

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, 2022, 556, pp.113403. 10.1016/j.fluid.2022.113403. hal-03585804

HAL Id: hal-03585804 https://ifp.hal.science/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 \mathbb{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 1. Introduction

In the context of global warming, conclusions of research dedicated to the 2 reduction of greenhouse gases emissions advocate the use of alternative fuels 3 [1]. In particular, advanced fuels and biofuels including conventional renew-4 able fuels, respecting environmental criteria at a reasonable cost are of pri-5 mary interest [2]. Biofuels are issued from organic raw materials and they can 6 be seen as blends of renewable molecules such as normal- and iso-paraffins, 7 naphtenic and aromatic compounds, normal- and iso-olefins, alcohols, and/or esters [3]. Normal- and iso-paraffins can be obtained by industrial processes 9 such as Fisher-Tropsch (FT) [4] and hydrotreatment of vegetable oils (HVO) 10 [5]. In the same way, naphtenic and aromatic compounds can be synthesized 11 from the liquefaction or pyrolysis of biomass [6, 7]. Moreover, compositions 12 constantly evolving and being different from one country to another, it has 13 become essential to understand the impact of the introduction of these com-14 pounds on the physical properties of alternative fuels. The presence of these 15 families of molecules with different chemistry requires extensive research and 16 development activities. Indeed, it drives the conditions for storage, trans-17 portation, and combustion quality. 18

In combustion engine vehicles many pieces of the fuel-delivery systems 19 are composed all or part of polymers. Polymeric materials in contact with 20 fuels and/or biofuels may be subject to deformations such as swelling, caused 21 by solvent ingress within their structure and leading to strong modifications 22 and loss of their initial physical and chemical properties [8, 9]. To address 23 this problem, one solution consists in using multilayer structures containing 24 interleaved barrier polymers [10]. Up to date, only few works have been pub-25 lished in the litterature dealing with the compatibility between polymeric 26 materials and fluids, and there is a lack of available experimental data [11]. 27 The group of Izák *et al.* have investigated experimentally and theoretically 28 gases and liquids sorption into polymers over the last decades [12, 13, 14] 29 ; and more recently, Krajakova *et al.* focused on sorption of liquids into 30 poly(ethylene) samples of different densities [15]. Regarding fuels, Haseeb 31 et al. immersed some elastomeric materials such as nitrile butadiene rub-32 ber (NBR) and $Viton(\hat{\mathbf{R}})$ a fluorinated elastomer (FKM) in diesel and palm 33 biodiesel to compare the degradation of physical properties like weight and 34 volume changes, hardness and tensil strength [16, 17]. Kaas et al. studied 35 the compatibility of elastomeric materials with gasoline blends containing 36 ethanol and isobutanol, followed evolution of some polymer's properties, and 37

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

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

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

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

98 2. Materials and methods

99 2.1. Experimental procedure

100 2.1.1. Materials and Samples

Three polymers commonly considered for the design of fuel-delivery sys-101 tems were selected for this study: NBR, FKM, and FVMQ. Polymers raw 102 materials - plane square sheets with 0.3m size and 2.10⁻³m thickness -103 were supplied either by Zodiac Aerotechnics or by Stacem. Rectangular 104 parallelepiped shapes with $60 \times 10 \times 2 \text{ mm}^3$ were extracted from the polymer 105 sheets using a cutting shape, and samples were subsequently used for the 106 sorption tests. Some characteristics (grades for aerospace applications) for 107 these materials are presented in Table 1. We also performed measurements, 108 the Dynamic Mechanical Analysis (DMA) was used to determine the glass 109 transition temperature for the three polymers. A sinusoidal stress was ap-110 plied to each sample while the strain was measured, allowing one to determine 111

Dolumoon	Type	Standards	Hardness	T_g	Plasticizer
Polymer			$(IRHD^a)$	$(^{\circ}C)$	(% 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

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

^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.

the complex modulus and the loss factor. So-obtained T_q values, reported 112 in Table 1, correspond to the peak value of tan δ , the damping, a measure 113 of the energy dissipation of a material. Additionally, the Thermal Gravi-114 metric Analysis (TGA) was used, it consists in following the mass variations 115 of a sample with the time as the temperature changes. This measurement 116 provides information about chemical phenomena including thermal decom-117 position but also physical phenomena, such as the desorption of additives. 118 This technique has allowed the determination of the plasticizer amount ini-119 tially present in each studied polymer. It led to plasticizer amounts of 0.5~%120 wt. for FKM, 1.2 % wt. for FVMQ, and 11 % wt. for NBR, indicating that 121 the amount in NBR is not negligible and may lead to measurement artifacts. 122 Fluids under consideration in this study are pure liquids and aviation 123 fuels. Naphthenic and aromatic hydrocarbons (decaline (labeled D), xylene 124 (labeled X), tetraline (labeled T), iso-propylbenzene (or cumene, labeled C), 125 n-propylbenzene (labeled P), and methylnaphtalene (labeled M)), with high 126 purity grades were purchased from Merck, and no additional purification was 127 performed. One Jet A-1 (labeled J) being one of the fuels most commonly 128 used in commercial aviation was selected for sorption measurements. Addi-129 tionally, we considered alternative jet fuels approved for certification such as: 130 Synthetic Paraffinic Kerosene (SPK) and Hydroprocessed Esters and Fatty 131 Acids (HEFA). For instance, SPK can be FT fuels – composed of normal 132 and isoparaffins, or Alcohol to Jet Synthetic Paraffinic Kerosene (ATJ-SPK) 133 created from isobutanol which is derived from feedstocks. HEFA – similar 134 to HVO - includes hydrocarbon-based jet fuels (100% paraffinic) produced 135 from animal or vegetable oils by hydroprocessing [45]. The current certifica-136 tion for the use of HEFA in mixture with jet fuel allows a maximum of 50%137

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

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

157 2.1.2. Sorption measurements

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

¹⁷⁵ the initial polymer weight, as follows:

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

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

188 2.2. Modelling Method

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

195 2.2.1. Data Sets

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

During last decades of QSPR model developments, the use of external validation has been shown as necessary to ensure its ability to extrapolate 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	$\Delta M (\%)$		
Laber	Fluid	NBR	FVMQ	FKM
F01	Х	120.0	9.6	6.0
F02	Т	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

model [49]. Its popular version is the *n*-fold cross-validation (n-CV) in which 208 the data set is randomly divided in approximately equal n portions. An 209 aggregate of (n-1) portions forms the Training set – used to train models, 210 and the remaining portion constitutes the Test set - used to evaluate model's 211 performance. We emphasize that no data point belonging to external sets 212 is used to derived models. This procedure is repeated n times choosing at 213 each new fold another portion of data as a Test set. The subject of external 214 validation for QSPR analysis of mixtures has been addressed by Muratov et215 al. [50], and the authors-defined "mixture out" strategy was applied in this 216 study. 217

218 2.2.2. Fluids characterisation and representation

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

240 2.2.3. Molecular and mixture descriptors

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

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

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



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]		

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

The calculation of descriptors for mixtures has been addressed similarly as in previous works [44, 56]. We assumed mixture descriptors X_{mix} as linear combinations of pure component descriptors weighted with the associated molar fractions x_i . This approach has already been shown effective in predicting sorption values for some alternative fuels in a poly(ethylene) [44]. For instance, in the case of descriptor X1, the corresponding descriptor for a mixture $X1_{mix}$, is defined as follows:

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

where i runs over the N constituants in the mixture.

263 2.2.4. Chemical space representation

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

281 2.2.5. Machine learning algorithm

In the frame of past studies [47], we have observed that QSPR models de-282 rived from Support Vector Machine (SVM) algorithms frequently outperform 283 others evaluated learning algorithms such as neural networks, partial least 284 squares, genetic algorithm... Table 2 shows that the number of data points 285 - fluids candidates - is quite limited and the application of SVM does not 286 seem appropriate in this case. We focused on developing multilinear equa-287 tions which moreover have the advantage to be explicit models and easily 288 implemented in a spreadsheet. Such multilinear models can be, for instance, 280 generated by means of Evolutionary Algorithms (EA) techniques inspired 290



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.

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

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

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

Models are evaluated according to their capability in predicting fluids properties. Predicted values are compared to reference experimental data, and performances of models are evaluated by means of metrics such as MAE (Mean Absolute Error, equation (4)), RMSE (Root Mean Squared Error, equation (5)) or R² (Coefficient of determination, equation (6)), defined as follows:

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



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	$+, -, \times, \div, \sqrt{2}, \exp, \ln 2$			
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			

15

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

$$R^{2} = \frac{\sum_{i=1}^{N} (\widehat{y}_{i} - \overline{y})^{2}}{\sum_{i=1}^{N} (y_{i} - \overline{y})^{2}},$$
(6)

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

325 3. Results and discussion

326 3.1. Experimental results

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



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 (Δ M) A1, B1, and C1 into NBR, FVMQ, and FKM.

Fluid		$\Delta M \ (\%)$	
Fluid	NBR	FVMQ	FKM
A1	14.2	4.7	0.3
B1	2.7	4.4	0.1
C1	30	3.9	0.4

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

359 3.2. Machine learning models

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

368 3.2.1. Modelling plateau values

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

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
\mathbf{R}^2	0.786	0.992	0.995	0.614	0.929

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

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

$$\lambda_{0} = \text{Intercept} = -30.12$$

$$\lambda_{1}G_{1} = -36.70 * \exp(-\exp(-X4))$$

$$\lambda_{2}G_{2} = 70.87 * \exp(X26)$$

$$\lambda_{3}G_{3} = -12.38 * \sqrt[4]{X21}$$

$$\lambda_{4}G_{4} = 7.26 * \exp(X33 - X10)$$
(7)

³⁹⁶ where Xi 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 \bigcirc represent fuels A1, B1, and C1.

each gene non-linearly contributes to the predicted sorption value, and Equa-397 tion (7) highlights some interesting contributions of chemical function to the 398 amount of fluid sorbed into NBR. For instance, Equation (7) indicates that 399 increasing the number of $-CH_2$ - groups (X4) in fluid decreases the sorption 400 value. On the contrary, increasing the number of aromatic rings (X26) in 401 fluid increases its amount sorbed into NBR. These elements are in line with 402 the analysis of Figure 4. G_4 is a combination between numbers of saturated 403 6-rings (X33) and hydrogenated aromatic carbon atom bonded to two atoms 404 by aromatic bonds (X10). Figure 5 presents scatterplots of experimental 405 sorption values vs. predicted sorption values using equation (7). All data 406 points are roughly located on the bisector (dashed line) indicating that pre-407 dicted plateau values are in excellent agreement with reference experimental 408 data. Moreover, values predicted for fuels A1, B1, and C1 (14.2, 4.1, and 409 31.1, respectively) are in excellent agreement with corresponding experimen-410 tal values as reported in Table 6. 411

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

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

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

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
\mathbb{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
\mathbb{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
\mathbb{R}^2	0.990	0.992	0.992	0.992	0.993

$$\lambda_{0} = \text{Intercept} = -38.40$$

$$\lambda_{1}G_{1} = 12.23 * \exp(\exp(X26))$$

$$\lambda_{2}G_{2} = 0.86 * (X9^{4} + X22^{3})$$

$$\lambda_{3}G_{3} = 4.51 * \sqrt{\exp(X10) * \ln(t)}$$

$$\lambda_{4}G_{4} = -2.35 * \exp(X10)$$
(8)

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



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

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

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

464 4. Conclusion

We proposed here an investigation of fuels' sorption into polymers by 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.

commonly considered for the design of fuel-delivery systems were selected 467 for this study: NBR, FKM, and FVMQ. Polymers samples were immersed 468 into liquids, and fluids under consideration were pure liquids and aviation 469 fuels – conventional and synthetic jet fuels. Sorption measurements were 470 performed for polymer/fluid couples, and experimental values were analysed 471 with chemoinformatics tools, and a machine learning method (i.e. MGGP)472 and molecular descriptors (*i.e.* FGCD) were used to derive predictive models. 473 Performed sorption experiments to evaluate NBR, FVMQ, and FKM 474 compatibility with a series of hydrocarbons mixtures, have shown that FKM 475 can be assumed as a barrier polymer in case of such fluids, and that ΔM 476 values obtained for the FVMQ are small as compared to those for NBR. If 477 n- and iso-paraffins are fewly ingress into the NBR matrix, we demonstrated 478 that the swelling of NBR is strongly related to the amount of aromatics in 479 the studied liquids. 480

Machine learning techniques were used to derive two types of predictive 481 models. The first type of models aimed in predicting plateau ΔM values, 482 the maximum mass percentage gains. Models successfully reproduced exper-483 imental data, and indicate that increasing the number of $-CH_2$ - groups and 484 aromatic rings in the fluid leads to decreasing and increasing the amount of 485 liquid sorbed into NBR, respectively. Application of the models to external 486 multi-component mixtures (not considered during the training procedure) 487 have demonstrated their predictive capabilities. The second type of models 488 aimed in predicting the sorption kinetics, *i.e.* the time evolution of ΔM . 489 Models reasonably reproduced experimental data, and in these models too, 490 increasing the number of aromatic rings in fluid contributes in increasing pre-491 dicted values of ΔM , in NBR. Application of these models to external fluids 492 have demonstrated their capabilities in predicting both the kinetics and the 493 maximum ΔM values. 494

The determination of gases and liquids sorption into polymers is funda-495 mental in many applications: fuels, lubricants, packaging, gas and liquids 496 transport and storage, among others. Our work shows that when using a 497 good quality database and relevant descriptions of fluids, machine learning 498 approaches are capable to catch soprtion phenomenon, and the so-obtained 490 predictive models are powerful tools to accurately estimate the sorption of 500 chemicals into a polymer. Moreover, such a modelling approach contributes 501 to drastically reduce the time necessary to quantify polymeric materials com-502 patibility with a fluid candidate only knowing some of its structural charac-503 This work is to be extended to other families of polymers and teristics. 504

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

514 Acknowledgement

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

523 References

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

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

- [15] L. Krajakova, M. Laskova, J. Chmelar, K. Jindrova, J. Kosek, Sorption of liquid diluents in polyethylene: Comprehensive experimental data for slurry polymerization, Industrial & Engineering Chemistry Research 58 (2019) 7037–7043.
- ⁵⁷³ [16] A. Haseeb, M. A. Fazal, M. I. Jahirul, H. H. Masjuki, Compatibility of 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 phys⁵⁷⁶ 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.
- [19] L. M. Silva, E. G. Filho, A. J. Simpson, M. R. Monteiro, T. Venâncio, Comprehensive multiphase NMR spectroscopy: A new analytical method to study the effect of biodiesel blends on the structure of commercial rubbers, Fuel 166 (2016) 436–445.
- [20] W. Trakarnpruk, S. Porntangjitlikit, Palm oil biodiesel synthesized with
 potassium loaded calcined hydrotalcite and effect of biodiesel blend on
 elastomer properties, Renewable Energy 33 (2008) 1558–1563.
- [21] M. Weltschev, F. Heming, M. Haufe, M. Heyer, The influence of the age of biodiesel and heating oil with 10 % biodiesel on the resistance of sealing materials at different temperatures, Materialwissenschaft und Werkstofftechnik 48 (2017) 837–845.
- [22] A. Plota, A. Masek, Lifetime prediction methods for degradable poly meric materials? a short review, Materials 13 (2020) 4507.
- [23] N. Nosengo, Can artificial intelligence create the next wonder material?,
 Nature 533 (2016) 22–25.
- Y. Liu, Z. Hu, Z. Suo, L. Hu, L. Feng, X. Gong, Y. Liu, J. Zhang, High-throughput experiments facilitate materials innovation: A review, Science China Technological Sciences 62 (2019) 521–545.

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

- [35] F. Jabeen, M. Chen, B. Rasulev, M. Ossowski, P. Boudjouk, Refractive
 indices of diverse data set of polymers: A computational QSPR based
 study, Computational Materials Science 137 (2017) 215–224.
- [36] A. G. Mercader, P. R. Duchowicz, Encoding alternatives for the prediction of polyacrylates glass transition temperature by quantitative structure-property relationships, Materials Chemistry and Physics 172 (2016) 158-164.
- [37] A. G. Mercader, D. E. Bacelo, P. R. Duchowicz, Different encoding
 alternatives for the prediction of halogenated polymers glass transition
 temperature by quantitative structure-property relationships, International Journal of Polymer Analysis and Characterization (2017) 1–10.
- [38] A. Karuth, A. Alesadi, W. Xia, B. Rasulev, Predicting glass transition
 of amorphous polymers by application of cheminformatics and molecular
 dynamics simulations, Polymer 218 (2021) 123495.
- [39] A. Toropov, A. Toropova, V. Kudyshkin, N. Bozorov, S. Rashidova,
 Applying the monte carlo technique to build up models of glass transition temperatures of diverse polymers, Structural Chemistry 31 (2020)
 1739–1743.
- [40] F. Cravero, M. J. Martnez, I. Ponzoni, M. F. Daz, Computational modelling of mechanical properties for new polymeric materials with high molecular weight, Chemometrics and Intelligent Laboratory Systems 193 (2019) 103851.
- [41] F. Cravero, M. J. Martnez, G. Vazquez, M. F. Daz, I. Ponzoni, Feature
 learning applied to the estimation of tensile strength at break in polymeric material design, Journal of Integrative Bioinformatics 13 (2016)
 15–29.
- [42] T. Zhu, Y. Jiang, C. Haomiao, R. P. Singh, B. Yan, Development of
 pp-LFER and QSPR models for predicting the diffusion coefficients of
 hydrophobic organic compounds in ldpe, Ecotoxicology and Environmental Safety 190 (2020) 110179.
- [43] M. Li, H. Yu, Y. Wang, J. Li, G. Ma, X. Wei, QSPR models for predicting the adsorption capacity for microplastics of polyethylene, polypropylene and polystyrene, Scientific Reports 10 (2020) 14597.

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

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

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