Application of Monte Carlo techniques to LCO gas oil hydrotreating: Molecular reconstruction and kinetic modelling

Maria Lopez Abelairas, Luis Pereira de Oliveira, Jan Verstraete

HAL Id: hal-01338408
https://hal.science/hal-01338408

Submitted on 28 Jul 2016

1   **Application of Monte Carlo techniques to LCO gas oil hydrotreating: molecular**

2   **reconstruction and kinetic modelling.**

3

4   Maria Lopez Abelairas, Luis Pereira de Oliveira, Jan J. Verstraete[*]

5

6   ***Jan J. Verstraete: corresponding author**

7   E-mail address: jan.verstraete@ifpen.fr

8   Tel: +33 4 78 02 20 53, ext. 16014; Fax: +33 4 78 02 20 08

9   Postal address: IFP Energies nouvelles, Rond-point de l'échangeur de Solaize, BP 3, 69360

10   Solaize, France

11

12   **Abstract**

13   The increasing pressure to satisfy the environmental criteria concerning oil products leads to

14   the need for developing accurate models to predict the performances of the refining processes.

15   In the current study, a stochastic two-step procedure using Monte Carlo techniques is applied

16   to and validated on the hydrotreating of Light Cycle Oil (LCO) gas oils. In the first step, a

17   mixture of molecules representative of the LCO gas oils is generated using a molecular

18   reconstruction method termed SR-REM. Subsequently, the Stochastic Simulation Algorithm

19   (SSA) is applied to simulate the evolution of the mixture composition during hydrotreating.

20   The results show that an accurate representation of eleven different LCO gas oils was

21   obtained by the application of the molecular reconstruction method. The hydrotreating

22   simulations of three LCO gas oils at different operating conditions showed a good agreement

23   with the experimental data obtained at laboratory scale. The current stochastic procedure has

24   demonstrated to be a valid tool for the reconstruction of the composition of LCO gas oils and

25   the simulation of the hydrotreating process.

26

27   **Keywords**

30   **1. Introduction**

31   The pressure on the quality and the maximum impurities content of refinery products has been

32   increased in the last decades due to the environmental concerns associated to their use. The

33   suitable performance of the processes aimed to remove these impurities is essential to satisfy

34   the specifications fixed for oil products. Among these processes, hydrotreating (HDT) is one

35   of the most mature technologies. The use of hydrogen at high temperature over a catalyst bed

36   allows to remove sulphur, nitrogen and other undesirable elements from the petroleum

37   distillates such as naphthas or gas oils.

38   The correctness in the prediction of HDT process performance directly depends on the

39   reliability of the used kinetic model. A lumping strategy, in which molecular components are

40   grouped according to their global properties, has usually been applied in the kinetics models

41   of complex hydrocarbons [1–3]. However, this approach is no longer manageable due to the

42   elevated number of generated groups and reaction pathways. The limited applicability of

43   lumped models has resulted in the development of molecular-based kinetic models [2,4–9]. In

44   this latter approach, reactant species and products are described by a selection of molecules or

45   small groups of them which react following networks formed by reactions at the molecular

46   level and distinguished by a global kinetic constant. The elemental steps are regrouped within

47   the global reactions scheme, thereby reducing the number of reactive species and the size of

48   reaction network. Following this approach, Lopez Garcia *et al.* [8] have simulated the

49   hydrotreating of LCO gas oils and, more recently, Pereira *et al.* [9–11] have proposed a kinetic

50   methodology for its application on conversion of petroleum cuts based on the Stochastic

51   Simulation Algorithm (SSA) developed by Gillespie [12].

52   These kinetics models required a molecular description of the feedstock. However, the high

53   complexity of the petroleum cuts hinders their molecular characterisation even by using the

54   most advanced analytical techniques. To surpass this drawback, a molecular representation of

55   the feedstock can be determined by the application of a molecular reconstruction algorithm.

56   This kind of algorithms provides a set of molecules from petroleum analyses and chemical

57   knowledge. Several approaches have been performed in this sense, from the first model

58   proposed by Liguras and Allen [13] to the stochastic reconstruction (SR) developed by Neurock

59   *et al.* [14] in which a Monte Carlo method is used to sample the objective function.

60   Hudebine [15,16] has characterised different petroleum cuts by the application of a two-steps

61   algorithm based on SR algorithm and on reconstruction entropy maximisation (REM). The

62   combination of both methods (SR-REM) overcomes the drawbacks of its separated

63   application, such as noise of the objective function and computational effort. Posteriorly,

64   Pereira *et al.* [17,18] have modified Hudebine's method to introduce a genetic algorithm which

65   performs the minimisation of the objective function.

66   The current study is focused on the validation of the SR-SEM and SSA algorithms developed

67   by Pereira *et al.* on the hydrotreating of Light Cycle Oil (LCO) gas oils. A brief description of

68   the stochastic methodology applied to both molecular reconstruction and reaction simulation

69   has been developed. Then, the molecular reconstruction of eleven LCO gas oil has been

70   carried out and the hydrotreating of three of them has been simulated at different operational

71   conditions.

72   **2. Methodology**

73   The stochastic methodology applied in the current work consists of two steps. In the first one,

74   a molecular reconstruction algorithm is used to model the feedstock by means of a set of

75   molecules which mixture properties are similar to those of the feedstock. This algorithm

76   comprises two coupled methods: stochastic reconstruction (SR) and reconstruction by entropy

77   maximization (REM). The application of the SR method enables to obtain a set of molecules

78   typical for a given type of oil fraction. Then, the molar fraction of molecules is adjusted using

79   the REM method in order to improve the predicted mixture properties.

80   Finally, the generated set of molecules is introduced as input in the second step concerning

81   the molecule-based kinetic modelling of the conversion process. A kinetic Monte Carlo

82   (kMC) method is applied to model the chemical reaction system at a molecular level. This

83   approach is an alternative to traditional deterministic kinetic methods in which a serial of

84   differential equations (ODE) are involved to obtain a temporal evolution of the species in the

85   system.

86   *2.1. Molecular reconstruction*

87   The SR method, from which a representative set of molecules is generated, uses probability

88   distributions functions (PDF's) of structural attributes, such as number of rings or number of

89   side chains, to characterise the oil fraction to be reconstructed [14,19]. Feedstock analyses, which

90   provide information about atomic abundance (elemental analysis) or structural classes (GC-

91   MS), and the previous knowledge about petroleum cuts should be taken into account to select

92   the proper structural attributes. The selected PDF's are sampled *N* times using a Monte Carlo

93   method in order to generate an initial set of *N* molecules (Figure 1). To carry out the transition

94   from PDF's to the set of molecules, the application of a building diagram and a serial of

95   chemical rules is required to avoid the creation of impossible and unlikely molecules. The

96   building diagram is a decision tree that defines the sequence in which PDF's should be

97   sampled to proceed to the posterior assembly of the different attributes, which provides the

98   structure of the molecule. During the assembly process, the application of chemical rules

99   discards the molecules which do not satisfy the thermodynamic or likelihood criteria.

100   The application of the Monte Carlo sampling procedure led to the creation of a mixture of N

101   molecules, each one with a molar fraction equal to $1/N$. The average properties of this mixture

102   are calculated and compared with the experimental data by means of an objective function.

103   An optimisation method is used to minimise this objective function though the modification

104   of PDF's parameters. The simulated annealing [15,19–21] and the genetic algorithms [22–24] are both

105   usual methods applied to the optimisation of the objective function in the field of the

106   stochastic reconstruction. However, Schnongs [24] has demonstrated that the genetic algorithms

107   are more adapted to the stochastic reconstruction than the simulated annealing since the

108   genetic algorithms are less sensitive to the initial value of the distribution parameters and it

109   has also showed a lower oscillation of the value of the objective function.

110   Once the initial set of molecules is obtained, the second step based on the reconstruction by

111   entropy maximisation (REM) is applied. During this step, the molar fraction of the molecules

112   in the mixture is modified to obtain a mixture with properties closer to those experimentally

113   determined [15].To carry out the adjustment of the parameters, the REM method, based on

114   Shannon's information theory [25], uses the maximisation of an information entropy criterion

115   ($E(x_i)$) given by the Eq. (1):

116
$$E(x_i) = -\sum_{i=1}^{N} x_i \cdot \ln(x_i) \qquad (1)$$

117   with

118
$$\sum_{i=1}^{N} x_i = 1 \qquad (2)$$

119   where x_i corresponds to the molar fraction of molecule i and N is the number of molecules

120   present in the initial mixture determined by applying the SR method. This criterion guaranties

121   that, without constrains or available information, a given molecule cannot be preferred to

122   others, so that the distribution of the predefined set of molecules remains uniform. The

123   introduction of a matrix of constrains from analytical data, mixing rules and the initial set of

124   molecules distorts this uniformity in order to match this information. These constrains are

125   introduced in the entropy maximisation criterion by means of the Lagrange multiplier method.

126   In the case of linear constrains were used, the problem is reduced to an optimisation of J

127   parameters of a no-linear equation that can be performed by the conjugate gradient method.

128   Once the Lagrange parameters are determined, the molar fraction of the molecules is
129   calculated in order to obtain the representative mixture of molecules.

130   *2.2. Reaction modelling*

131   The goal of the kMC method is to describe the chemical reaction system at the molecular
132   level by following the transformations of a discrete population of molecules. Contrary to
133   deterministic methods, the temporal evolution of the system is not described by a set of
134   differential equations, but by a single probabilistic function termed the chemical master
135   equation (CME) [26,27]. This equation determines the different status of the system in a
136   sequence of discrete-time steps. The evolution of the system is a result from chemical
137   reactions, those have associated a probability function, which depends on the reactant
138   molecule and the rate constant.

139   However, CME cannot usually be solved by analytical or numerical methods due to its high
140   complexity. In this case, a Monte Carlo procedure developed by Gillespie [12] and termed
141   Stochastic Simulated Algorithm (SSA) is used to solve CME. This method is based on an
142   event-space approach, i.e., the evolution of the system is followed event by event, resulting in
143   disparate time intervals between two consecutive events or, in this particular case, reactions.
144   The scheme followed in the application of the kMC method has been illustrated in Figure 2.

145   The initial step of the procedure consists in the "molecular discretisation" of the set of
146   molecules generated by applying the SR-REM method. During this step, the replication of the
147   molecules in the mixture is performed according to their molar fraction and it is based on the
148   same principle as the procedure developed by Hudebine and Verstraete [21] for the creation of a
149   representative set of molecules for gasoline fractions. A replication factor (F_{rep}), defined as
150   the maximum number of molecules in the discrete mixture, is multiplied by the mole fraction
151   of each molecule in order to calculate the number of replications. F_{rep} also defines the lowest
152   mole fraction to be retained since the molecules with a mole fraction lower than $0.5/F_{rep}$ are
153   discarded. The factor 0.5 arises from the rounding off to the closest integer.

154   In the next step of the kMC procedure, all potential reaction events should be identified and
155   listed taking into account a set of reaction rules that generates the reactions by inspecting the
156   structure of the molecules. Once all potential reaction events are identified, the normalised
157   probability for each reaction event is calculated as the ratio of its stochastic rate (R_v) to the
158   sum of the rates of all reaction events (M), as shown in Eq. (3):

$$P_v = \frac{R_v}{\sum_i^M R_i} \qquad (3)$$

159

with

160

$$R_v = h_v \cdot c_v \qquad (4)$$

161

where h_v is the number of distinguishable combinations of the reactant molecule and c_v is the stochastic rate parameter. For monomolecular reactions, the number of distinguishable combinations of the reactant molecules h_v equals 1, while for bimolecular reactions, it is equal to the number of available molecules of the second reactant. The stochastic rate parameter c_v is closely linked to the deterministic rate parameter (k_v) of each reaction. In the case of monomolecular reactions, k_v and c_v are equal, while for bimolecular reactions, c_v is equal to the ratio of k_v by reaction volume V.[12]

162
163
164
165
166
167
168

The effect of the reaction temperature on the stochastic rate is taking into account by means of the Arrhenius expression reflected in Eq. (5):

169
170

$$c_v = c_v(T_{ref}) \cdot \exp(-\frac{E_a}{R} \cdot (\frac{1}{T} - \frac{1}{T_{ref}})) \qquad (5)$$

171

where $c_v(T_{ref})$ is the stochastic rate constant of the reaction type v at the reference temperature T_{ref}, E_a is the activation energy, R is the ideal gas constant, T is the system temperature and T_{ref} is the reference temperature.

172
173
174

The subsequent step of the algorithm is aimed to the determination of the cumulative reaction probability distribution (D_R), which grouped all M reactions that can take place in the mixture at a given time t.

175
176
177

A first random number (RN_1) is obtained to determine the reaction time step (Δt) according to the Eq. (6), as proposed Gillespie [12]:

178
179

$$\Delta t = -\frac{\ln(RN_1)}{\sum_i^M R_i} \qquad (6)$$

180

The distribution D_R is randomly sampled in order to select the next reaction (μ) to be executed in the next reaction time step. For this purpose, a second random number between 0 and 1 is drawn (RN_2) and used to select the next reaction from the cumulative distribution D_R, as shown in Eq. (7):

181
182
183
184

$$D_R(\mu - 1) < RN_2 \leq D_R(\mu) \qquad (7)$$

185

6

186 Once the reaction and the time step have been defined, the system is updated by the execution

187 of the selected reaction, which implies the replace of the reactant(s) by the product(s) and the

188 increase of simulation time in Δt. The simulation continues reaction by reaction until the final

189 simulation time is reached and a set of product molecules is obtained.

190 Due to its stochastic nature, the algorithm should be executed several times in order to

191 provide several sets of molecules which average can be a proper representation of the

192 molecules in the reaction system.

193 **3. Application to the molecular reconstruction of LCO gas oils**

194 LCO gas oils are gas oil fractions obtained from the catalytic cracking process and composed

195 by a mixture of saturates (paraffins and naphthenes), olefins and aromatic hydrocarbons[8].

196 These mixtures also show a marginal presence of nitrogen but an important content in

197 sulphur, which is present in heterocyclic structures, such as thiophenes and

198 benzothiophenes[28]. Their boiling point is in the range of 120°C to 450°C, which corresponds

199 to the molecules with a number of carbons between C_8 and C_{30} [29]. According to this data and

200 the mass spectrum, LCO gas oils are formed by molecules with none or one poly-aromatic

201 structure with a number of rings which usually ranges from 0 to 4 [30–32].

202 In the current work, it has been performed the molecular reconstruction of eleven LCO gas

203 oils using the SR and REM algorithms. The PDF's of molecular attributes in Table 1 is used

204 to generated the set of molecules that would represent the LCO gas oils. Two kind of PDF's

205 (histograms and gamma functions) were used to defined molecular attributes. The use of

206 histograms is limited to the attributes with a narrow range of possible values (less than three)

207 while the gamma functions describe the attributes with a higher number of possible values.

208 These PDF's have been defined taking into account the previous knowledge about LCO gas

209 oils composition. Firstly, it is considered that the molecules in mixture are only constituted by

210 carbon , hydrogen and sulphur due to the negligible presence of nitrogen species. Then,

211 nitrogen content is fully assigned to carbon fraction, the most abundant element in mixture.

212 Another hypothesis is that paraffins, which would also include the marginal olefins fraction,

213 are formed by alkyl chains with at least 8 carbons. It is also considered that, according to

214 experimental data, each molecule is formed by a maximum of 4 cycles, in which linear alkyl

215 chains can be the unique substituents, and that the sulphur containing compounds present in

216 the mixture only belong to the thiophenes family and its derivatives.

217 In order to define the relationships between distributions and also identify the sampling steps,
218 a building diagram is needed for the reconstruction of the molecules in LCO gas oils (Figure
219 3). First, the type of molecule to be constructed (saturate or aromatic) is determined by
220 distribution 1. In the case of saturate molecules, distribution 3 allows to determine the number
221 of cyclohexane rings (ranged from 0 to 4). When the number of rings present is null, the
222 molecule to be constructed is a paraffin, which chain length is determined by the application
223 of distribution 8. Contrary, in the case of a naphthenic units with between 1 and 4
224 cyclohexane rings, it is necessary to determine the number (distribution 6) and type
225 (distribution 7) of the side chains. In the case of aromatic molecules, the number of benzenes,
226 cyclohexanes and thiophenes is determined by distributions 2, 3 and 4, respectively. These
227 three distributions are sampled until the total number of rings is inferior or equal to 4. The
228 number of benzenes is always 1 or higher to avoid the creation of naphthenic molecules in
229 this route. Besides, according to experimental data, most of the thiophene rings do not contain
230 cyclohexane rings, so that distribution 3 is not sampled when thiophene rings are present.
231 Once the number and type of cycles is defined, the distributions 5 and 6 are used to find the
232 number of side chains in naphthenic and aromatic rings, respectively, and the distribution 7 is
233 applied to determine the type of chain.

234 Elemental analysis (carbon, hydrogen, sulphur and nitrogen), average molecular weight,
235 specific gravity, simulated distillation, mass spectrometry (MS), ^{1}H NMR and ^{13}C NMR were
236 used to characterise the LCO gas oils analytically. Elemental analysis is used to remove the
237 molecules which are composed by an element with a null experimental value. For example, if
238 sulphur presence has not been detected, all sulphureted molecules (benzothiophenes and
239 dibenzothiophenes) are eliminated. In a similar way, all molecules belonging to a
240 spectrometric family which has not been found in mass spectrometry analyses are suppressed.
241 Spectrometric families are determined according to the number of aromatic and thiophene
242 rings in compounds. Thus the following families can be distinguished: saturates (paraffins and
243 cycloparaffins), monoaromatics (alkylbenzenes, tetralins…), diaromatics (naphthalenes,
244 diphenyls…), triaromatics and compounds with more than three aromatic rings (anthracenes,
245 phenanthrenes…), benzothiophenes and dibenzothiophenes. Meanwhile, simulated distillation
246 enables to discard the molecules with a boiling temperature 5°C under the initial or 5°C over
247 the final boiling point. This correction is introduced due to the bias between the real boiling
248 point and the value obtained from simulated distillation curve [33]. The "experimental"

249 molecular weight was calculated using an API correlation based on specific gravity and

250 simulated distillation [34].

251 *3.1. Sensitivity of the stochastic method*

252 The average properties of the simulated mixture of molecules are calculated by the application

253 of a number of linear mixing rules. An objective function (*OF*) is used to perform the

254 comparison between calculated and measured properties. This function is the sum of the

255 relative deviations between calculated and experimental data and it is given by Eq. (8):

256
$$OF = \frac{1}{N_p} \cdot \sum_{i=1}^{N_p} \delta_i \qquad (8)$$

257 where N_p is the number of properties present in objective function and δ_i is the relative

258 deviation between the calculated and experimental values of property *i* which is calculated

259 following Eq. (9):

260
$$\delta_i = \frac{1}{N_{M,i}} \cdot \sum_{j=1}^{N_{M,j}} \frac{\left| X_{j,i}^{exp} - X_{j,i}^{calc} \right|}{X_{j,i}^{exp}} \qquad (9)$$

261 where $N_{M,i}$ is the number of measurements of property *i* (for example, the number of points in

262 the simulated distillation) and $X_{j,i}^{exp}$ and $X_{j,i}^{calc}$ are, respectively, the experimental and

263 calculated values of measurement *j* of property *i*.

264 An elitist genetic algorithm developed by Pereira *et al.* [18] was applied in order to minimise

265 this objective function by the modification of the parameters of the PDF's for the structural

266 attributes. Genetic algorithms [22–24], as well as simulated annealing [15,19–21] and particle swarm

267 optimization [35], are robust global optimizers, a characteristic required for their application in

268 stochastic models. Besides, according to its elitist nature, the applied methodology allows to a

269 certain percentage of individuals to move to the following generation (in this case, a

270 percentage of 50% was considered), with which the best individuals (i.e. those with the lowest

271 *OF* values) remain in the next generation.

272 The number of molecules is an important parameter in molecular reconstruction since an

273 excessively low number would lead to an inadequate representation of the attributes of the

274 represented mixture. Otherwise, the use of a large number of molecules would imply a

275 significant increase in calculation time. In order to determine the optimal value for this

276 parameter, the sensibility of the stochastic method has been tested by the generation of 100

277 mixtures of *N* molecules from a set of parameter distributions determined for a LCO gas oil.

278   *N* value was varied in a range between 10 and 20000 molecules and the average, minimum,

279   maximum and standard deviation values were determined for the objective function in each

280   point (Figure 4). In addition, the time necessary to perform the calculation was also

281   represented.

282   It was observed that the increase of the size of the mixture leads to a drop of the objective

283   function until attaining an asymptotic value at $N = 5000$ (Figure 4.a). The increase of the

284   number of molecules also implies a rapprochement of the minimum and maximum values and

285   a reduction in the standard deviation of the objective function. However, the drop of the

286   standard deviation is very mild at $N > 5000$, while the calculation time continues to rise

287   linearly with the size of the mixture (Figure 4.b). Then, a set of 5000 molecules was

288   considered as optimal to perform the stochastic reconstruction of LCO gas oils.

289   *3.2. Construction of the molecular representation of LCO gas oils*

290   The analyses used in this case for the molecular reconstruction were the elemental analysis,

291   MS and simulated distillation. The elemental analysis provides information on the number of

292   thiophenes, while the information on the chemical structure is provided by MS and by

293   simulated distillation in the case of the average molecule size and size distribution. The

294   experimental and simulated properties were compared in Table 2 for a LCO gas oil.

295   Properties that were not used in reconstruction step (specific gravity, molecular weight, [1]H

296   NMR and ^{13}C NMR) were calculated from the predicted molecular mixture and also

297   compared with the corresponding experimental values.

298   The properties of the mixture obtained after the application of the SR step are mostly in good

299   agreement with the analytical results. Some deviations are nevertheless observed in the

300   simulated distillation curve. These differences may be explained by the fact that the

301   distillation curve is calculated under the assumption that all molecules are perfectly separated

302   by the use of increasing boiling points. However, this is not strictly true (especially in the

303   region near the initial and final distillation points), since the chromatographic columns are not

304   completely apolar and separation is influenced by the polarity of the molecules [29].

305   On the other hand, the coupled application of SR and SEM schemes led to a significant

306   improvement in the agreement between predicted and experimental data for most of the

307   properties (Table 2). The most meaningful effects were observed on more specific and

308   detailed analyses, such as NMR and MS, while only small improvements could be detected in

309   the more global measurements, such as in the case of elemental analysis.

310   On the following step, a total of eleven LCO gas oils have been reconstructed, and the

311   properties of the predicted mixtures have been compared with the corresponding experimental

312   data in order to perform the validation of the current methodology (Figure 5). It was found a

313   good agreement with the analytical results for most of the evaluated properties. However, a

314   deviation was detected for the predicted average molecular weight, which was mostly placed

315   below the value calculated from experimental data. This may be explained by the alterations

316   in the determination of the distillation curve, which imply a reduction in the boiling point of

317   the molecules with respect to their normal value. It leads to the prediction of molecules

318   smaller than their counterpart and, consequently, to reduce the average molecular weight of

319   the mixture.

320   These results show that the predicted set of molecules are proper molecular representations of

321   the corresponding LCO gas oils. Some of these simulated mixtures have been used as inputs

322   for the kinetic model of the hydrotreating of LCO gas oil.

323   **4. Application to the hydrotreating of LCO gas oils**

324   Once the representative set of molecules has been generated for each LCO gas oil, the

325   hydrotreating reactions are simulated by applying the kMC method. The experimental data

326   used in this step has been collected from previous studies developed at IFP Energies nouvelles

327   about the hydrotreating of these oil fractions [36].

328   Three main type of reactions are involved in the hydrotreating process: hydrodesulphurisation

329   (HDS), hydrodenitrogenation (HDN) and hydrogenation of benzene and cyclohexane rings

330   and of olefins. However, the reactions of hydrodenitrogenation and hydrogenation of olefins

331   will be neglected due to the lack of nitrogen and olefins in the mixture. Similarly, the HDS

332   reactions are limited to heterocyclic structures since it has been considered that sulphur is

333   only present in the mixture as thiophenes and their derivatives.

334   Hydrogenation of aromatic rings (HDA), an exothermic and reversible reaction, is

335   thermodynamically limited at typical hydrotreating conditions, so that the reverse reaction of

336   dehydrogenation should also be considered. Korre *et al.* [37] have indicated that no partially

337   hydrogenated ring compounds, such as dihydronaphthalene and hexahydrophenanthrene, were

338   found in the product stream, which shows that hydrogenation proceeds in a ring-by-ring

339   manner. Besides it was detected that the higher number of aromatic rings, the higher reactivity

340   of the molecule. This fact implies that polyaromatic species, such as pyrene, would be the

341   most reactive whereas that the HDA of monoaromatics, such as alkyl benzenes and tetralines,

342   would be the slowest. Other qualitative trend observed during the process was that, for the

343   molecules with more than two aromatic rings, the hydrogenation of the rings in the middle of

344   the molecule goes faster than in the case of the external rings [38,39]. To this, it should be added

345   the fact that the present of alkyl substituents and of cyclohexane rings increases the

346   hydrogenation rate [37].

347   HDS is an exothermic and basically irreversible reaction at the hydrotreating conditions [40]

348   through which sulphur is removed from the petroleum cuts in form of hydrogen sulphide. The

349   removal of this element is one of the main aims of hydrotreating process due to its catalyst

350   poisoning and pollution potential [11]. The HDS reactivity of sulphureted species closely

351   depends on their structure. According with this, non-heterocyclic compounds present a very

352   higher reactivity at hydrotreating conditions than heterocyclic species, such as thiophenes

353   family and derivatives. Different reactivities are also found within this last compounds family,

354   in which thiophenes (Ts) and benzothiophenes (BTs) show a higher reactivity than

355   dibenzothiophenes (DBTs) [41]. The hydrotreating rates of the heterocyclic species are also

356   influenced by the presence of substituent groups in the positions adjacent to the sulphur

357   atom [9]. The desulphurisation process in these species goes through two different pathways:

358   (1) the direct removal of the sulphur atom without ring hydrogenation (hydrogenolysis

359   pathway) or (2) the saturation of the thiophene and the fused benzene rings prior to the

360   removal of the sulphur atom from the thiophene ring (hydrogenation pathway). The HDS of

361   Ts, BTs and unsubstituted DBTs mostly follows the hydrogenolysis pathway whereas

362   hydrogenation becomes the preferred pathway for DBTs at the same time that the degree of

363   substitution increases [40,41].

364   Taking into account the above information, the reaction types reflected in Table 3 has been

365   considered during the stochastic modelling of LCO hydrotreating. A set of reaction rules is

366   also established in order to identify the reaction events from the structure of the molecule.

367   Hydrogenation of aromatic rings and dehydrogenation of saturated rings have been included

368   in the current model as ring-by-ring reactions while it have been considered that sulphur is

369   removed from thiophene rings as H_2S. Besides, the rate constants of the HDT reactions must

370   be specified.

371   In the case of hydrogenation and dehydrogenation reactions, their reaction rates can be

372   calculated from Eq. (10) and (11), respectively:

$$R_{Hydro} = c_{Hydro} \cdot p_{H_2} \qquad (10)$$

$$374 \qquad R_{Dehydro} = c_{Dehydro} = \frac{c_{Hydro}}{K_{eq}} \qquad (11)$$

375   where c_{Hydro} and $c_{Dehydro}$ are the stochastic reaction constants for hydrogenation and

376   dehydrogenation (s^{-1}), K_{eq} is the equilibrium constant (mol·(mol·atm)$^{-1}$), and p_{H_2} is the partial

377   pressure of hydrogen (atm).

378   The expressions for the calculation of hydrogenation rate constants and the equilibrium

379   constants have been developed in previous studies [17], starting from the quantitative

380   structure/reactivity correlations (QS/RCs) proposed by Korre *et al.*[37]. They are given in Eq.

381   (12) and (13), respectively:

$$
\begin{aligned}
382 \qquad \ln(c_{Hydro}) \;=\; & -4.65 - 7.06 \cdot n_H^{1.18} - 1.08 \cdot \left| \frac{\Delta H_R^0}{R} \right| \\
& +0.616 \cdot N_{AR} + 0.330 \cdot N_{SR} - 2.80 \cdot N_{TR} \qquad (12) \\
& -\frac{E_a}{R} \cdot \left(\frac{1}{T} - \frac{1}{T_{Ref}} \right)
\end{aligned}
$$

$$
\begin{aligned}
383 \qquad \ln(K_{eq}) \;=\; & 2.952 - 13.215 \cdot n_H + 5.196 \cdot 10^{-3} \cdot \left| \frac{\Delta H_R^0}{R} \right| \\
& -0.784 \cdot N_{SR} \qquad (13) \\
& -\frac{\Delta H_R}{R} \cdot \left(\frac{1}{T} - \frac{1}{T_{Ref}} \right)
\end{aligned}
$$

384   where n_H is the stoichiometric coefficient for molecular hydrogen, ΔH_R^0 is the heat of

385   reaction at 25°C (J·mol^{-1}), N_{AR} is the number of aromatic rings in the molecule, N_{SR} is the

386   number of saturated rings, N_{TR} is the number of thiophenes fused to the aromatic ring, E_a is

387   the activation energy (J·mol^{-1}), R is the ideal gas constant (J·(K·mol)$^{-1}$), T is the temperature of

388   the system (K), and T_{ref} is the reference temperature (K).

389   The values of the coefficients in the K_{eq} expression proposed by Korre *et al.* [37] have been

390   modified in order to obtain a better agreement between predicted and experimental

391   equilibrium data. The experimental data of a set of 21 compounds, including anthracene,

392   phenanthrene and benzene, have been used to determine the new constants. For the

393   hydrogenation rate constant c_{Hydro}, the N_{TR} term was added to reflect the effect of the presence

394   of thiophene rings on the hydrogenation of aromatic rings, as observed in

395   hydrodesulphurisation kinetics [40]. An exponential coefficient was also added to the n_H term to

396   improve the prediction of the kinetics of benzoic compounds. The hydrogen exponential

397   coefficient, the coefficient for the thiophene correction and the independent term in the

398   expression for k_{SR} were determined by minimising the difference between the predicted and

399 experimental data for the gas oil LCO1 (Table 4) during its hydrotreating at 300 °C (Figure
400 6.a).

401 A very good agreement between experimental and predicted data was obtained for
402 monoaromatic compounds. Besides, the fast hydrogenation of the aromatic compounds with
403 more than two aromatic rings was also properly reflected but some differences could still be
404 observed for saturates and diaromatic compounds. Despite they followed the same tendency
405 than the predicted data, the experimental data showed a larger removal of diaromatic
406 compounds and also a higher concentration of saturated compounds. The introduction of
407 hydrogen exponential coefficient may be the cause for this deviation since, although it
408 provides a better prediction of benzoic compounds kinetics, it worsens the results for other
409 molecules, such as naphthalenes.

410 HDS reactions have been classified in five families according to the structure of the reacting
411 molecule and a rate parameter was assigned to each one (Table 4). The reaction rates are the
412 component directly given by the corresponding rate constant following Eq. (15):

$$R_{HDS} = c_{HDS} \qquad (15)$$

414 The rate constants c_{HDS} were obtained from the fit to the experimental data of the LCO gas oil
415 which was previously used to determine the coefficients in Eq. (15) (Figure 6.b). In this case,
416 the predicted curves perfectly fit to experimental data of both BT and DBT compounds.

417 In order to validate the current modelling methodology, the hydrotreating of the three LCO
418 gas oils in Table 4 has been simulated at different conditions of temperature and partial H_2
419 pressure. These LCO gas oils present different composition profiles. LCO1 and LCO2 both
420 present a significant content in sulphur aromatics (mainly DBT in LCO1 and BT in LCO2)
421 while, in the case of LCO3, the content in these compounds is very lower (less than 1.5% wt)
422 and the addition of sulphides such as dimethyl disulphide (DMDS) is required in order to
423 reach an adequate partial pressure of H_2S over the process. Besides, LCO3 is mainly
424 composed by saturated compounds (almost 60% wt) unlike the other two gas oils, in which
425 diaromatics are the most abundant compounds.

426 The predicted results after 1 h of operation have been compared with analytical data in the
427 parity plots depicted in Figure 7. A good agreement with experimental data was found for
428 saturated and monoaromatic compounds with an average relative error lower than 14%. The
429 prediction of the species with three or more aromatic rings was accurate in most of the

430 experimental points and a very good fit was also found for sulphur aromatic compounds.
431 Performance variations due to modifications in operational conditions, i.e. temperature and
432 hydrogen partial pressure, were contemplated and well-reflected on the current model. On the
433 other hand, as in the case of LCO1, some differences were also detected for diaromatic
434 compounds in LCO3. High complexity of mixtures, which leads to the use of correlations to
435 determine the reactivity of each molecule in hydrogenation/dehydrogenation processes,
436 implies a greater difficulty to accurately predict the content in all aromatic species, although a
437 good global agreement between predicted and experimental data was reached.

438 **5. Conclusions**

439 A stochastic two-steps procedure has been described and validated on the hydrotreating of

440 three LCO gas oils at different operational conditions. In the first step, a molecular

441 reconstruction algorithm is used to represent the feedstock by means of a set of molecules

442 which mixture properties are similar to those of feedstock. This algorithm is formed by the

443 combination of two methods: stochastic reconstruction (SR) and reconstruction by entropy

444 maximization (REM). The SR method leads to obtain a set of molecules typical of a given

445 feedstock while the molar fraction of molecules is adjusted using the REM method in order to

446 improve the predicted mixture properties. Then, the generated set of molecules is used as

447 input in the second step concerning the molecule-based kinetic modelling of the hydrotreating

448 process and in which a kinetic Monte Carlo (kMC) method is applied. The expressions for the

449 calculation of adsorption, equilibrium and hydrogenation/dehydrogenation rate constants have

450 been obtained by applying the quantitative structure/reactivity correlations (QS/RCs).

451 The current stochastic methodology has demonstrated to be a valid tool for the reconstruction

452 of LCO gas oils by combining their available analytical data. This procedure can be used not

453 only to get the molecular information that is not provided by analytical data but also to obtain

454 an accurate molecular representation which can be used as input in kinetic models. Monte

455 Carlo techniques have also given good results in their application on the kinetic simulation of

456 the hydrotreating of LCO gas oils. Moreover, the simulation of the reactions does not require

457 a pre-defined kinetic network due to it is generated as the reactions proceed. Besides, the use

458 of the corrected QS/RCs led to a good prediction of molecular reactivities in complex

459 mixtures of hydrocarbons such as LCO gas oils.

460

461

References

1. Quann, R. & Jaffe, S. Structure-Oriented Lumping: Describing the Chemistry of Complex Hydrocarbon Mixtures. *Ind. Eng. Chem. Res.* **31,** 2483–2497 (1992).

2. Quann, R. & Jaffe, S. Building useful models of complex reaction systems in petroleum refining. *Chem. Eng. Sci.* **5,** 1615–1635 (1996).

3. Vynckier, E. & Froment, G. in *Kinet. Thermodyn. Lumping Multicomponent Mix.* (Astarita, G. & Sandler, S.) 131–161 (Elsevier B.V., 1991).

4. Klein, M. T., Hou, G., Bertolacini, R. J., Broadbeit, L. J. & Kumar, A. *Molecular modeling in heavy hydrocarbon conversions*. (CRC Press Taylor & Francis Group, 2006).

5. Weekman, V. W. Lumps, models, and kinetics in practice. *AIChE Monogr. Ser.* **75,** 1–29 (1979).

6. Sanchez, S., Rodriguez, M. A. & Ancheyta, J. Kinetic Model for Moderate Hydrocracking of Heavy Oils. *Ind. Eng. Chem. Res.* **44,** 9409–9413 (2005).

7. Verstraete, J. J., Le Lannic, K. & Guibard, I. Modeling fixed-bed residue hydrotreating processes. *Chem. Eng. Sci.* **62,** 5402–5408 (2007).

8. López García, C., Hudebine, D., Schweitzer, J.-M., Verstraete, J. J. & Ferré, D. In-depth modeling of gas oil hydrotreating: From feedstock reconstruction to reactor stability analysis. *Catal. Today* **150,** 279–299 (2010).

9. Pereira de Oliveira, L., Verstraete, J. J. & Kolb, M. A Monte Carlo modeling methodology for the simulation of hydrotreating processes. *Chem. Eng. J.* **207-208,** 94–102 (2012).

10. De Oliveira, L. P., Verstraete, J. J. & Kolb, M. Molecule-based kinetic modeling by Monte Carlo methods for heavy petroleum conversion. *Sci. China Chem.* **56,** 1608–1622 (2013).

488   11.   De Oliveira, L. P., Verstraete, J. J. & Kolb, M. Simulating vacuum residue
489   hydroconversion by means of Monte-Carlo techniques. *Catal. Today* **220-222,** 208–220
490   (2014).

491   12.   Gillespie, D. T. A general method for numerically simulating the stochastic time
492   evolution of coupled chemical reactions. *J. Comput. Phys.* **22,** 403–434 (1976).

493   13.   Liguras, D. K. & Allen, D. T. Structural models for catalytic cracking. 2. Reactions of
494   simulated oil mixtures. *Ind. Eng. Chem. Res.* **28,** 674–683 (1989).

495   14.   Neurock, M., Nigam, A., Trauth, D. & Klein, M. T. Molecular representation of
496   complex hydrocarbon feedstocks through efficient characterization and stochastic
497   algorithms. *Chem. Eng. Sci.* **49,** 4153–4177 (1994).

498   15.   Hudebine, D. Reconstruction moléculaire de coupes pétrolières. PhD thesis. (2003).

499   16.   Hudebine, D. & Verstraete, J. J. Reconstruction of Petroleum Feedstocks by Entropy
500   Maximization. Application to FCC Gasolines. *Oil Gas Sci. Technol. – Rev. d'IFP*
501   *Energies Nouv.* **66,** 437–460 (2011).

502   17.   Pereira de Oliveira, L. Développement d'une méthodologie de modélisation cinétique
503   de procédés de raffinage traitant des charges lourdes. PhD thesis. (2013).

504   18.   Oliveira, L. P. De, Vazquez, A. T., Verstraete, J. J. & Kolb, M. Molecular
505   Reconstruction of Petroleum Fractions : Application to Vacuum Residues from
506   Different Origins. *Energy & Fuels* **27,** 3622–3641 (2013).

507   19.   Trauth, D. M., Stark, S. M., Petti, T. F., Neurock, M. & Klein, M. T. Representation of
508   the Molecular Structure of Petroleum Resid through Characterization and Monte Carlo
509   Modeling. *Energy & Fuels* **8,** 576–580 (1994).

510   20.   Trauth, D. M. Structure of complex mixtures through characterization, reaction and
511   modelling. PhD thesis. (1993).

512   21.   Hudebine, D. & Verstraete, J. J. Molecular reconstruction of LCO gasoils from overall
513   petroleum analyses. *Chem. Eng. Sci.* **59,** 4755–4763 (2004).

514   22.   Verstraete, J. J., Revellin, N. & Dulot, H. Molecular reconstruction of vacuum gasoils.
515   *Prepr. Pap. Chem. Soc. Div. fuel Chem.* **49,** 20–21 (2004).

516   23.   Verstraete, J. J., Schnongs, P., Dulot, H. & Hudebine, D. Molecular reconstruction of
517   heavy petroleum residue fractions. *Chem. Eng. Sci.* **65,** 304–312 (2010).

518   24.   Schnongs, P. *Reconstruction moléculaire de coupes pétrolières lourdes*. (2005).

519   25.   Shannon, C. E. A Mathematical Theory of Communication. *bell Syst. Tech. J.* **27,** 379–
520   423, 623–656 (1948).

521   26.   Gillespie, D. T. A rigorous derivation of the chemical master equation. *Phys. A Stat.*
522   *Mech. its Appl.* **188,** 404–425 (1992).

523   27.   Gillespie, D. T. Stochastic simulation of chemical kinetics. *Annu. Rev. Phys. Chem.* **58,**
524   35–55 (2007).

525   28.   López García, C., Becchi, M., Grenier-Loustalot, M. F., Païsse, O. & Szymanskit, R.
526   Analysis of aromatic sulfur compounds in gas oils using GC with sulfur
527   chemiluminescence detection and high-resolution MS. *Anal. Chem.* **74,** 3849–57
528   (2002).

529   29.   Hudebine, D. & Verstraete, J. J. Molecular reconstruction of LCO gas oils from overall
530   petroleum analyses. *Chem. Eng. Sci.* **59,** 4755–4763 (2004).

531   30.   Cooper, B. H. & Donnis, B. B. L. Aromatic saturation of distillates : an overview. *Appl.*
532   *Catal. A Gen.* **137,** 203–223 (1996).

533   31.   Castex, H., Boulet, R., Juguin, J. & Lepinasse, A. Analyse des kérosènes et des gas oils
534   moyens par spectrométrie de masse à moyenne résolution. *Oil Gas Sci. Technol.* **38,**
535   523–532 (1983).

536   32.   Fafet, A., Bonnard, J. & Prigent, F. New Developments in Mass Spectrometry for
537   Group-Type Analysis of Petroleum Cuts. *Oil Gas Sci. Technol.* **54,** 429–452 (1999).

538   33.   ASTM D2887. in *Annu. B. ASTM Stand.* (2007).

539   34.   API 282.1. in *API Tech. Handb.* (1987).

540   35.   Kennedy, J. & Eberhart, R. C. *Swarm intelligence*. (Morgan Kaufmann Publishers,
541   2001).

542   36.   Lopez Garcia, C. Analyse de la réactive des composés soufrés dans les coupes
543   pétrolières: cinétique et modélisation de l'hydrotraitement. PhD thesis. (2000).

544   37.   Korre, S., Neurock, M., Klein, M. T. & Quann, R. J. Hydrogenation of polynuclear
545   aromatic hydrocarbons.2. Quantitative Structure/Reactivity Correlations. *Chem. Eng.*
546   *Sci.* **49,** 4191–4210 (1995).

547   38.   Stanislaus, A. & Cooper, B. H. Aromatic hydrogenation catalysis: a review. *Catal. Rev.*
548   **36,** 75–123 (1994).

549   39.   Korret, S. C., Klein, M. T. & Quann, R. J. Polynuclear Aromatic Hydrocarbons
550   Hydrogenation . 1 . Experimental Reaction Pathways and Kinetics. *Ind. Eng. Chem.*
551   *Res.* **34,** 101–117 (1995).

552   40.   Girgis, M. J. & Gates, B. C. Reactivities, Reaction Networks, and Kinetics in High-
553   Pressure Catalytic Hydroprocessing. *Ind. Eng. Chem. Res.* 2021–2058 (1991).

554   41.   Whitehurst, D. D., Isoda, T. & Mochida, I. Present State of the Art and Future
555   Challenges in the Hydrodesulfurization of Polyaromatic Sulfur Compounds. *Adv.*
556   *Catal.* **42,** 345–471 (1998).

557

558

559

560

561

562

563

Figure captions

565   **Figure 1.** Flowchart of the SR-REM algorithm

566   **Figure 2.** Flowchart of the kMC method

567   **Figure 3.** Building diagram for LCO gas oils

568   **Figure 4.** Evolution as a function of the number of molecules in the predicted mixture of **(a)**
569   the average value of the objective function (white diamonds) and the maximum (black
570   triangles) and minimum (black squares) values **(b)** the CPU time required (white diamonds)
571   and the standard deviation of the objective function (black crosses)

572   **Figure 5.** Parity plots for the predicted properties of eleven gas oils after SR-REM
573   application: **(a)** density (g/cm^3), **(b)** molecular weight (g/mol), **(c)** ^{13}C NMR signatures (%
574   mole): primary saturated CH$_3$ (black diamonds), secondary saturated CH$_2$ (white squares),
575   tertiary saturated CH (white rounds), aromatic CH (black crosses), condensed aromatic C
576   (white triangles) and substituted aromatic C (black rounds), and **(d)** ^{1}H NMR signatures (%
577   mole): diaromatic H (black diamonds), monoaromatic H (white squares), H alpha (white
578   rounds), H beta (black crosses) and H gamma (white triangles)

579   **Figure 6.** Comparison between experimental (points) and predicted values (lines) for LCO1:
580   **(a)** for aromatic families: saturates (black diamonds and black continuous line),
581   monoaromatics (black squares and black discontinuous line), diaromatics (white squares and
582   grey discontinuous line) and compounds with more than two aromatic rings (white diamonds
583   and grey continuous line); and **(b)** for sulphur families: benzothiophenes (black squares and
584   continuous line) and dibenzothiophenes (white squares and discontinuous line). All data are
585   expressed as weight fraction (%)

586   **Figure 7.** Parity plots for the compounds concentration predicted by means of kMC method
587   after 1 h of hydrotreating: **(a)** saturates, **(b)** monoaromatics, **(c)** diaromatics, **(d)** compounds
588   with more than two aromatic rings and **(e)** sulphur compounds: benzothiophenes (black
589   diamonds) and dibenzothiophenes (white rounds). In the case of benzothiophenes all points
590   overlap each other due to their similar values close to point (0.0, 0.0)

591

Table 1. Structural attributes of LCO gas oil molecules

	Structural attribute	Values	Distribution	Parameter names
1	Type of molecule	0 or 1	Histogram	0
2	Number of benzene rings per core	1, …, 4	Gamma	1 and 2
3	Number of naphthenic rings per core	0, …, 4	Gamma	3 and 4
4	Number of thiophenes per core	0 or 1	Histogram	5
5	Acceptance probability for aromatic carbon CH	0 or 1	Histogram	6
6	Acceptance probability for naphthenic carbon CH_2	0 or 1	Histogram	7
7	Type of the side chains	0, 1 or 2	Histogram	8 and 9
8	Length of the paraffinic chains	> 0	Gamma	10 and 11

Table 2. Comparison between experimental and calculated properties after the SR and REM steps

		Experimental	Simulated (after SR)	Simulated (after SR-REM)
Elemental analysis (%wt)	Carbon	89.95	89.85	89.99
	Hydrogen	9.21	9.33	9.15
	Sulphur	0.84	0.82	0.86
Simulated distillation (°C)	Initial boiling point	143	143	143
	5%	214	175	211
	10%	231	211	232
	20%	250	250	250
	30%	259	268	267
	40%	273	288	270
	50%	284	314	284
	60%	299	334	299
	70%	314	351	315
	80%	330	369	334
	90%	349	385	355
	95%	363	393	383
	Final boiling point	398	403	403
Mass spectrometry (%wt)	Saturates	11.00	10.49	11.12
	Monoaromatics	12.90	17.07	12.98
	Diaromatics	58.40	49.25	58.36
	Triaromatics	8.50	16.40	8.48
	Tetraaromatics	3.00	1.53	2.97
	Benzothiophenes	4.00	1.39	3.96
	Dibenzothiophenes	2.20	3.86	2.13
C NMR (%wt)	Aromatic CH	39.10	34.23	36.62
	Fused aromatic C	8.00	12.17	12.03
	Substituted aromatic C	15.30	14.69	14.95
	Saturated CH_3	15.20	15.62	14.96
	Saturated CH_2	19.60	19.35	18.19
	Saturated CH	2.70	3.93	3.25
H NMR (%wt)	Diaromatic H	25.70	24.26	27.64
	Monoaromatic H	2.30	3.40	2.59
	Saturated alpha H	33.90	27.39	27.61
	Saturated beta H	27.30	29.33	27.80
	Saturated gamma H	10.80	15.61	14.37
Others	Molecular weight (g/mol)	192	194	184
	Density 20°C (g/cm^3)	0.9701	0.9710	0.9729

Table 3. Reaction types and their parameters (*c*)
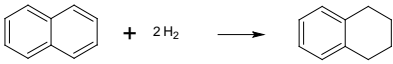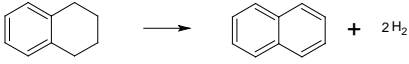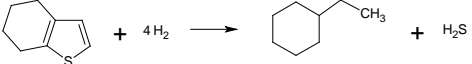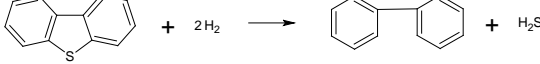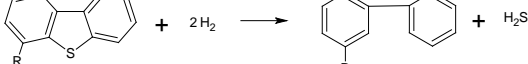
Type of reaction	Reaction	c (s^{-1})
Hydrogenation		Eq. (10)
Dehydrogenation		Eq. (11)
Hydrodesulphurisation	Hydrogenolysis of thiophene and their derivatives	138.90
	Hydrogenolysis of BT and their derivatives	55.60
	Hydrogenolysis of DBT and their derivatives	19.40
	Hydrogenolysis of 4-alkyl-DBT and their derivatives	0.56
	Hydrogenolysis of 4,6-dialkyl-DBT and their derivatives	0.06

Table 4. Hydrotreating conditions (partial H_2 pressure and temperature) and properties from SR-REM algorithm for the three LCO gas oils used as input in the kMC method

	LCO1	**LCO2**	**LCO3**
Hydrotreating conditions			
pH_2 *(atm)*	69.1	95.1	46.1
T (°C)	300/330	300/330	300/310/320
Composition			
C (%wt)	88.01	88.57	87.23
H (%wt)	10.51	9.74	12.57
S (%wt)	1.48	1.69	1.45
Saturates (%wt)	28.58	19.85	59.84
Monoaromatics (%wt)	16.14	22.23	9.55
Diaromatics (%wt)	34.24	38.96	26.20
Triaromatics and more (%wt)	11.49	8.95	3.19
BT (%wt)	2.70	6.39	0.67
DBT (%wt)	6.86	3.61	0.56

Figure 1.
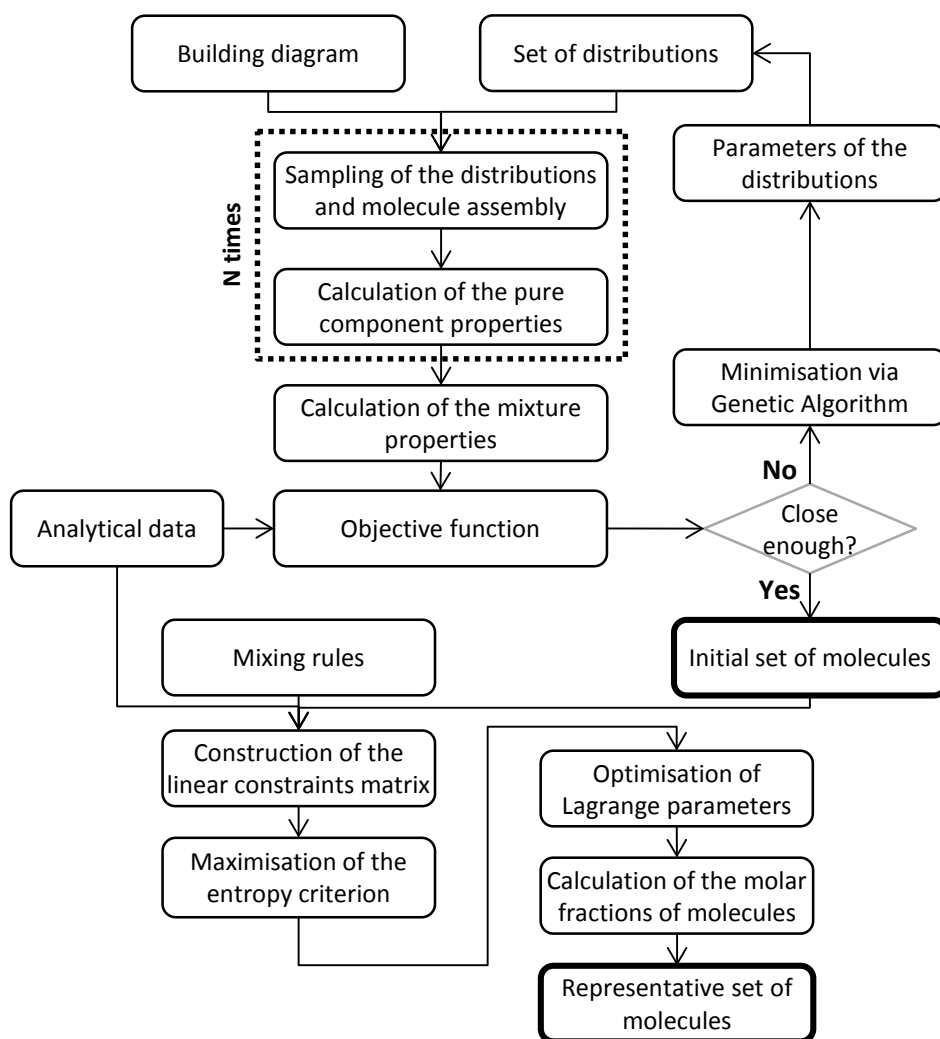
Figure 2.
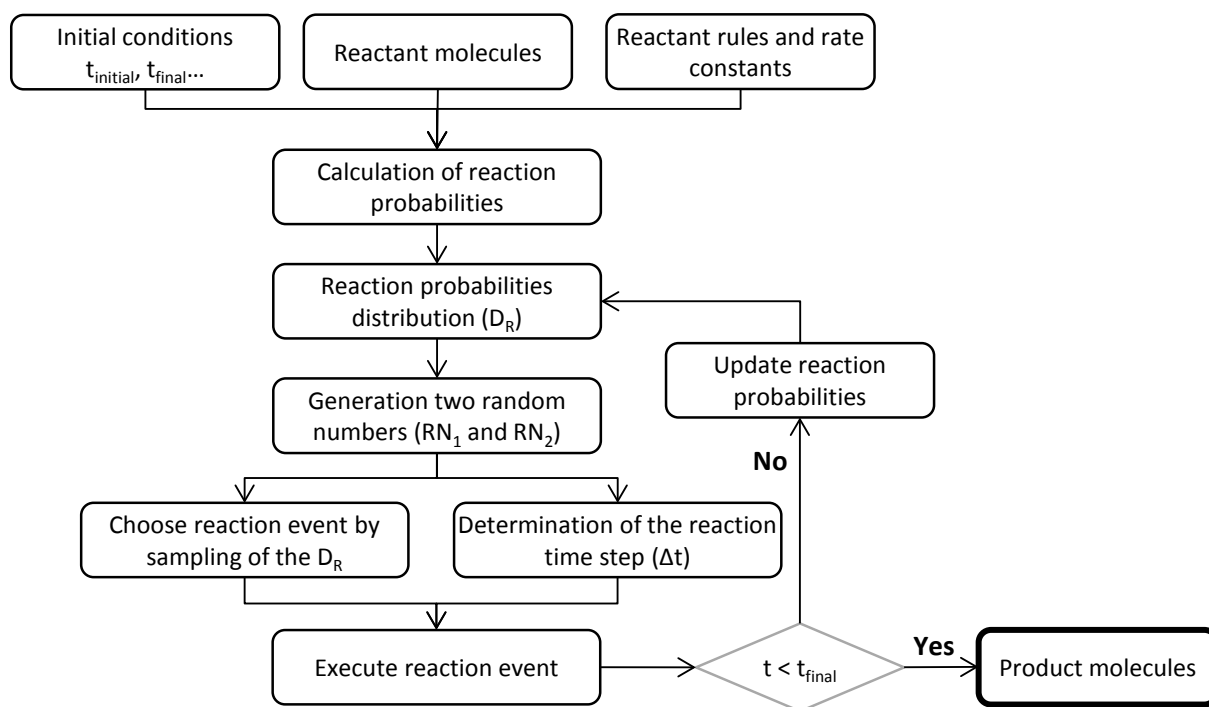
Figure 3.
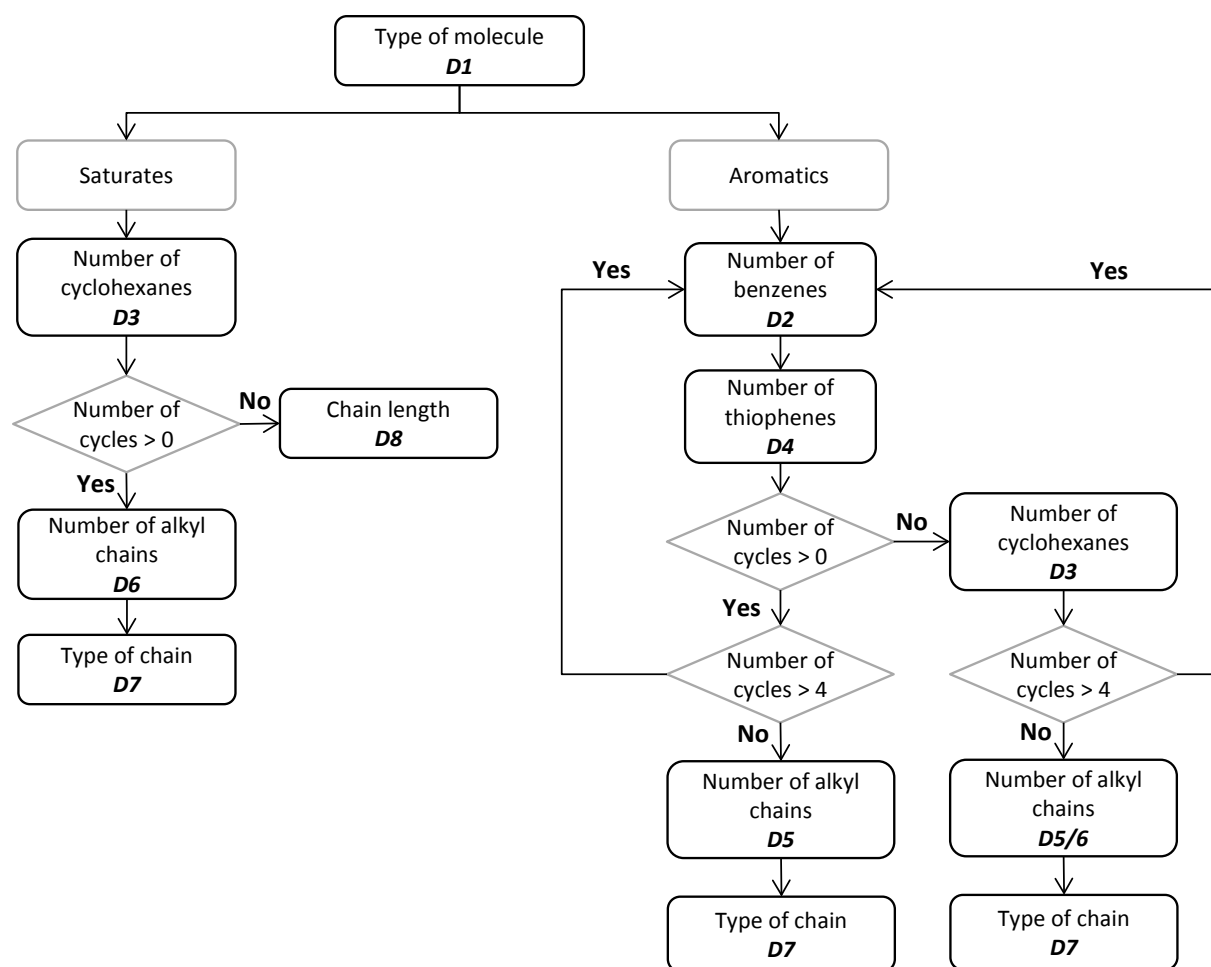
Figure 4.
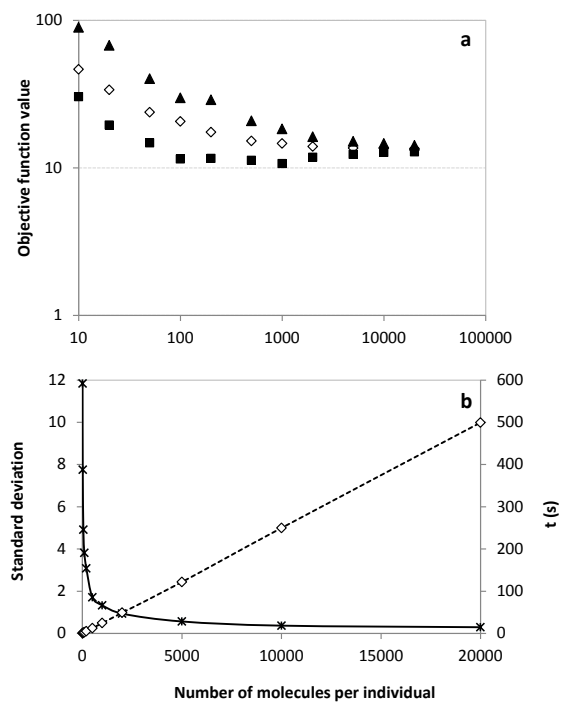
Figure 5.

Figure 6.

Figure 7.