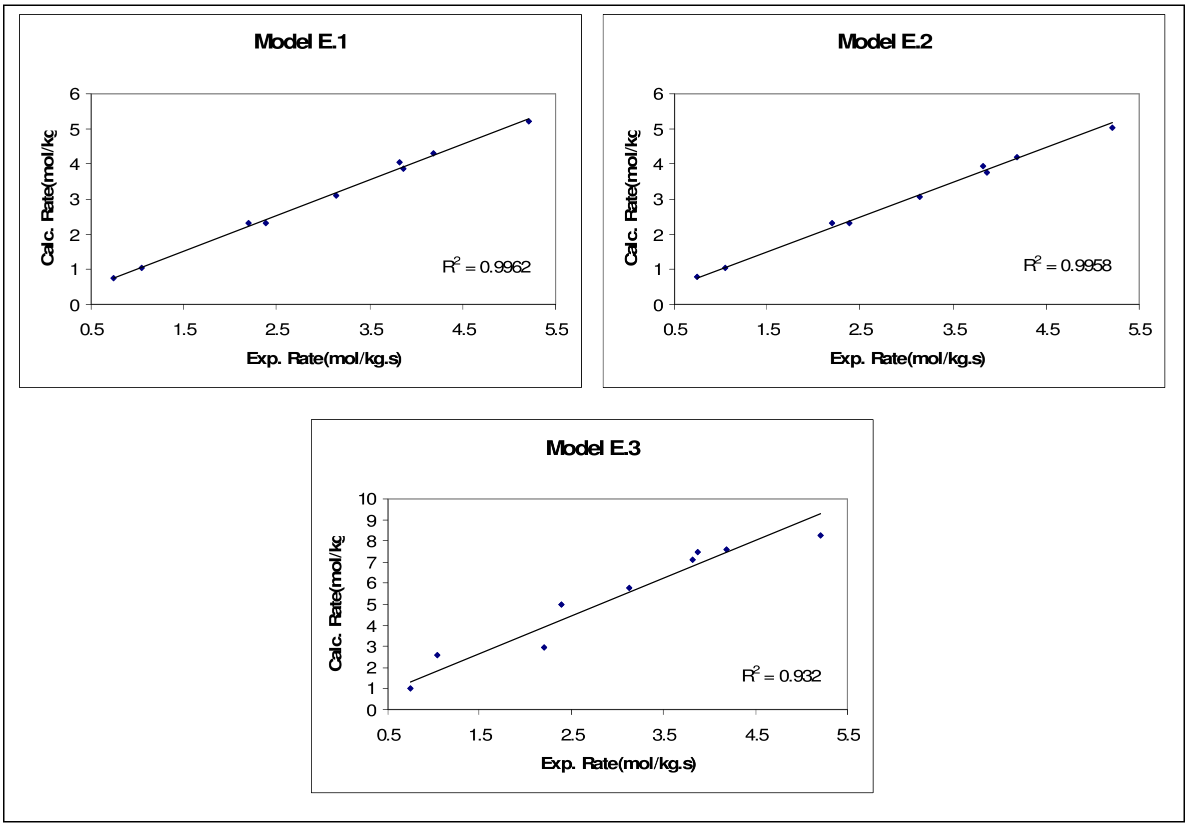

Figure 4.10. Calculated rates vs. experimental rates for hydrogenation of ethylene

4.5. Modeling of Experimental Data taken from Özyönüm (2005)

The kinetics of selective CO oxidation in hydrogen rich streams over Pt-Co-Ce/Al_2O_3 catalyst was studied by Özyönüm (2005). The intrinsic kinetic data were obtained in the initial rate region in a micro flow reactor operating in a differential mode using eight different sets of CO and O_2 concentrations, each at two space times and at 110 °C in the presence of 25 percent CO_2 and 10 percent H_2O. Two different reaction pathways were found to be satisfying the positive constant requirement in regression analysis. First pathway was:

$$2CO + 2* \leftrightarrow 2CO* \tag{4.32}$$

$$O_2 + s \leftrightarrow O_2s \tag{4.33}$$

$$O_2s + s \leftrightarrow 2Os \tag{4.34}$$

$$2CO + 2Os \rightarrow 2CO_2 + 2* + 2s \tag{4.35}$$

In this bifunctional pathway, the reaction occurs at the interface between oxygen adsorbed on cobalt-ceria surface and CO adsorbed on the noble metal surface. Cobalt-ceria and noble metal surfaces are denoted respectively as “s” and “*”. While the oxygen chemisorption is assumed to be slower than the dissociation and the noble surface is covered completely with CO, the rate expression was suggested as:

$$-r_{CO} = \frac{k_1 P_{O_2}}{K_2 P_{O_2} + 1} = \frac{6.66\times10^{-6} P_{O_2}}{9.65\times10^{-4} P_{O_2} + 1} \tag{4.36}$$

The second path which is also a bifunctional path includes reversible adsorption of CO on both noble metal and cobalt-ceria sites, the adsorption of oxygen on cobalt ceria sites in two steps and the formation of CO_2 via two Langmuir-Hinshelwood surface reactions.

$$CO + * \leftrightarrow CO* \tag{4.37}$$

$$CO + s \leftrightarrow COs \tag{4.38}$$

$$O_2 + s \rightarrow O_2s \tag{4.39}$$

$$O_2s + s \rightarrow 2Os \tag{4.40}$$

$$CO* + Os \rightarrow CO_2 + * + s \tag{4.41}$$

$$COs + Os \rightarrow CO_2 + 2s \tag{4.42}$$

If the oxygen chemisorption is slower than the dissociation and CO* is assumed to be the most abundant reaction intermediate in addition to the CO adsorption on cobalt-ceria sites being in equilibrium, the rate expression and the parameters were proposed as:

$$-r_{CO} = \frac{K_1 P_{O_2}}{K_2 P_{CO} + 1} = \frac{3.5\times10^{-6} P_{O_2}}{8.31\times10^{-4} P_{CO} + 1} \tag{4.43}$$

As in the first generic examples discussed above, ”addition” in denominator and “multiplication” in both numerator and denominator were used as function set of our algorithm for evaluating CO oxidation rate law expressions from the data taken from the work of Özyönüm (2005). After fifteenth generation, the diversity in the population of solution candidates suddenly decreased and the best so far individual did not change for the rest of the run. The model evaluated and the experimental vs. calculated rate graph with R^2 value are given in equation 4.44. and in Figure 4.11. respectively.

$$-r_{CO} = \frac{0.0232 P_{O_2}}{1 + \frac{P_{O_2}}{P_{CO}} + 2.328 P_{CO} + 2.328 P_{O_2}} \tag{4.44}$$

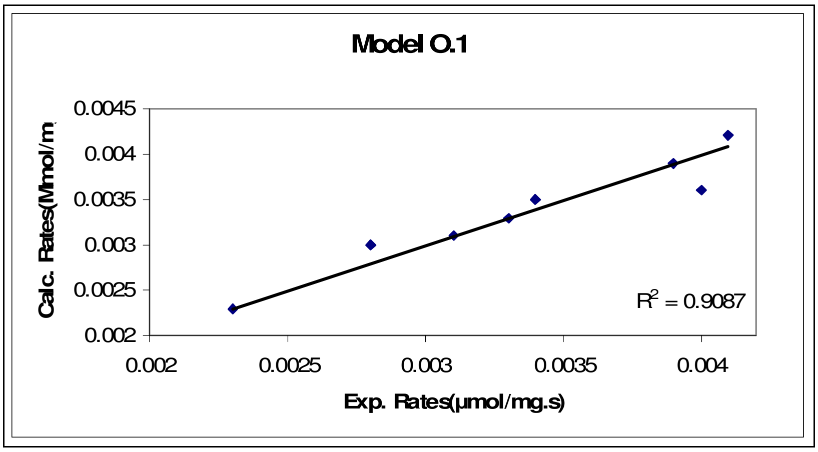

Figure 4.11. Calculated rates vs. experimental rates for CO oxidation

Contrary to the proposed rate law equations and reaction mechanisms of Özyönüm (2007), the evaluated rate law in our work resembles more to the one that rate law equation of the monofunctional reaction path. One of the possible reaction mechanisms mentioned in Özyönüm (2007) that our solution candidate belongs to as follows:

$$O_2 + S \rightarrow O_2.S \tag{4.45}$$

$$O_2.S + S \rightarrow 2O.S \tag{4.46}$$

$$2CO + 2O.S \leftrightarrow OCO.S \tag{4.47}$$

$$2OCO.S \rightarrow 2CO_2 + 2S \tag{4.48}$$

The monofunctional reaction path takes place via the Eley-Rideal step between adsorbed oxygen and CO in the gas phase including the OCO.S surface specie formation on platinum. With the assumption of the molecular chemisorption of oxygen to be rate determining, the rate law expression for the monofunctional reaction path is:

$$-r_{CO} = \frac{2\, k_3^f\, k_7^f\, k_8^f\, P_{CO}\, P_{O_2}\, [S]_0}{2\, k_3^f\, P_{O_2}\left(k_7^f + k_8^f\right) + k_7^f\, k_8^f\, P_{CO} + 2\, k_3^f\, k_7^f\, P_{CO}\, P_{O_2}} \tag{4.49}$$

As confronted with the model representing the monofunctional reaction path with the rate determining step being chemisorption of oxygen, the evaluated model in our work matches except for an additional term in the denominator of the expression. The possible reason for this kind of result can be the kinetic constant values of the desired rate law expression being so close to each other while there's no big difference in the input pressure of the components.

4.6. Modeling of Experimental Data taken from Demir (2009)

Kinetic study of selective CO oxidation was performed by Demir (2009) over Au/Al_2O_3 catalyst prepared using homogeneous deposition-precipitation method. Only the single-site mechanisms proceeding on Au surface were considered since the Al_2O_3 support is believed to be inert for CO oxidation. The presence of H_2, H_2O and CO_2 in the feed stream was assumed to have a great effect on the rate parameter, but not on the reaction mechanism. The rates of CO consumption were evaluated in a microflow reactor operating in differential mode using ten sets of CO and O_2 concentrations both in the absence and presence of 10 per cent H_2O and 25 per cent CO_2 for two residence times at the temperatures ranging from 90°C to 130°C. Finally, the model discrimination was performed for the most plausible models among the mechanisms proposed using the experimentally measured and calculated data. The monofunctional mechanism including the adsorption of O_2 and CO on the gold sites concluded to be the best fit for the experimental data. The mentioned mechanism that is proposed in the work of Haruta and Date (2001) is given in Table.18. “a” denotes an adsorption site on Au surface.

Table 4.18. Elementary steps of proposed reaction mechanism for
CO oxidation over $AuAl_2O_3$

Elementary Step	Step Number
$CO + a \underset{k_{-1}}{\overset{k_1}{\Leftrightarrow}} COa$	(1)
$O_2 + a \xrightarrow{k_2} O_2a$	(2)
$COa + O_2a \xrightarrow{k_3} CO_2 + Oa + a$	(3)
$COa + Oa \xrightarrow{k_4} CO_2 + 2a$	(4)
$COa + O_2a \xrightarrow{k_5} CO_3a + a$	(5)
$CO_3a \xrightarrow{k_6} CO_2 + Oa$	(6)

Two rate law equtions were proposed in the study of Demir (2009) based on the mechanism in Table 4.18. The reaction rate in equation 4.50. is derived by the help of the first four steps while the reaction rate given in equation 4.51. comprises all of them.

$$-r_{CO} = \frac{(k_2 P_{O_2})(K_1 k_3 P_{CO} [S]_0 - k_2 P_{O_2})}{K_1 k_3 P_{CO} (K_1 P_{CO} + 1)} \tag{4.50}$$

$$-r_{CO} = \frac{2 K_1 k_2 k_5 k_6 P_{CO} P_{O_2} [S]_0}{K_1 k_5 k_6 P_{CO} (K_1 P_{CO} + 1) + k_2 k_6 P_{O_2} + K_1 k_2 k_5 P_{CO} P_{O_2}} \tag{4.51}$$

In our work, the experimental results that were attained in the conditions including the absence of CO_2, H_2O and in the presence of 60% H_2, and balance He at 130°C were modeled with the proposed algorithm. The reaction rate is as in equation 4.52.

$$-r_{CO} = \frac{0.03541 P_{CO} P_{O_2}}{1 + 0.6317 P_{CO} + 0.2827 P_{CO}^2} \tag{4.52}$$

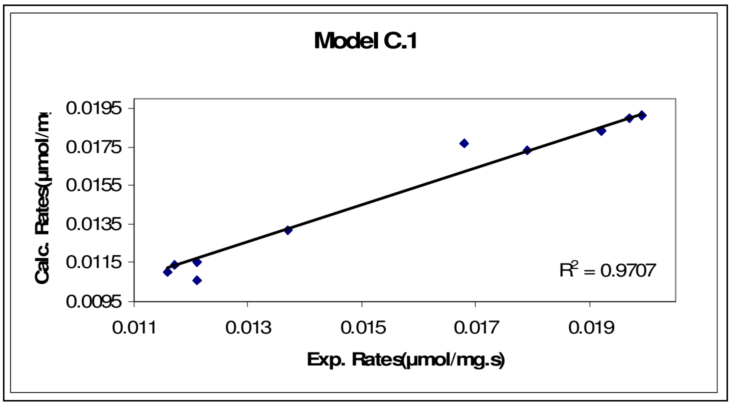

Figure 4.12. Calculated rates vs. experimental rates for CO oxidation

It can be suspected a dual site mechanism while the denominator of the expression resembles to the expended form of a mathematical statement.

5. CONCLUSIONS AND RECOMMENDATIONS

5.1. Conclusions

The kinetic modeling of various generic and real catalytic reactions was performed by using strongly typed genetic programming, and the models were compared with the models that are derived from the mechanisms to test the usability of the generic programming method to predict the reaction mechanism. It was found that the strongly typed generic programming techniques are quite successful predicting the simple model arising from a simple mechanisms or complex mechanisms with simplifying assumptions. It should be also noted that the 75% of the catalytic reactions are controlled by the surface reaction step, and our approach is especially successful in this type of reactions. However, the following complexities reduce the ability to predict the mechanism although the models obtained are still well representing the experimental data. The existence of too many terms in the denominator increases the number of mathematically possible solutions that the program may converge to a mathematically perfect but not physically meaningful solution.

- If the “subtraction” function in the numerator exists, the program may not converge due to the uncontrollable increase of the numerator.

- The models that include the square of the whole denominators (dual site with the surface reaction is rate determining) could not be perfectly predicted because the program is working on an expended expression, and this expression could not be fully converted to the whole square form although sometimes the terms in the denominator implies the square structure. Also it was observed that the utilization of square function did not help the solution of the problem.

- The proposed algorithm in this work is questionable for the evaluation of rate laws for mechanisms including dissociation with rate determining steps being adsorption (requiring the use of square root terms). The square root, as a function, complicates the rate expression unrealistically while the addition of the square roots as a terminal make the conversion more difficult by reducing the population diversity, although the models obtained this way are more realistic.

5.2. Recommendations

The weaknesses of the proposed algorithm showed that there can be further improvement for the purpose of perfection. In order to reach that goal, following recommendations for upcoming studies can be done:

- In order to derive rate law equations in a more accurate way, an algorithm giving credit to the certain combinations of functions and terminals can be beneficial under the main program structure of our work.

- The denominator structure in the generation of initial population and the implementation procedure of genetic operators can be changed as an alternative to our algorithm: The variables can be set as untouchable elements and the constants, or the combination of the constants can be exchanged or mutated resulting a large number of terms in the denominator with different weights. Then the small terms can be neglected to reach the real structure of the model. The numerator part of expression can be formed with the same way as it was in our work (all the possible variable combinations multiplied with randomly combined) since our approach seems to work there.

REFERENCES

Androulakis, I. P., V. Venkatasubramanian, 1991, “A Genetic Algorithm Framework for Process Design and Optimization”, *Computers and Chemical Engineering*, Vol.15, pp. 217-228.

Cao, H., J. Yu, L. Kang, Y. Chen and Y. Chen, 1999, “The Kinetic Evolutionary Modeling of Complex Systems of Chemical Reactions”, *Computers and Chemistry*, Vol. 23, pp. 143-151.

Dawkins, R. and P. J. Bentley (editors), 1999, *Evolutionary Design by Computers*, Morgan Kaufmann Publication, San Francisco, pp. 8-13.

Demir, M., 2009, *Kinetic Study of Selective CO Oxidation on Au / Al_2O_3 Catalyst*, M. S. Thesis, Boğaziçi University.

Eiben, A. E. and J. E. Smith, 2003, *Introduction to Evolutionary Computing*, Springer, New York, pp. 15-18.

Filippi, C., J. L. Greffe, J. Border, J. Villermaux, J. L. Barnay, P. Bonte and C. Georgakis, 1986, “Tendency Modeling of Semibatch Reactors for Optimization and Control”, *Chemical Engineering Science*, Vol. 41, pp. 913-920.

Fogler, H. S., 1992, *Elements of Chemical Reaction Engineering*, 2nd ed., Prentice-Hall, Englewood Cliffs, New Jersey, pp. 241-337.

Galvan, I. M., J. M. Zaldivar, H. Hernandez and E. Molga, “The Use of Neural Networks for Fitting Complex Kinetic Data”, 1996, *Computers and Chemical Engineering*, Vol. 20, No. 12, pp. 1451-1465.

Gianoli, S. I., G. Puxty , U. Fisher , M. Maeder and K. Hungerbühler , 2007, “Empirical Kinetic Modeling of on Line Simultaneous Infrared and Calorimetric Measurement Using a Pareto Optimal Approach and Multi-objective Genetic Algorithm”, *Chemometrics and Intelligent Laboratory Systems* , Vol.85, pp. 47–62.

Goldberg, D. E., 1989, *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison Wesley, Reading, MA.

Haynes, T., R. Wainwright, S. Sen and D. Schoenefeld, 1995, "Strongly Typed Genetic Programming in Evolving Cooperation Strategies", Proceedings of the Sixth International Conference on Genetic Algorithms, Pittsburg, July 1995, pp. 271-278, Morgan Kauffman.

Hinchliffe, M. P. and M. J. Willis, 2003, "Dynamic Systems Modelling using Genetic Programming", *Computers and Chemical Engineering*, Vol.27, pp. 1841-1854.

Hougen, A. O. and M. K. Watson, 1961, *Chemical Process Principles: Part Three Kinetics and Catalysis,* 9th ed., John Wiley and Sons, Inc., New York, pp. 902-972.

Koza, J., 1992, *Genetic Programming: On the Programming of Computers by Means of Natural Selection*, 6th ed., MIT Press, Cambridge, MA, pp. 479-526.

Koza, J., 1994, *Genetic Programming II: Automatic Discovery of Reusable Programs*, The MIT Press, USA, section 2.

Luke, S. and L. Panait, 2002, "Lexicographic Parsimony Pressure", Proceedings of GECCO, Morgan Kauffman, pp. 829-836.

Masel, R. I., 1996, *Principles of Adsorption and Reaction on Solid Surfaces*, Wiley, New York, pp. 493-496.

McKay, B., M. Willis and G. Barton, 1997, "Steady State Modeling of Chemical Process Systems Using Genetic Programming", *Computer Chemical Engineering*, Vol. 21, pp. 981-996.

Mitchell, M., 1996, "*An Introduction to Genetic Algorithms*", MIT Press, Cambridge, MA.

Montana, D. J., 1994, "Strongly Typed Genetic Programming", Technical Report 7866, Bolt Baranek and Newman, Inc., Cambridge, MA.

Moros, R., H. Kalies, H. G. Rex and St. Schaffarczyk, 1996, "A Genetic Algorithm for Generating Initial Parameter Estimations for Kinetic Models of Catalytic Processes", *Computers and Chemical Engineering,* Vol. 20, No. 10, pp. 1257-1270.

Nordvik, J. P. and J. M. Renders, 1991, "Genetic Algorithms and Their Potential for Use in Process Control", *Proc. of the 4th International Conference on Genetic Algorithms*, pp. 480-486.

Özyönüm, G. N., 2005, *Kinetics of Selective CO Oxidation in Hydrogen – Rich Streams over Pt-Co-Ce/ Al_2O_3 Catalyst*, M. S. Thesis, Boğaziçi University.

Özyönüm, G. N., A. N. Akın and R. Yıldırım, 2007, "Kinetic Study of Selective CO Oxidation over Pt-Co-Ce/Al_2O_3 Catalyst in Hydrogen rich Streams", *Turkish Journal of Chemistry*, Vol. 31, pp. 445-453.

Pursley, J. A., 1952, *An Investigation of the Reaction between Carbon Monoxide and Hydrogen on a Nickel Catalyst above One Atmosphere*, Ph.D. Thesis, University of Michigan.

Ramin, B. B. and M. Masoori.,2007, " Which method is Better for the Kinetic Modeling: Decimal Encoded or Binary Genetic Algorithm?", *Chemical Engineering Journal*, Vol. 130, pp. 29-37.

Rein, G., C. Lautenberger, A. C. Fernandez-Pello, J. L. Torero and D. L. Urban, 2006, "Application of Genetic Algorithms and Thermogravimetry to determine the Kinetics of Polyurethane Foam in Smoldering Combustion", *Combustion and Flame*, Vol.146, pp. 95-108.

Schulze-Kremer, S., 1992, "Genetic Algorithms for Prediction of Protein Conformations", *IEE Colloquium Digest*, Vol.119, pp. 1.

Sinfelt, J. H., H. Hurwitz and R. A. Shulman, 1960, "Kinetics of Methylcyclohexane Dehydrogenation over Pt/Al_2O_3", *The Journal of Physical Chemistry*, Vol.64, pp.1559-1562.

Silva, S., 2007, "GPLAB-A Genetic Toolbox for MATLAB", Version 3, Evolutionary and Complex Systems Group

Silva, S. and J. Almeida, 2004, "Dynamic Limits for Bloat Control – Variations on Size and Depth", in K. Deb (eds.), Proceedings of GECCO, Springer-Verlag, Berlin, pp. 666-677.

Steinfield, I. J., J. S. Francisco and L. W. Hase, 1989, *Chemical Kinetics and Dynamics*, Prentice Hall, Englewood Cliffs, New Jersey, pp. 109-110.

Van Der Laan, G. P. and A. C. M. Beenackers, 2000, "Intrinsic Kinetics of the Gas –Solid Fischer-Tropsch and WGS Reactions over a Precipitated Iron Catalyst", *Applied Catalysis*, Vol. 193, pp. 39-53.

Willis, M., H. Hiden, M. Hinchliffe, B. McKay, G. W. Barton, 1997, "Systems Modeling using Genetic Programming", *Computers and Chemical Engineering*, Vol. 21, pp.1161-1166.