



HAL
open science

Fuel Sorption into Polymers : Experimental and Machine Learning Studies

Benoit Creton, Benjamin Veyrat, Marie-Hélène Klopffer

► **To cite this version:**

Benoit Creton, Benjamin Veyrat, Marie-Hélène Klopffer. Fuel Sorption into Polymers : Experimental and Machine Learning Studies. Fluid Phase Equilibria, Elsevier, 2022, 556, pp.113403. 10.1016/j.fluid.2022.113403 . hal-03585804

HAL Id: hal-03585804

<https://hal-ifp.archives-ouvertes.fr/hal-03585804>

Submitted on 23 Feb 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Fuel sorption into polymers: experimental and machine learning studies.

Benoit Creton*, Benjamin Veyrat, Marie-Hélène Klopffer

IFP Energies nouvelles, 1 et 4 avenue de Bois-Préau, 92852 Rueil-Malmaison, France.

Abstract

In the automotive industry, the introduction of alternative fuels in the market or even the consideration of new fluids such as lubricants requires continuous efforts in research and development to predict and evaluate impacts on materials (e.g., polymers) in contact with these fluids. We address here the compatibility between polymers and fluids by means of both experimental and modelling techniques. Three polymers were considered: a nitrile butadiene rubber (NBR), a fluorinated elastomer (FKM) and a fluorosilicon rubber (FVMQ), and a series of hydrocarbons mixtures were formulated to study the swelling of the polymers. The swelling of samples has been investigated in terms of weight and not volume variations as the measure of this former is assumed to be more accurate. Multi-gene genetic programming (MGGP) was applied to experimental data obtained in order to derive models to predict: (i) the maximum value of the mass gain (ΔM) and (ii) the sorption kinetics, *i.e.* the time evolution of ΔM . Predicted values are in excellent agreement with experimental data (with R^2 greater than 0.99), and models have demonstrated their predictive capabilities when applied to external fluids (not considered during the training procedure). Combining experiments and modelling, as proposed in this work, leads to accurate models which drastically reduce the time necessary to quantify polymeric materials compatibility with a fluid candidates as compared to experiments.

Keywords: Polymer, Fuels, Machine Learning, Sorption

*Corresponding author: benoit.creton@ifpen.fr

1. Introduction

In the context of global warming, conclusions of research dedicated to the reduction of greenhouse gases emissions advocate the use of alternative fuels [1]. In particular, advanced fuels and biofuels including conventional renewable fuels, respecting environmental criteria at a reasonable cost are of primary interest [2]. Biofuels are issued from organic raw materials and they can be seen as blends of renewable molecules such as normal- and iso-paraffins, naphthenic and aromatic compounds, normal- and iso-olefins, alcohols, and/or esters [3]. Normal- and iso-paraffins can be obtained by industrial processes such as Fisher-Tropsch (FT) [4] and hydrotreatment of vegetable oils (HVO) [5]. In the same way, naphthenic and aromatic compounds can be synthesized from the liquefaction or pyrolysis of biomass [6, 7]. Moreover, compositions constantly evolving and being different from one country to another, it has become essential to understand the impact of the introduction of these compounds on the physical properties of alternative fuels. The presence of these families of molecules with different chemistry requires extensive research and development activities. Indeed, it drives the conditions for storage, transportation, and combustion quality.

In combustion engine vehicles many pieces of the fuel-delivery systems are composed all or part of polymers. Polymeric materials in contact with fuels and/or biofuels may be subject to deformations such as swelling, caused by solvent ingress within their structure and leading to strong modifications and loss of their initial physical and chemical properties [8, 9]. To address this problem, one solution consists in using multilayer structures containing interleaved barrier polymers [10]. Up to date, only few works have been published in the literature dealing with the compatibility between polymeric materials and fluids, and there is a lack of available experimental data [11]. The group of Izák *et al.* have investigated experimentally and theoretically gases and liquids sorption into polymers over the last decades [12, 13, 14]; and more recently, Krajakova *et al.* focused on sorption of liquids into poly(ethylene) samples of different densities [15]. Regarding fuels, Haseeb *et al.* immersed some elastomeric materials such as nitrile butadiene rubber (NBR) and Viton® a fluorinated elastomer (FKM) in diesel and palm biodiesel to compare the degradation of physical properties like weight and volume changes, hardness and tensile strength [16, 17]. Kaas *et al.* studied the compatibility of elastomeric materials with gasoline blends containing ethanol and isobutanol, followed evolution of some polymer's properties, and

38 proposed a ranking of elastomer specimens according to their swelling [18].
39 Silva *et al.* ordered some rubbers as a function of their compatibility with
40 biodiesels, and revealed that the mobility of chains of NBR in biodiesel in-
41 creases without change in their chemical structures [19]. In the article by
42 Trakarnpruk *et al.*, authors studied elastomer properties after immersion
43 in biodiesel, focusing among others on NBR, copolymers, and terpolymer
44 FKM. Authors concluded that among tested polymers fluoroelastomers un-
45 dergo fewer physical degradation [20]. Weltshev *et al.* focused their research
46 on the resistance of sealing materials such as FKM, NBR and fluorosilicon
47 rubber (FVMQ), immersed during hours in biodiesel based fuels, as a func-
48 tion of the age and the temperature of fluids [21]. Authors noted that the
49 percentage of degradation is proportional to the temperature and the age
50 of the fuels. In regards to available experimental results, it appears that
51 current methods used for the data acquisitions are time consuming, and the
52 development of robust predictive models is of high relevance.

53 Plota and Masek recently reviewed kinetic based models used to predict
54 the lifetime of polymeric materials and conclude to the necessity of develop-
55 ing new methods [22]. During last decades materials informatics has emerged
56 as a new approach for the conception of new materials [23, 24, 25]. It consists
57 in training learning algorithms on database content, in order to allow pre-
58 dictions for materials having structures similar to those contained within the
59 database, or even to propose promizing candidates for specific applications.
60 Polymer informatics necessitates relevant databases which integrate knowl-
61 edge about properties related to thermodynamics, mechanics, optics, and
62 transport [26, 27]. Litterature reviews report developments of quantitative
63 structure property relationships (QSPR) for polymer properties [28, 29, 30].
64 In the case of transparent polymeric materials, QSPR methods have been
65 used to model optical properties such as the refractive index, n [28, 30, 31, 32].
66 Holder *et al.* have shown that the use of dimeric repeating units for descriptor
67 calculation leads to the most accurate models [33]. Duchowicz *et al.* used a
68 Simplified Molecular Input Line Entry System (SMILES) – not dependent of
69 3D-molecular geometries – based model to predict n for 234 structurally di-
70 verse polymers [34]. Jabeen *et al.* developed a four-descriptor QSPR model
71 with accurate predictions for a highly diverse set of 133 organic polymers
72 [35]. Numerous works reported attempts to predict polymer properties such
73 as glass transition temperature, T_g using QSPR [29, 36, 37, 38, 39]. The
74 knowledge of T_g defines domains of rigid structure or rubber-like properties
75 for polymeric materials, and thus is of utmost importance for many appli-

76 cations. Mercader *et al.* have demonstrated that T_g can be well predicted
77 with QSPR and advocated the use of trimeric moieties for descriptor cal-
78 culation [37]. QSPR were also developed to predict mechanical properties
79 for polymeric materials [40], and Cravero *et al.* proposed QSPR models to
80 estimate tensile strength of polymers [41]. Another possible application of
81 QSPR modelling is to predict sorption of chemicals into polymer matrices.
82 Zhu *et al.* proposed a QSPR based model for the prediction of diffusion
83 coefficients of hydrophobic organic contaminants in low density polyethylene
84 [42]. Li *et al.* proposed models for predicting polymer/brine partition co-
85 efficients for chemicals, with polymers such as polyethylene, polypropylene
86 and polystyrene [43]. Our group has previously proposed QSPR models to
87 predict sorption values for neat compounds and up to quinary mixtures of
88 hydrocarbons, alcohols, and ethers, and demonstrated their applicability to
89 predict sorption values for some alternative fuels into a poly(ethylene) [44].

90 In the present work, we report the acquisition of new experimental sorp-
91 tion values at room temperature for neat compounds and alternative jet fuels
92 based fluids into three polymers. Additionally, we present QSPR based mod-
93 els developed using machine learning methods, and its application to model
94 new experimental data. The paper is organized as follows: we present exper-
95 imental data methods and the strategy followed to build new QSPR based
96 models, new experimental data and the predictive performance of models are
97 then exposed and discussed, and the last section gives our conclusions.

98 2. Materials and methods

99 2.1. Experimental procedure

100 2.1.1. Materials and Samples

101 Three polymers commonly considered for the design of fuel-delivery sys-
102 tems were selected for this study: NBR, FKM, and FVMQ. Polymers raw
103 materials – plane square sheets with 0.3m size and $2 \cdot 10^{-3}$ m thickness –
104 were supplied either by Zodiac Aerotechnics or by Stacem. Rectangular
105 parallelepiped shapes with $60 \times 10 \times 2$ mm³ were extracted from the polymer
106 sheets using a cutting shape, and samples were subsequently used for the
107 sorption tests. Some characteristics (grades for aerospace applications) for
108 these materials are presented in Table 1. We also performed measurements,
109 the Dynamic Mechanical Analysis (DMA) was used to determine the glass
110 transition temperature for the three polymers. A sinusoidal stress was ap-
111 plied to each sample while the strain was measured, allowing one to determine

Table 1: Characteristics for polymeric materials considered in this study.

Polymer	Type	Standards	Hardness (IRHD ^a)	T _g (°C)	Plasticizer (% wt.)
NBR	20B8	NF L 17-120	78	-36	11.0
FVMQ	61D8	NF L 17-261	80	-55	1.2
FKM	60C8	NF L 17-164	80	-1	0.5

^aIRHD: International Rubber Hardness Degree. The dial of the durometer is graduated according to the Shore D scale, from 0 (soft) to 100 (hard) IRHD, with uncertainties associated to measurements of +5/-4 IRHD.

112 the complex modulus and the loss factor. So-obtained T_g values, reported
 113 in Table 1, correspond to the peak value of tan δ, the damping, a measure
 114 of the energy dissipation of a material. Additionally, the Thermal Gravi-
 115 metric Analysis (TGA) was used, it consists in following the mass variations
 116 of a sample with the time as the temperature changes. This measurement
 117 provides information about chemical phenomena including thermal decom-
 118 position but also physical phenomena, such as the desorption of additives.
 119 This technique has allowed the determination of the plasticizer amount ini-
 120 tially present in each studied polymer. It led to plasticizer amounts of 0.5 %
 121 wt. for FKM, 1.2 % wt. for FVMQ, and 11 % wt. for NBR, indicating that
 122 the amount in NBR is not negligible and may lead to measurement artifacts.

123 Fluids under consideration in this study are pure liquids and aviation
 124 fuels. Naphthenic and aromatic hydrocarbons (decaline (labeled D), xylene
 125 (labeled X), tetraline (labeled T), iso-propylbenzene (or cumene, labeled C),
 126 n-propylbenzene (labeled P), and methylnaphtalene (labeled M)), with high
 127 purity grades were purchased from Merck, and no additional purification was
 128 performed. One Jet A-1 (labeled J) being one of the fuels most commonly
 129 used in commercial aviation was selected for sorption measurements. Addi-
 130 tionally, we considered alternative jet fuels approved for certification such as:
 131 Synthetic Paraffinic Kerosene (SPK) and Hydroprocessed Esters and Fatty
 132 Acids (HEFA). For instance, SPK can be FT fuels – composed of normal
 133 and isoparaffins, or Alcohol to Jet Synthetic Paraffinic Kerosene (ATJ-SPK)
 134 created from isobutanol which is derived from feedstocks. HEFA – similar
 135 to HVO – includes hydrocarbon-based jet fuels (100% paraffinic) produced
 136 from animal or vegetable oils by hydroprocessing [45]. The current certifica-
 137 tion for the use of HEFA in mixture with jet fuel allows a maximum of 50%

138 vol. We considered three HEFA with different cold flow properties such as
139 crystallization temperatures: $-50\text{ }^{\circ}\text{C}$, $-30\text{ }^{\circ}\text{C}$, and $-20\text{ }^{\circ}\text{C}$ labeled H(50), H(30),
140 and H(20), respectively. We considered three additional fuels (labeled A1,
141 B1, and C1) to assess their compatibility with the polymers through sorption
142 measurements, and later used to assess the predictive capability of models.
143 A1 is a Jet A-1 fuel, noting that its composition slightly differs from that of J.
144 B1 is an ATJ-SPK mainly composed of i-paraffins. C1 is a jet fuel surrogate
145 with high aromatics content (*ca.* 20 vol%). Note that Hall *et al.* recently
146 considered these conventional and synthetic fuels [46], and representations
147 for these fluids are proposed hereafter.

148 In order to deeply explore effects of the fluid composition on polymer
149 mass variations when the polymer is immersed in a fluid, we defined differ-
150 ent mixtures varying compositions for instance, in terms of naphthenes and
151 aromatics content, paraffins chain length... Mixtures containing J and 25,
152 50, and 75 % vol. of H(50) were formulated. Nine mixtures containing H(50)
153 and amounts 1, 5, and 10 % vol. of X, T, and D were also elaborated. Three
154 blends of 90 % vol. of H(30) and 10 % vol. of C, P, and M were designed. J
155 was mixed with X in 75, and 25 % vol. proportions, and a ternary mixtures
156 containing J, X, and H(30) in equal volumetric proportions was formulated.

157 2.1.2. Sorption measurements

158 The term sorption is commonly used to describe the dissolution of a pen-
159 etrant into a polymer matrix. Measurements of liquid sorption into polymers
160 were performed using a gravimetric method, as detailed in our previous works
161 [44]. Experiments consists in recording the mass variation (weight gain or
162 loss) of a polymer sample with time when immersed into a large excess of
163 the studied liquid. Noting that from sorption values, at equilibrium or sat-
164 uration, it is possible to derive the solubility coefficient, and measurements
165 must be accurately performed as the absorbed quantities are often very small.
166 Rectangular parallelepiped polymer samples were first weighted ($m_{t=0}$) us-
167 ing an analytical balance METTLER TOLEDO (capacity up to 30 g, with
168 a precision of 0.026 g), and then immersed in a large excess of studied liquid
169 in a closed 100 ml glass vessel. Glass vessels were placed at ambient temper-
170 ature ($20 \pm 1\text{ }^{\circ}\text{C}$) in an air-conditioned laboratory, for all the duration of the
171 sorption experiments. Polymer samples were regularly removed from glass
172 vessels, wiped carefully, and weighted (m_t) in order to follow mass variations
173 of each polymer materials in considered liquids. The mass variation (ΔM)
174 is expressed in percent as the ratio between the amount of sorbed fluid and

175 the initial polymer weight, as follows:

$$\Delta M = 100 \times \frac{m_t - m_{t=0}}{m_{t=0}}, \quad (1)$$

176 It has been checked that the repeatability of the sorption values is excel-
177 lent, with less than 2% of variation coefficients. Measurements are performed
178 until the curve ΔM as a function of time reaches its equilibrium value, ex-
179 hibiting a plateau. According to the considered polymer-fluid couple up to 40
180 days were needed to reach the plateau value. We emphasize that the sample
181 swelling has been investigated in terms of weight variations and not volume
182 swellings as this former is assumed to be more accurate. We previously noted
183 the presence of plasticizers in NBR which can cause a weight loss of the sam-
184 ples during sorption tests, and can produce misleading results. Therefore, all
185 NBR samples were pretreated to remove plasticizers, as follows: they have
186 been washed with toluene during 3 weeks (at 50 °C to speed up the diffusion
187 mechanism) and then dried.

188 *2.2. Modelling Method*

189 These last years, we have devoted large efforts in the development of
190 QSPR based models for the prediction of various property values [47]. These
191 approaches aim at identifying non-obvious correlations between property val-
192 ues of the matter and some features rendering information about the matter.
193 Reviews have been published dealing with developments and applications of
194 QSPR based models, and best practices in developing such models [29, 48].

195 *2.2.1. Data Sets*

196 The accuracy of predictive QSPR is related to the quality of data, and
197 thus a keystone of such works is the database used to develop models. The
198 used database contains reference sorption values, ΔM measured following the
199 experimental procedure described above. The database contains 521 sorption
200 values measured at room temperature, for neat compounds and mixtures.
201 Table 2 presents an extract of our database, *i.e.* the maximum amounts of
202 each fluid sorbed into NBR, FVMQ, and FKM. Indeed, the database contains
203 the complete isotherms – evolution with time of the amount of sorbed fluid
204 through NBR –, with in between 20 and 25 data points for each isotherm.

205 During last decades of QSPR model developments, the use of external
206 validation has been shown as necessary to ensure its ability to extrapolate
207 to new fluids, *i.e.* not considered within the database used to train the

Table 2: Maximum amounts of sorbed fluid (ΔM , in %) into NBR, FVMQ, and FKM. Fluids are labeled as follows: X, T, D, C, P, and M stand for xylene, tetraline, decaline, iso-propylbenzene, n-propylbenzene, and methylnaphtalene, respectively ; fluids mixtures are labeled as follows, for instance, H(50)90-T10 contains 90 % vol. of H(50) in mixture with 10 % vol. of tetraline.

Label	Fluid	ΔM (%)		
		NBR	FVMQ	FKM
F01	X	120.0	9.6	6.0
F02	T	97.4	4.8	0.5
F03	D	18.7	3.6	0.0
F04	J	16.8	4.3	0.2
F05	H(20)	3.5	1.5	0.1
F06	H(50)	4.1	1.8	0.0
F07	J75-H(50)25	12.3	3.6	0.1
F08	J50-H(50)50	9.1	3.1	0.1
F09	J25-H(50)75	6.8	2.5	0.1
F10	H(50)99-X01	4.5	1.9	0.1
F11	H(50)95-X05	6.2	2.5	0.1
F12	H(50)90-X10	8.4	3.0	0.2
F13	H(50)99-T01	4.5	1.9	0.0
F14	H(50)95-T05	6.7	2.3	0.1
F15	H(50)90-T10	9.4	2.6	0.1
F16	H(50)99-D01	4.1	1.9	0.0
F17	H(50)95-D05	4.6	2.0	0.0
F18	H(50)90-D10	5.1	2.1	0.0
F19	H(30)90-C10	7.6	2.6	0.1
F20	H(30)90-P10	7.9	2.6	0.1
F21	H(30)90-M10	14.8	3.0	0.2
F22	J75-X25	30.5	6.2	0.8
F23	J25-X75	81.6	8.8	3.7
F24	J33-X33-H(30)33	27.2	5.9	0.8

208 model [49]. Its popular version is the n -fold cross-validation (n -CV) in which
209 the data set is randomly divided in approximately equal n portions. An
210 aggregate of $(n-1)$ portions forms the Training set – used to train models,
211 and the remaining portion constitutes the Test set – used to evaluate model’s
212 performance. We emphasize that no data point belonging to external sets
213 is used to derived models. This procedure is repeated n times choosing at
214 each new fold another portion of data as a Test set. The subject of external
215 validation for QSPR analysis of mixtures has been addressed by Muratov *et*
216 *al.* [50], and the authors-defined ”mixture out” strategy was applied in this
217 study.

218 *2.2.2. Fluids characterisation and representation*

219 A fuel contains thousands of diverse chemicals and its exact composition is
220 never known. The characterization of such complex fluids and identification
221 of representative compounds or surrogates are of utmost importance when
222 developing predictive models for application in the industry [51, 52]. The use
223 of modern analytical instruments such as chromatography, helps in obtaining
224 information about the composition and structure of fluids components. The
225 two-dimensional gas chromatography (labeled GC-2D or GCxGC) has been
226 proved as an interesting analysis technique for detailed characterisation of
227 petroleum products [53]. Fuel candidates considered in this study were ana-
228 lyzed by means of GCxGC, and their compositions expressed as distributions
229 of mass fractions as a function of the number of carbon atoms for hydrocarbon
230 families such as n-paraffins, i-paraffins, naphthenes, aromatics. . . A molecu-
231 lar structure is attributed for each hydrocarbon family/number of carbon
232 atom bin, and each fuel is thus represented by a maximum of 120 molecular
233 structures. Figure 1 presents compositions of fluids A1, B1, C1, J, H(20),
234 H(30), and H(50), simplified to four chemical families: n-paraffins, i-paraffins,
235 naphthenes, and aromatics. It shows that ATJ-SPK (B1) and HEFA fuels
236 are clearly mainly paraffinics, with B1 purely n-paraffinics. A1 and J have
237 similar compositions with for J, slightly (*ca.* 3 %) lower and higher i-paraffins
238 and aromatics contents, respectively. The surrogate C1 is poor in paraffins
239 and rich in naphthenes as compared to other fluids.

240 *2.2.3. Molecular and mixture descriptors*

241 From conclusions drawn in previous studies [2, 44], we chose to solely
242 consider functional group count descriptors (FGCD). Such a simple repre-
243 sentation of compounds has been shown to provide relevant descriptors us-

Table 3: Ranges of number of carbon atoms to represent jet fuel candidates.

Formulae	Family	Number of C atoms
C_nH_{2n+2}	n-paraffins	5 to 20
C_nH_{2n+2}	i-paraffins	5 to 30
C_nH_{2n}	mono-naphthenes	6 to 18
C_nH_{2n-2}	di-naphthenes	9 to 29
C_nH_{2n-4}	tri-naphthenes	13 to 16
C_nH_{2n-6}	mono-aromatics	7 to 17
C_nH_{2n-8}	naphthenes mono-aromatics	9 to 16
C_nH_{2n-10}	naphthenes mono-aromatics	10 to 15
C_nH_{2n-12}	di-aromatics	10 to 16
C_nH_{2n-14}	naphthenes di-aromatics	12 to 16
C_nH_{2n-16}	naphthenes di-aromatics	13 to 15

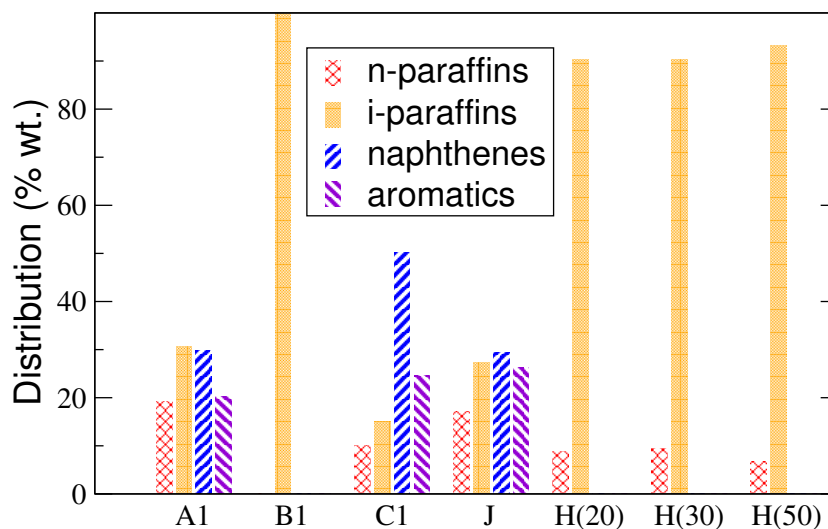


Figure 1: Simplified chemical compositions described in terms of n-paraffins, i-paraffins, naphthenes, and aromatics, for fluids A1, B1, C1, J, H(20), H(30), and H(50).

Table 4: List of the Functional Group Count Descriptors (FGCD) used to describe fluids in the database and associated SMARTS codes or definitions.

Label	SMARTS/Definition	Label	SMARTS/Definition
X1	[H]	X18	[C][CR](![C])(![C])[C]
X2	[C,c]	X19	[C][CR](![C])([C])[C]
X3	[CX4H3]	X20	[C]=[C]([C])![C]
X4	[CX4H2]	X21	[CX3H1]=[CX3H1]
X5	[CX4H1]	X22	[c][CX4H3]
X6	[CX4H0]	X23	[c][CX4H2]
X7	[CX3H1]	X24	[c][CX4H1]
X8	[CX4H2R]	X25	[R]
X9	[CX4H1R]	X26	aromatic_rings
X10	[cX3H1](:*):*	X27	non-aromatic_rings
X11	[cX3H0](:*)(:*)*	X28	aliphatic_rings
X12	[cX3H0](:*)(:*):*	X29	number_of_rings
X13	[cX3H0]-[cX3]	X30	MM
X14	[cX3H0](:*)(:*)(-[CX4H2R])	X31	[C;R]
X15	[CX4H2]-[CX4H1]-[CX4H2]	X32	[c;R]
X16	[C][C](!CX1)(!CX1)![CX1]	X33	C1CCCCC1
X17	![C][C]([C])([C])[C]		

244 able in QSPR procedure [3, 44]. This family of molecular descriptors gathers
 245 some counts of groups identified as relevant under chemical aspects. Table 4
 246 gives the list of FGCD under consideration in this study and labelled from
 247 X1 to X33. For instance, the FGCD labelled X25 denotes the number of
 248 carbon atoms involved in a ring. As Villanueva *et al.* did [44], we have
 249 also computed the molar mass (MM) of neat compounds, this information
 250 being used as an additional descriptor (labelled X30). Simplified molecular
 251 input line entry specification (SMILES) notations were assigned to each neat
 252 compound considered in this study. FGCD were counted using the RDKit’s
 253 SMILES arbitrary target specification (SMARTS) matching functionalities
 254 [54, 55], and SMARTS codes corresponding to FGCD are given in Table 4.

255 The calculation of descriptors for mixtures has been addressed similarly
 256 as in previous works [44, 56]. We assumed mixture descriptors X_{mix} as linear
 257 combinations of pure component descriptors weighted with the associated
 258 molar fractions x_i . This approach has already been shown effective in pre-

259 dicting sorption values for some alternative fuels in a poly(ethylene) [44].
260 For instance, in the case of descriptor X1, the corresponding descriptor for a
261 mixture X1_{mix}, is defined as follows:

$$X1_{mix} = \sum_{i=1}^N x_i \times X1_i, \quad (2)$$

262 where i runs over the N constituents in the mixture.

263 2.2.4. Chemical space representation

264 We preprocessed the data by applying a principal component analysis
265 (PCA) on fluid descriptor values. Figure 2 represents the projections of F01
266 to F24 in the space formed by the three main principal components resulting
267 from the PCA, providing one approximated representation of the chemical
268 space for our database. Some of fluid candidates are at edges of the domain,
269 isolated from all other samples, this is typically the case for fluids F01, F03,
270 F04, and F05. These latter datapoints appear as outliers for the following
271 reasons: (i) F01, xylene, is a pure compound with the highest property
272 value ; (ii) F03, decaline, is a pure compound and has the highest value on
273 PC2 axis ; (iii) F04, Jet-A1, has the highest value on PC3 axis ; (iv) F05,
274 HEFA-20, has the lowest property value. The presence of outliers in external
275 sets during the CV procedure may induce applicability domain violations.
276 Fluids F01, F03, F04, and F05 were fixed, meaning they are placed in a fold
277 always used to form Training sets. We used a 5-CV procedure applied on the
278 remaining 20 fluids candidates – four fluids per fold. The Training and Test
279 sets thus represent 83% and 17% of the database, respectively. In Figure 2,
280 each symbol is filled according to the fold the fluid belongs to.

281 2.2.5. Machine learning algorithm

282 In the frame of past studies [47], we have observed that QSPR models de-
283 rived from Support Vector Machine (SVM) algorithms frequently outperform
284 others evaluated learning algorithms such as neural networks, partial least
285 squares, genetic algorithm. . . Table 2 shows that the number of data points
286 – fluids candidates – is quite limited and the application of SVM does not
287 seem appropriate in this case. We focused on developing multilinear equa-
288 tions which moreover have the advantage to be explicit models and easily
289 implemented in a spreadsheet. Such multilinear models can be, for instance,
290 generated by means of Evolutionary Algorithms (EA) techniques inspired

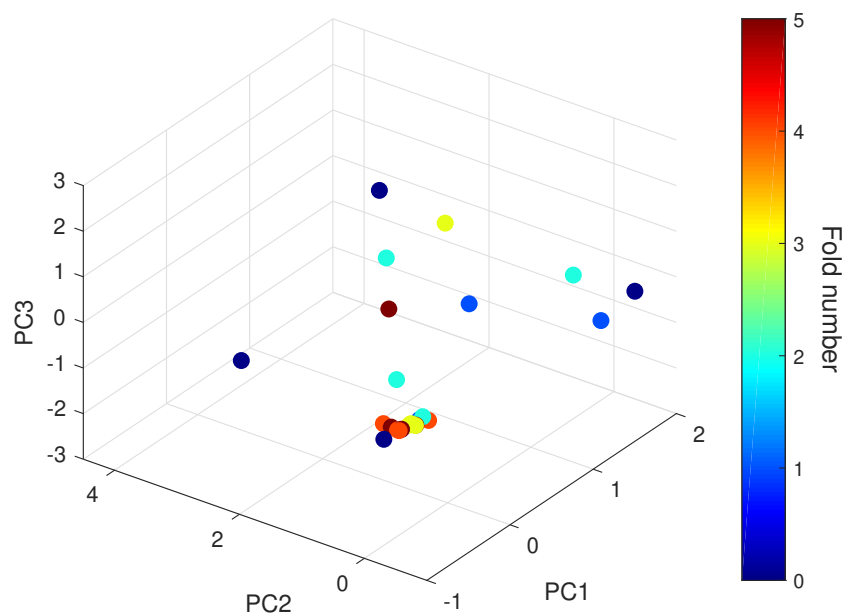


Figure 2: Projections of jet fuel based fluids on the space formed by PC1, PC2 and PC2, the three first principal components resulting from the PCA. Symbols are filled according to a gradient of colors, as legended in the colorbar each of the six folds is represented by one color.

291 from the Darwinian evolution theory of biological species. The application
 292 of EA to regression problems consists in an iterative evolution of a population
 293 of equations initially randomly set. Equations can be summarized under the
 294 general following form:

$$Property = \lambda_0 + \sum_{i=1}^N \lambda_i G_i, \quad (3)$$

295 where λ_0 is the intercept, λ_i denotes a weight associated to the gene i (G_i),
 296 and N is the total number of genes in the model. Each gene consists in a
 297 combination of descriptors (see Table 4) and mathematical functions (see Ta-
 298 ble 5), and can be thought as a tree with nodes and branches (Figure 3). Such
 299 construction allows to catch non-linearity in property variations. Multi-Gene
 300 Genetic Programming (MGGP) was applied to generate models, using the
 301 genetic programming toolbox for the identification of physical systems (GP-
 302 TIPS) coded in the MATLAB environment [57, 58, 59]. The evolution of the
 303 initial population – initial equations – is ensured by survival of fitter individ-
 304 uals, and reproduction of individuals consists in applying crossover as well as
 305 mutation operations to produce child equations. Genetic operators act upon
 306 sub-tree elements, thus making the structure of trees evolve during the itera-
 307 tive procedure. The procedure ends when one of the pre-defined criteria such
 308 as maximum number of generations, best fitness values... is reached. Some
 309 of GPTIPS parameters such as the maximum numbers of genes and nodes
 310 per tree, must be lowered to prevent any overfitting problems. Similarly, the
 311 maximum numbers of generations and runs have to be optimized to ensure
 312 convergence of calculations for reasonable computational resources [60, 61].
 313 Parameters of performed GPTIPS calculations were optimized according to
 314 the procedure defined by Creton *et al.* [60]. Table 5 reports details about
 315 values and/or ranges of investigated GPTIPS settings in this work.

316 Models are evaluated according to their capability in predicting fluids
 317 properties. Predicted values are compared to reference experimental data,
 318 and performances of models are evaluated by means of metrics such as MAE
 319 (Mean Absolute Error, equation (4)), RMSE (Root Mean Squared Error,
 320 equation (5)) or R^2 (Coefficient of determination, equation (6)), defined as
 321 follows:

$$MAE = \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i|, \quad (4)$$

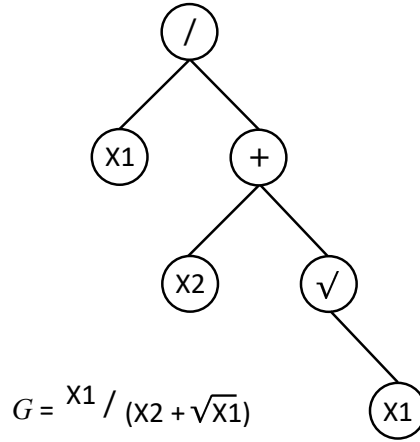


Figure 3: Example of a gene G , and its tree-like architecture as considered in MGGP.

Table 5: Investigated parameter settings for the MGGP based method.

Parameter	Corresponding values
Function set	+, -, ×, ÷, √, exp, ln
Population size	250
Number of runs	1, 5, 10, 15, 20, 25, 30, 40
Tournament size	25
Maximum tree depth	4
Number of generations	100, 500, 1000, 2000
Maximum number of genes	1 to 5
Maximum number of nodes per tree	1 to 8
Mutation events	0.1
Crossover events	0.85
Reproduction events	0.05

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2}, \quad (5)$$

$$R^2 = \frac{\sum_{i=1}^N (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^N (y_i - \bar{y})^2}, \quad (6)$$

322 where in equations (4) to (6), \hat{y}_i stands for the predicted value, y_i represents
 323 the experimental value, \bar{y} denotes the mean property value calculated on
 324 experimental data set candidates, and N is the number of data.

325 3. Results and discussion

326 3.1. Experimental results

327 We performed sorption experiments to evaluate polymers (*i.e.*, NBR,
 328 FVMQ, and FKM) compatibility with a series of hydrocarbons mixtures and
 329 more specially, to mixtures containing different amounts and types of aro-
 330 matics. Details about tested fluids mixtures are given in the Table 2. The
 331 weight variation of the polymer is very dependent on the considered system,
 332 the chemical compositions of both the polymer and the fluid. The measured
 333 maximum uptakes of hydrocarbons in each polymer (or maximum ΔM) are
 334 presented in Table 2. From tested polymer/fluid couples, ΔM values range
 335 from *ca.* 0% (FKM immersed in HEFA) to 120% (NBR immersed in xylene).
 336 Clearly, none of tested hydrocarbons is significantly absorbed into FKM, and
 337 from measured values, FKM can be assumed as a barrier polymer in case of
 338 hydrocarboned fluids. ΔM values obtained for the FVMQ polymer are neg-
 339 ligible as compared to those of NBR. Within the fluids matrix (Table 2), we
 340 considered mixtures having from about 0% vol. aromatics (*e.g.*, HEFA) to
 341 100% vol. aromatics (*e.g.*, xylene). In Figure 4, we plot the ΔM plateau
 342 value for NBR as a function of mono-aromatics content in the fluid tested.
 343 Considering all these systems, deplasticized NBR presents a higher level of
 344 sorption and is very sensitive to the aromatics content of the fluids. Figure 4
 345 shows that mixtures rich in paraffins such as HEFA fuels (left part of the di-
 346 agram) lead to low ΔM values as compared to mono-aromatics rich mixtures
 347 (right part of the diagram). From these elements a quadratic (of the % vol.
 348 of mono-aromatics) trend could describe the observed behavior. Noting that
 349 fluids containing polyaromatic compounds, *e.g.* the fluid F21 containing 10%
 350 vol. methylnaphtalene, deviate from this trend with upper ΔM values.

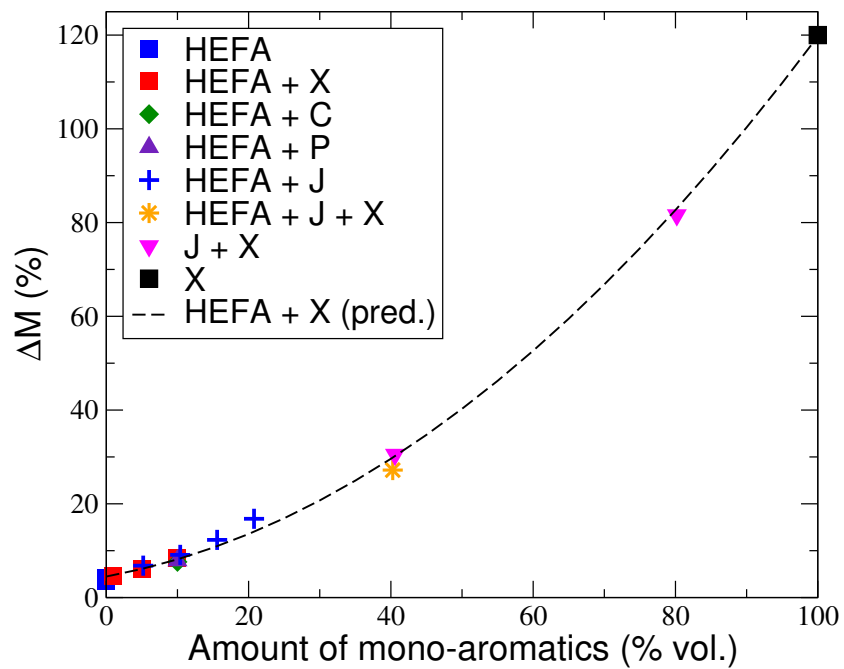


Figure 4: Evolution of the ΔM plateau value for NBR, with the mono-aromatics volumetric percent of the fluid. Fluids are labeled as in Table 2: X, C, P, and J stand for xylene, iso-propylbenzene, n-propylbenzene, and Jet-A1, respectively. The dashed line stands for values predicted using equation 7.

Table 6: Maximum amounts of sorbed (ΔM) A1, B1, and C1 into NBR, FVMQ, and FKM.

Fluid	ΔM (%)		
	NBR	FVMQ	FKM
A1	14.2	4.7	0.3
B1	2.7	4.4	0.1
C1	30	3.9	0.4

351 Fluids A1, B1 and C1 – conventional and synthetic jet fuels – were
 352 considered to experimentally assess their compatibility with the three poly-
 353 mers of interest, and Table 6 presents measured ΔM values. In agreement
 354 with conclusions drawn previously, amounts of fluids adsorbed into FKM or
 355 FVMQ are roughly much lower than that measured for NBR. On the basis
 356 of compositions proposed in Figure 1 for A1, B1, and C1, ΔM plateau val-
 357 ues appear to follow the previously observed relationships with paraffins and
 358 aromatics contents.

359 3.2. Machine learning models

360 Obtained experimental values were used to feed machine learning tech-
 361 niques in order to derive predictive models. Based on the conclusions drawn
 362 in the previous section, small quantities of hydrocarbons were adsorbed into
 363 FKM and FVMQ and therefore, we only focus on modelling of the sorption of
 364 hydrocarbons into NBR. We hereafter report the development of two types of
 365 predictive models: models which predict the maximum mass gain (maximum
 366 ΔM) and models which predict the sorption kinetics *i.e.* the time evolution
 367 of ΔM , in NBR.

368 3.2.1. Modelling plateau values

369 As a first attempt, we considered a subpart of our database extracting
 370 for each fluid the sorption plateau value – the maximum mass percentage
 371 gain –, and data are presented in Table 2. Parameters of the MGGP such as
 372 numbers of runs, generations, genes, and nodes that will further be used to
 373 develop models were optimized using a 5-CV and according to the procedure
 374 proposed by Creton *et al.* [60]. This procedure can be summarized as follows:
 375 The numbers of genes and nodes are first set to their respective maximum
 376 allowed value to consider models having the highest complexity. Numbers of

Table 7: Performance characteristics (statistical indices) of MGGP based models applied to plateau values. Fold- i stands for performances calculated on the fold i when it is external to the learning procedure.

Indices	Fold-01	Fold-02	Fold-03	Fold-04	Fold-05
MAE	9.21	1.90	0.63	1.60	0.42
RMSE	17.23	2.87	0.75	2.80	0.54
R ²	0.786	0.992	0.995	0.614	0.929

377 generations and runs are optimized within the response surface with bound-
378 aries as defined in Table 5. Then, numbers of generations and runs are set to
379 optimized values, and numbers of genes and nodes are optimized within the
380 response surface defined according to boundaries indicated in Table 5. The
381 optimization procedure applied to our regression problem led to numbers of
382 runs, generations, genes, and nodes of 20, 500, 4, and 3, respectively.

383 Five MGGP based models were developed using the GPTIPS code and
384 following a 5-fold cross-validation procedure. All models exhibit excellent
385 performances over the Training sets. Performances of models evaluated on
386 external fluids are presented in Table 7. Values returned by indices for Fold-
387 01 and Fold-04 indicate overfitting trends for these two models that can
388 originate from fold constitution. For instance, Fold-01 contains the fluid F02
389 having the second highest property value in the database, and the model
390 fails in predicting this value. Among the three remaining models, best per-
391 formances on external sets are obtained for Fold-05, Fold-03, and then Fold-
392 02. However, the chemical diversity is not similar for these three folds. The
393 model that best generalizes the database has been developed using Fold-02
394 as Test set. Details about this latter model such as the four weighted genes
395 and the intercept value are presented in Equation (7).

$$\begin{aligned}
\lambda_0 &= \text{Intercept} = -30.12 \\
\lambda_1 G_1 &= -36.70 * \exp(-\exp(-X4)) \\
\lambda_2 G_2 &= 70.87 * \exp(X26) \\
\lambda_3 G_3 &= -12.38 * \sqrt[4]{X21} \\
\lambda_4 G_4 &= 7.26 * \exp(X33 - X10)
\end{aligned}
\tag{7}$$

396 where X_i stands for descriptors as defined in the Table 4. In Equation (7),

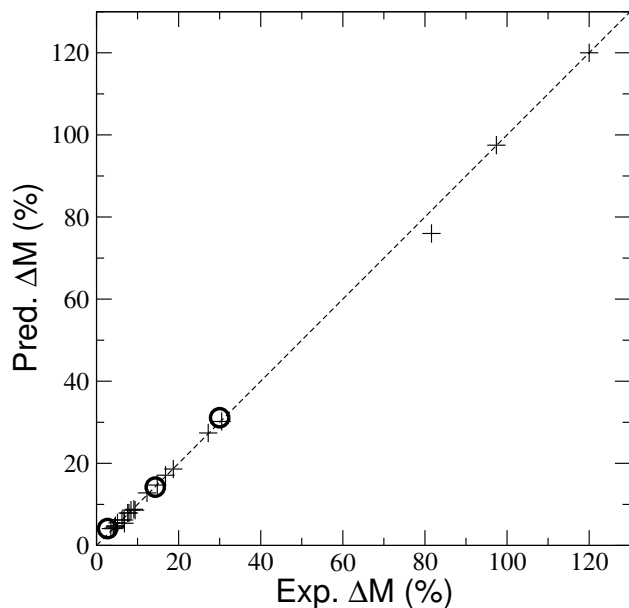


Figure 5: Scatterplots of experimental sorption values *vs.* predicted sorption values using Equation (7). Symbols + stand for fluids in Table 2, and \circ represent fuels A1, B1, and C1.

397 each gene non-linearly contributes to the predicted sorption value, and Equa-
 398 tion (7) highlights some interesting contributions of chemical function to the
 399 amount of fluid sorbed into NBR. For instance, Equation (7) indicates that
 400 increasing the number of $-\text{CH}_2-$ groups (X4) in fluid decreases the sorption
 401 value. On the contrary, increasing the number of aromatic rings (X26) in
 402 fluid increases its amount sorbed into NBR. These elements are in line with
 403 the analysis of Figure 4. G_4 is a combination between numbers of saturated
 404 6-rings (X33) and hydrogenated aromatic carbon atom bonded to two atoms
 405 by aromatic bonds (X10). Figure 5 presents scatterplots of experimental
 406 sorption values *vs.* predicted sorption values using equation (7). All data
 407 points are roughly located on the bisector (dashed line) indicating that pre-
 408 dicted plateau values are in excellent agreement with reference experimental
 409 data. Moreover, values predicted for fuels A1, B1, and C1 (14.2, 4.1, and
 410 31.1, respectively) are in excellent agreement with corresponding experimen-
 411 tal values as reported in Table 6.

Table 8: Optimized parameter settings used to train MGGP based models on our database.

Parameter	Optimized values
Number of runs	20
Number of generations	1000
Maximum number of genes	4
Maximum number of nodes	6

412 *3.2.2. Modelling the sorption kinetic*

413 We then considered the whole content of our database with for each fluid,
 414 the time evolution of the maximum mass percentage gain. Parameters of the
 415 MGGP as implemented in the GPTIPS code were optimized according to
 416 the procedure proposed by Creton *et al.* [60]. Additionally, we applied a
 417 5-CV procedure together with folds' chemistry – the same fluids in each
 418 fold – associated to the above described development of models to predict
 419 plateau values. Table 8 presents obtained optimized parameter values subse-
 420 quently used in GPTIPS to develop QSPR models. The number of nodes is
 421 twice higher as compared to parameters values optimized to derive Equation
 422 (7), and most probably due to this increase in complexity, the number of
 423 generations is here 1000.

424 Five MGGP based models were developed removing for each, one of the
 425 five folds defined for the 5-CV procedure. Performances of models evaluated
 426 on the Training and Test (external fluids) sets are presented in Table 9. All
 427 models exhibit excellent performances over the Training sets with RMSE
 428 lower than 2.6 (in ΔM unit) and R^2 greater than 0.99. Indices calculated for
 429 Test sets of models indicate various trends regarding their predictive capabil-
 430 ities. However, as discussed previously, the chemical diversity is not similar
 431 within the five folds, and external validation performed for these five scenar-
 432 ios are difficult to compare with each other. Values taken by indices over
 433 the database are presented in Table 9. Considering these latter values, the
 434 model developed using Fold-01 as Test set leads to a greater RMSE value
 435 as compared to others. Although none of models outperforms others, the
 436 model that best generalizes the database was obtained using Fold-05 as Test
 437 set. Details about this latter model such as the four weighted genes and the
 438 intercept value are presented in Equation (8).

Table 9: Performance characteristics (statistical indices) of MGGP based models applied to sorption curves. Fold- i stands for performances calculated on the fold i when it is external to the learning procedure.

Metrics	Fold-01	Fold-02	Fold-03	Fold-04	Fold-05
Training:					
MAE	1.02	1.29	1.52	1.49	1.38
RMSE	1.82	2.45	2.56	2.52	2.53
R ²	0.995	0.992	0.993	0.993	0.993
Test:					
MAE	2.96	1.42	1.27	1.54	0.69
RMSE	5.54	2.63	1.64	2.24	0.96
R ²	0.975	0.992	0.978	0.776	0.859
Database:					
MAE	1.34	1.31	1.48	1.50	1.26
RMSE	2.80	2.48	2.42	2.48	2.34
R ²	0.990	0.992	0.992	0.992	0.993

$$\begin{aligned}
 \lambda_0 = \text{Intercept} &= -38.40 \\
 \lambda_1 G_1 &= 12.23 * \exp(\exp(X26)) \\
 \lambda_2 G_2 &= 0.86 * (X9^4 + X22^3) \\
 \lambda_3 G_3 &= 4.51 * \sqrt{\exp(X10) * \ln(t)} \\
 \lambda_4 G_4 &= -2.35 * \exp(X10)
 \end{aligned} \tag{8}$$

439 where t is the time (expressed in hours) and X_i stands for descriptors as
440 defined in Table 4. In Equation (8), G_1 reveals that increasing the number
441 of aromatic rings (X26) in fluid increases its amount sorbed into NBR. G_2
442 can be considered as a sum of contributions of branchings on saturated (X9)
443 and aromatic (X22) rings. The descriptor X10 – number of hydrogenated
444 aromatic carbon atom bonded to two atoms by aromatic bonds – is involved
445 both in genes G_3 and G_4 where in this former, X10 acts as a weight for the
446 time evolution. Figure 6 presents scatterplots of experimental sorption values
447 *vs.* predicted sorption values using Equation (8). All data points are not too
448 scattered on both sides of the bisector indicating that predicted values are in
449 good agreement with reference experimental data. However, Figure 6 shows

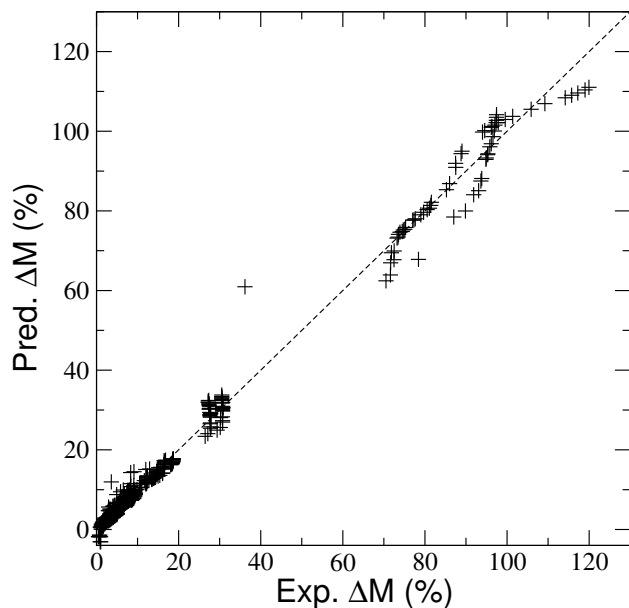


Figure 6: Scatterplots of experimental sorption values *vs.* predicted sorption values using Equation (8).

450 that one point is poorly predicted, the sorption value measured after 5 hours
 451 immersion in tetraline is 36.1 (%) while the model returns 61 (%). Noting
 452 that for tetraline, only this data point is poorly predicted.

453 We performed consensus modelling to investigate whether combining mod-
 454 els' predictions can lead to more accurate predicted values. It reveals that
 455 combining predictions of models obtained using Fold-04 and Fold-05 as Test
 456 sets improve performances as compared to individual models. We used this
 457 combinaison to predict the time evolutions of sorption values for real fuels
 458 A1, B1, and C1. Figure 7 presents comparisons between predicted values and
 459 experimental data measured in this study. The models successfully reproduce
 460 the sorption kinetics for the fluids A1 and C1, with however significant devi-
 461 ations from experimental values for the first hours. The models well predict
 462 B1 as a low ingress fluid in NBR but a shift of few percents is observed with
 463 reference experimental values.

464 4. Conclusion

465 We proposed here an investigation of fuels' sorption into polymers by
 466 means of experimental and machine learning techniques. Three polymers

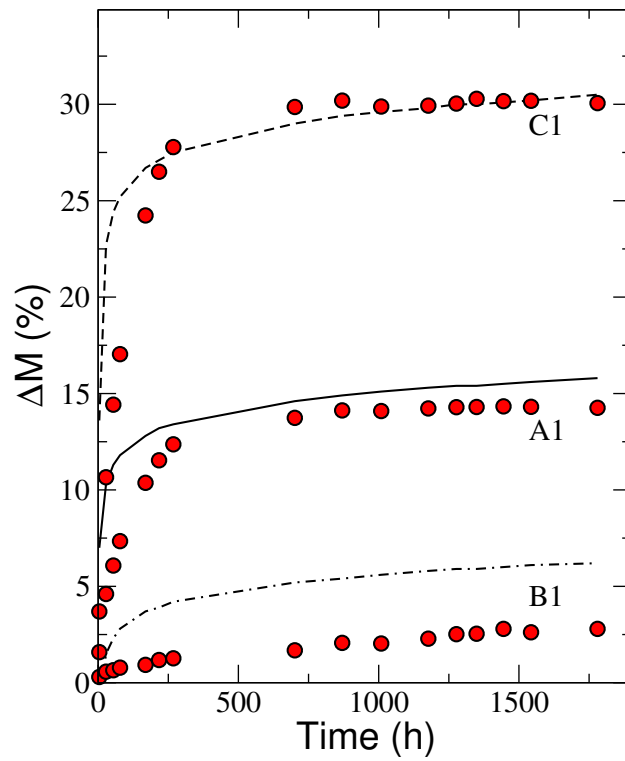


Figure 7: Time evolutions of sorption values for fuels A1, B1, and C1 predicted using the consensus model. Red circles stand for experimental values obtained in this work.

467 commonly considered for the design of fuel-delivery systems were selected
468 for this study: NBR, FKM, and FVMQ. Polymers samples were immersed
469 into liquids, and fluids under consideration were pure liquids and aviation
470 fuels – conventional and synthetic jet fuels. Sorption measurements were
471 performed for polymer/fluid couples, and experimental values were analysed
472 with chemoinformatics tools, and a machine learning method (*i.e.* MGGP)
473 and molecular descriptors (*i.e.* FGCD) were used to derive predictive models.

474 Performed sorption experiments to evaluate NBR, FVMQ, and FKM
475 compatibility with a series of hydrocarbons mixtures, have shown that FKM
476 can be assumed as a barrier polymer in case of such fluids, and that ΔM
477 values obtained for the FVMQ are small as compared to those for NBR. If
478 n- and iso-paraffins are fewly ingress into the NBR matrix, we demonstrated
479 that the swelling of NBR is strongly related to the amount of aromatics in
480 the studied liquids.

481 Machine learning techniques were used to derive two types of predictive
482 models. The first type of models aimed in predicting plateau ΔM values,
483 the maximum mass percentage gains. Models successfully reproduced exper-
484 imental data, and indicate that increasing the number of $-\text{CH}_2-$ groups and
485 aromatic rings in the fluid leads to decreasing and increasing the amount of
486 liquid sorbed into NBR, respectively. Application of the models to external
487 multi-component mixtures (not considered during the training procedure)
488 have demonstrated their predictive capabilities. The second type of models
489 aimed in predicting the sorption kinetics, *i.e.* the time evolution of ΔM .
490 Models reasonably reproduced experimental data, and in these models too,
491 increasing the number of aromatic rings in fluid contributes in increasing pre-
492 dicted values of ΔM , in NBR. Application of these models to external fluids
493 have demonstrated their capabilities in predicting both the kinetics and the
494 maximum ΔM values.

495 The determination of gases and liquids sorption into polymers is funda-
496 mental in many applications: fuels, lubricants, packaging, gas and liquids
497 transport and storage, among others. Our work shows that when using a
498 good quality database and relevant descriptions of fluids, machine learning
499 approaches are capable to catch soprtion phenomenon, and the so-obtained
500 predictive models are powerful tools to accurately estimate the sorption of
501 chemicals into a polymer. Moreover, such a modelling approach contributes
502 to drastically reduce the time necessary to quantify polymeric materials com-
503 patibility with a fluid candidate only knowing some of its structural charac-
504 teristics. This work is to be extended to other families of polymers and

505 fluids, as well as to explore new conditions of temperature and pressure. The
506 use of such models is interesting for assessing the impact of advanced fuels
507 formulations, for evaluating the impact of certain chemical families, or even
508 for determining the maximum amounts of biomass-based fluids into fuels.
509 In addition, these models could be used to replace the current qualitative
510 information – green, orange, and red symbols – in polymer compatibility
511 charts provided by resellers on their websites. The inversion of models based
512 on machine learning represents another interesting prospect for the design of
513 new polymers with desired properties [62, 63, 64].

514 **Acknowledgement**

515 Authors gratefully acknowledge Maira Alves-Fortunato, Axel Baroni, Arij
516 Ben Amara, Xavier Martin, Mickael Matrat, Laurie Starck for the fruitful
517 discussions. The research presented in this paper has been performed in the
518 framework of: (i) the European project JETSCREEN (JET fuel SCREENing
519 and optimization), and has received funding from the European Union Hori-
520 zon 2020 Programme under grant agreement no. 723525 ; (ii) and the French
521 national research program entitled CAER (Alternative Fuels for Aeronautics)
522 supported by French Directorate-General for Civil Aviation (DGAC).

523 **References**

- 524 [1] H. K. Jeswani, A. Chilvers, A. Azapagic, Environmental sustainability
525 of biofuels: a review, *Proceedings of the Royal Society A: Mathematical,*
526 *Physical and Engineering Sciences* 476 (2020) 20200351.
- 527 [2] D. A. Saldana, B. Creton, P. Mougín, N. Jeuland, B. Rousseau,
528 L. Starck, Rational formulation of alternative fuels using QSPR meth-
529 ods: Application to jet fuels, *Oil Gas Sci. Technol. - Rev. IFP Energies*
530 *nouvelles* 68 (2013) 651–662.
- 531 [3] D. A. Saldana, L. Starck, P. Mougín, B. Rousseau, L. Pidol, N. Jeu-
532 land, B. Creton, Flash point and cetane number predictions for fuel
533 compounds using quantitative structure property relationship (QSPR)
534 methods, *Energy & Fuels* 25 (2011) 3900–3908.
- 535 [4] H. Schulz, Short history and present trends of fischer-tropsch synthesis,
536 *Applied Catalysis A: General* 186 (1999) 3–12.

- 537 [5] K. Murata, Y. Liu, M. Inaba, I. Takahara, Production of synthetic
538 diesel by hydrotreatment of jatropha oils using Pt-Re/H-ZSM-5 catalyst,
539 *Energy & Fuels* 24 (2010) 2404–2409.
- 540 [6] W. Weiss, H. Dulot, A. Quignard, N. Charon, M. Courtiade, Direct coal
541 to liquids (DCL): High quality jet fuels, in: 27th Annual International
542 Pittsburgh Coal Conference, 2010.
- 543 [7] T. R. Carlson, G. A. Tompsett, W. C. Conner, G. W. Huber, Aromatic
544 production from catalytic fast pyrolysis of biomass-derived feedstocks,
545 *Topics in Catalysis* 52 (2009) 241–252.
- 546 [8] S. Akhlaghi, U. W. Gedde, M. S. Hedenqvist, M. T. Conde Braña,
547 M. Bellander, Deterioration of automotive rubbers in liquid biofuels:
548 A review, *Renewable and Sustainable Energy Reviews* 43 (2015) 1238–
549 1248.
- 550 [9] X.-F. Wei, L. De Vico, P. Larroche, K. J. Kallio, S. Bruder, M. Bellander,
551 U. W. Gedde, M. S. Hedenqvist, Ageing properties and polymer/fuel
552 interactions of polyamide 12 exposed to (bio)diesel at high temperature,
553 *npj Materials Degradation* 3 (2019) 1.
- 554 [10] P. M. Subramanian, *Polymer Blends*, American Chemical Society, Wash-
555 ington, DC, 1990, pp. 252–265. doi:10.1021/bk-1990-0423.ch013.
- 556 [11] S. M. Alves, V. S. E. Mello, F. K. Dutra-Pereira, Biodiesel com-
557 patibility with elastomers and steel, in: E. Jacob-Lopes, L. Queiroz
558 Zepka (Eds.), *Frontiers in bioenergy and biofuels*, InTech, Rijeka, 2017.
559 doi:10.5772/65551.
- 560 [12] P. Izák, L. Bartovská, K. Friess, M. Šípek, P. Uchytíl, Comparison of
561 various models for transport of binary mixtures through dense polymer
562 membrane, *Polymer* 44 (2003) 2679–2687.
- 563 [13] A. Randová, L. Bartovská, K. Friess, Š. Hovorka, P. Izák, Fundamental
564 study of sorption of pure liquids and liquid mixtures into polymeric
565 membrane, *European Polymer Journal* 61 (2014) 64–71.
- 566 [14] A. Randová, L. Bartovská, P. Izák, K. Friess, A new prediction method
567 for organic liquids sorption into polymers, *Journal of Membrane Science*
568 475 (2015) 545–551.

- 569 [15] L. Krajkova, M. Laskova, J. Chmelar, K. Jindrova, J. Kosek, Sorption
570 of liquid diluents in polyethylene: Comprehensive experimental data for
571 slurry polymerization, *Industrial & Engineering Chemistry Research* 58
572 (2019) 7037–7043.
- 573 [16] A. Haseeb, M. A. Fazal, M. I. Jahirul, H. H. Masjuki, Compatibility of
574 automotive materials in biodiesel: A review, *Fuel* 90 (2011) 922–931.
- 575 [17] A. Haseeb, T. S. Jun, M. A. Fazal, H. H. Masjuki, Degradation of phys-
576 ical properties of different elastomers upon exposure to palm biodiesel,
577 *Energy* 36 (2011) 1814–1819.
- 578 [18] M. D. Kass, T. Theiss, S. Pawel, J. Baustian, L. Wolf, W. Koch,
579 C. Janke, Compatibility assessment of elastomer materials to test fuels
580 representing gasoline blends containing ethanol and isobutanol, *SAE*
581 *Int. J. Fuels Lubr.* 7 (2014) 445–456.
- 582 [19] L. M. Silva, E. G. Filho, A. J. Simpson, M. R. Monteiro, T. Venâncio,
583 Comprehensive multiphase NMR spectroscopy: A new analytical
584 method to study the effect of biodiesel blends on the structure of com-
585 mercial rubbers, *Fuel* 166 (2016) 436–445.
- 586 [20] W. Trakarnpruk, S. Porntangjitlikit, Palm oil biodiesel synthesized with
587 potassium loaded calcined hydrotalcite and effect of biodiesel blend on
588 elastomer properties, *Renewable Energy* 33 (2008) 1558–1563.
- 589 [21] M. Weltshev, F. Heming, M. Haufe, M. Heyer, The influence of the
590 age of biodiesel and heating oil with 10 % biodiesel on the resistance
591 of sealing materials at different temperatures, *Materialwissenschaft und*
592 *Werkstofftechnik* 48 (2017) 837–845.
- 593 [22] A. Plota, A. Masek, Lifetime prediction methods for degradable poly-
594 meric materials? a short review, *Materials* 13 (2020) 4507.
- 595 [23] N. Nosengo, Can artificial intelligence create the next wonder material?,
596 *Nature* 533 (2016) 22–25.
- 597 [24] Y. Liu, Z. Hu, Z. Suo, L. Hu, L. Feng, X. Gong, Y. Liu, J. Zhang,
598 High-throughput experiments facilitate materials innovation: A review,
599 *Science China Technological Sciences* 62 (2019) 521–545.

- 600 [25] Y. Liu, O. C. Esan, Z. Pan, L. An, Machine learning for advanced energy
601 materials, *Energy and AI* 3 (2021) 100049.
- 602 [26] D. J. Audus, J. J. de Pablo, Polymer informatics: Opportunities and
603 challenges, *ACS Macro Letters* (2017) 1078–1082.
- 604 [27] L. Chen, G. Pilania, R. Batra, T. D. Huan, C. Kim, C. Kuenneth,
605 R. Ramprasad, Polymer informatics: Current status and critical next
606 steps, *Materials Science and Engineering: R: Reports* 144 (2021) 100595.
- 607 [28] A. R. Katritzky, S. Sild, M. Karelson, Correlation and prediction of the
608 refractive indices of polymers by QSPR, *Journal of Chemical Informa-
609 tion and Computer Sciences* 38 (1998) 1171–1176.
- 610 [29] A. R. Katritzky, M. Kuanar, S. Slavov, C. D. Hall, M. Karelson, I. Kahn,
611 D. A. Dobchev, Quantitative correlation of physical and chemical prop-
612 erties with chemical structure: utility for prediction, *Chemical reviews*
613 110 (2010) 5714–5789.
- 614 [30] T. Le, V. C. Epa, F. R. Burden, D. A. Winkler, Quantitative structure-
615 property relationship modeling of diverse materials properties, *Chemical
616 reviews* 112 (2012) 2889–2919.
- 617 [31] M. E. Erickson, M. Ngongang, B. Rasulev, A refractive index study of
618 a diverse set of polymeric materials by QSPR with quantum-chemical
619 and additive descriptors, *Molecules* 25 (2020).
- 620 [32] S. A. Schustik, F. Cravero, I. Ponzoni, M. F. Daz, Polymer informatics:
621 Expert-in-the-loop in QSPR modeling of refractive index, *Computa-
622 tional Materials Science* 194 (2021) 110460.
- 623 [33] A. J. Holder, L. Ye, J. D. Eick, C. C. Chappelow, A quantum-mechanical
624 QSAR model to predict the refractive index of polymer matrices, *QSAR
625 & Combinatorial Science* 25 (2006) 905–911.
- 626 [34] P. R. Duchowicz, S. E. Fioressi, D. E. Bacelo, L. M. Saavedra, A. P.
627 Toropova, A. A. Toropov, QSPR studies on refractive indices of struc-
628 turally heterogeneous polymers, *Chemometrics and Intelligent Labora-
629 tory Systems* 140 (2015) 86–91.

- 630 [35] F. Jabeen, M. Chen, B. Rasulev, M. Ossowski, P. Boudjouk, Refractive
631 indices of diverse data set of polymers: A computational QSPR based
632 study, *Computational Materials Science* 137 (2017) 215–224.
- 633 [36] A. G. Mercader, P. R. Duchowicz, Encoding alternatives for the pre-
634 diction of polyacrylates glass transition temperature by quantitative
635 structure–property relationships, *Materials Chemistry and Physics* 172
636 (2016) 158–164.
- 637 [37] A. G. Mercader, D. E. Bacelo, P. R. Duchowicz, Different encoding
638 alternatives for the prediction of halogenated polymers glass transition
639 temperature by quantitative structure–property relationships, *Internation-
640 al Journal of Polymer Analysis and Characterization* (2017) 1–10.
- 641 [38] A. Karuth, A. Alesadi, W. Xia, B. Rasulev, Predicting glass transition
642 of amorphous polymers by application of cheminformatics and molecular
643 dynamics simulations, *Polymer* 218 (2021) 123495.
- 644 [39] A. Toropov, A. Toropova, V. Kudyshkin, N. Bozorov, S. Rashidova,
645 Applying the monte carlo technique to build up models of glass transi-
646 tion temperatures of diverse polymers, *Structural Chemistry* 31 (2020)
647 1739–1743.
- 648 [40] F. Cravero, M. J. Martnez, I. Ponzoni, M. F. Daz, Computational mod-
649 elling of mechanical properties for new polymeric materials with high
650 molecular weight, *Chemometrics and Intelligent Laboratory Systems*
651 193 (2019) 103851.
- 652 [41] F. Cravero, M. J. Martnez, G. Vazquez, M. F. Daz, I. Ponzoni, Feature
653 learning applied to the estimation of tensile strength at break in poly-
654 meric material design, *Journal of Integrative Bioinformatics* 13 (2016)
655 15–29.
- 656 [42] T. Zhu, Y. Jiang, C. Haomiao, R. P. Singh, B. Yan, Development of
657 pp-LFER and QSPR models for predicting the diffusion coefficients of
658 hydrophobic organic compounds in ldpe, *Ecotoxicology and Environ-
659 mental Safety* 190 (2020) 110179.
- 660 [43] M. Li, H. Yu, Y. Wang, J. Li, G. Ma, X. Wei, QSPR models for predict-
661 ing the adsorption capacity for microplastics of polyethylene, polypropy-
662 lene and polystyrene, *Scientific Reports* 10 (2020) 14597.

- 663 [44] N. Villanueva, B. Flaconnèche, B. Creton, Prediction of alternative
664 gasoline sorption in a semicrystalline poly(ethylene), *ACS combinatorial*
665 *science* 17 (2015) 631–640.
- 666 [45] L. Starck, L. Pidol, N. Jeuland, T. Chapus, P. Bogers, J. Bauldreay,
667 Production of hydroprocessed esters and fatty acids (HEFA) - optimisa-
668 tion of process yield, *Oil Gas Sci. Technol. - Rev. IFP Energies nouvelles*
669 71 (2016) 10.
- 670 [46] C. Hall, B. Rauch, U. Bauder, P. Le Clercq, M. Aigner, Predictive ca-
671 pability assessment of probabilistic machine learning models for density
672 prediction of conventional and synthetic jet fuels, *Energy & Fuels* 35
673 (2021) 2520–2530.
- 674 [47] B. Creton, Chemoinformatics at IFP energies nouvelles: Applications in
675 the fields of energy, transport, and environment, *Molecular Informatics*
676 36 (2017) 1700028.
- 677 [48] C. Nieto-Draghi, G. Fayet, B. Creton, X. Rozanska, P. Rotureau, J.-
678 C. de Hemptinne, P. Ungerer, B. Rousseau, C. Adamo, A general
679 guidebook for the theoretical prediction of physicochemical properties of
680 chemicals for regulatory purposes, *Chemical Reviews* 115 (2015) 13093–
681 13164.
- 682 [49] P. Gramatica, Principles of QSAR models validation: Internal and ex-
683 ternal, *QSAR and Combinatorial Science* 26 (2007) 694–701.
- 684 [50] E. N. Muratov, E. V. Varlamova, A. G. Artemenko, P. G. Polishchuk,
685 V. E. Kuz’Min, Existing and developing approaches for QSAR analysis
686 of mixtures, *Molecular Informatics* 31 (2012) 202–221.
- 687 [51] T.-B. Nguyen, J.-C. de Hemptinne, B. Creton, G. M. Kontogeorgis,
688 Characterization scheme for property prediction of fluid fractions origi-
689 nating from biomass, *Energy & Fuels* 29 (2015) 7230–7241.
- 690 [52] D. Steinmetz, K. R. Arriola Gonzalez, R. Lugo, J. Verstraete, V. Lachet,
691 A. Mouret, B. Creton, C. Nieto-Draghi, Experimental and mesoscopic
692 modeling study of water/crude oil interfacial tension, *Energy & Fuels*
693 (2021).

- 694 [53] C. Vendevre, R. Ruiz-Guerrero, F. Bertoncini, L. Duval, D. Thiébaud,
695 Comprehensive two-dimensional gas chromatography for detailed char-
696 acterisation of petroleum products, *Oil Gas Sci. Technol. - Rev. IFP*
697 *Energies nouvelles* 62 (2007) 43–55.
- 698 [54] SMARTS - a language for describing molecular patterns; daylight chemi-
699 cal information systems inc.: Laguna niguél, ca, Accessed in 2020. URL:
700 <http://www.daylight.com/dayhtml/doc/theory/theory.smarts.html>.
- 701 [55] RDKit: Open-Source Cheminformatics Software, Accessed in 2020.
702 URL: <http://www.rdkit.org/>.
- 703 [56] C. Muller, A. G. Maldonado, A. Varnek, B. Creton, Prediction of opti-
704 mal salinities for surfactant formulations using a quantitative structure-
705 property relationships approach, *Energy Fuels* 29 (2015) 4281–4288.
- 706 [57] D. Searson, D. Leahy, M. Willis, GPTIPS: an open source genetic pro-
707 gramming toolbox for multigene symbolic regression, *Proceedings of*
708 *the International MultiConference of Engineers and Computer Scien-*
709 *tists 2010 (IMECS 2010)*, Hong Kong, 17-19 March (2010) 77–80.
- 710 [58] A. Gandomi, A. Alavi, C. Ryan, *Handbook of Genetic Programming*
711 *Applications*, Springer International Publishing, New York, 2015.
- 712 [59] D. Searson, *Handbook of Genetic Programming Applications*, in: [58],
713 2015.
- 714 [60] B. Creton, I. Lévêque, F. Oukhemanou, Equivalent alkane carbon num-
715 ber of crude oils: A predictive model based on machine learning, *Oil*
716 *Gas Sci. Technol. - Rev. IFP Energies nouvelles* 74 (2019) 30.
- 717 [61] O. Agwu, J. U. Akpabio, A. Dosunmu, Modeling the downhole density
718 of drilling muds using multigene genetic programming, *Upstream Oil*
719 *and Gas Technology* (2021) 100030.
- 720 [62] P. Gantzer, B. Creton, C. Nieto-Draghi, Inverse-QSPR for de novo
721 design: A review, *Molecular Informatics* 39 (2020) 1900087.
- 722 [63] P. Gantzer, B. Creton, C. Nieto-Draghi, Comparisons of molecular
723 structure generation methods based on fragment assemblies and genetic
724 graphs, *Journal of Chemical Information and Modeling* 61 (2021) 4245–
725 4258.

- ⁷²⁶ [64] K. Sattari, Y. Xie, J. Lin, Data-driven algorithms for inverse design of
⁷²⁷ polymers, *Soft Matter* 17 (2021) 7607–7622.