

Article

Estimating Equivalent Alkane Carbon Number Using Abraham Solute Parameters

William E. Acree, Jr.¹, Wei-Khiong Chong², Andrew S.I.D. Lang^{3,*} and Hamed Mozafari³

¹ Department of Chemistry, University of North Texas, Denton, TX 76203, USA

² Advent Polytech Co., Ltd., Taipao City 61249, Taiwan

³ Department of Computing & Mathematics, Oral Roberts University, Tulsa, OK 74136, USA

* Correspondence: alang@oru.edu

Abstract: The use of equivalent alkane carbon numbers (EACN) to characterize oils is important in surfactant-oil-water (SOW) systems. However, the measurement of EACN values is non-trivial and thus it becomes desirable to predict EACN values from structure. In this work, we present a simple linear model that can be used to estimate the EACN value of oils with known Abraham solute parameters. We used linear regression with leave-one-out cross validation on a dataset of $N = 80$ oils with known Abraham solute parameters to derive a general model that can reliably estimate EACN values based upon the Abraham solute parameters: E (the measured liquid or gas molar refraction at 20°C minus that of a hypothetical alkane of identical volume), S (dipolarity/polarizability), A (hydrogen bond acidity), B (hydrogen bond basicity), and V (McGowan characteristic volume) with good accuracy within the chemical space studied ($N = 80$, $R^2 = 0.92$, RMSE = 1.16, MAE = 0.90, $p < 2.2 \times 10^{-16}$). These parameters are consistent with those in other models found in the literature and are available for a wide range of compounds.



Citation: Acree, W.E., Jr.; Chong, W.-K.; Lang, A.S.I.D.; Mozafari, H. Estimating Equivalent Alkane Carbon Number Using Abraham Solute Parameters. *Liquids* **2022**, *2*, 318–326. <https://doi.org/10.3390/liquids2040019>

Academic Editor: Slobodan Žumer

Received: 19 August 2022

Accepted: 26 September 2022

Published: 2 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Considerable attention has been given in recent years to development of better surfactant-based microemulsions and foam systems for enhanced oil recovery in petroleum processes, for removal of oil from contaminated soil and industrial machinery surfaces, and for the solubilization of fragrances in water-based formulations. Many factors including the temperature, electrolyte concentration, and the hydrophobicities of both the surfactant and oil contribute to the overall efficiency of the extraction system. Experimental determination of the optimum set of conditions for a given surfactant-oil-water system is both expensive and very time-consuming. Fortunately, empirical equations have been proposed to describe how the various factors affect microemulsion formation. One such expression is based on the hydrophilic-lipophilic difference (HLD) framework [1,2].

$$\text{Ionic surfactant: } \text{HLD} = \ln(S) - k \cdot \text{EACN} + C_c - \alpha \cdot (T - T_{\text{ref}}) \quad (1)$$

$$\text{Nonionic surfactant: } \text{HLD} = b \cdot S - K \cdot \text{EACN} + C_c + CT \cdot (T - T_{\text{ref}}) \quad (2)$$

where S is the salinity (not to be confused with Abraham's S parameter) and the terms $\ln(S)$ and $b \cdot S$ take into account the electrolyte concentration (usually in grams per 100 mL) of the system, b is electrolyte and surfactant specific, EACN is the equivalent alkane carbon number of the oil phase, C_c represents the hydrophilicity of the surfactant, and the last two terms, $\alpha \cdot (T - T_{\text{ref}})$ and $CT \cdot (T - T_{\text{ref}})$ are related to the temperature effect. The application of HLD in predicting the type and microemulsion phase behavior is described in greater detail elsewhere [1–3].

Our interest in the current study is in developing a predictive method for the equivalent alkane carbon number (EACN), which for simple alkanes is numerically equal to the number of carbons (ACN), and for other liquids it is equal to the number of carbons of the n-alkane exhibiting a similar phase behavior in a reference surfactant-oil-water (SOW) system. EACNs can be determined by comparing the oil's fish-tail-temperature (T^*) in a reference SOW to standard calibration curves for n-alkanes [4]. The experimental determination of EACN values for novel oils may be time consuming and the ability to predict EACN values is advantageous. Bouton et al. [5] have developed a two-descriptor model based upon experimental data for 43 oils using the proprietary Molecular Operating Environment (MOE) software:

$$\text{EACN} = -19.84 + 2.88 \cdot \text{average negative softness} + 0.88 \cdot \text{KierA3} \quad (3)$$

where KierA3 is the third alpha modified shape index and “average negative softness”—which is related to polarizability. Lukowicz et al. [6] have developed a three-descriptor model based upon 70 oils using COSMO-RS σ -moments:

$$\text{EACN} = -4.85 - 0.23 \cdot M_2 - 0.33 \cdot M_{\text{acc}} + 0.06 \cdot M_0, \quad (4)$$

where M_2 is molecular polarity, M_{acc} is hydrogen bond basicity, and M_0 is total molecular surface area. This extends previous work in this area [7]. Most recently, Delforce et al. [8] have developed both a graph machine model using SMILES codes and a neural network model using COSMO-RS-computed σ -moments based on reliable EACN values for 111 molecules.

This work develops a model for EACN based upon experimental EACN values for 80 liquids using the five Abraham solute parameters E, S, A, B, and V which encode physicochemical properties related to those already found to be important—namely size and shape, polarizability, and hydrogen bond basicity, i.e., we propose the following model:

$$\text{EACN} = c + e \cdot E + s \cdot S + a \cdot A + b \cdot B + v \cdot V \quad (5)$$

where E is the solute excess molar refractivity—the measured liquid or gas molar refraction at 20 °C minus that of a hypothetical alkane of identical volume—in units of (cm³/mol)/10, S is the solute dipolarity/polarizability, A and B are the overall or summation hydrogen bond acidity and basicity, and V is the McGowan characteristic volume in units of (cm³/mol)/100.

2. Materials and Methods

Experimentally measured EACN values, collected by Aubry et al. [4,8], were combined with their experimentally determined Abraham solute parameters, primarily from the UFZ-LSER database [9], with the values for decylcyclohexane taken from a paper by Chung et al. [10], and the values for dodecylcyclohexane and bis(2-ethylhexyl) adipate new to this work (from an unpublished database of measured Abraham parameters original to Professor Abraham dated December 2020 shared with one of the authors before Professor Abraham’s passing on the 19 January 2021), see Table 1. A modeling dataset was created from these data by: 1. Using median EACN values for compounds with multiple experimental measurements. 2. Only keeping compounds with all 5 Abraham parameters available (measured, not predicted) [9]. This dataset of N = 86 compounds with EACN values and Abraham parameters is available under a CC0 license from figshare [11]. Modeling was performed using R v4.2.0 (R Core Team, Vienna, Austria) [12].

Table 1. Measured EACN values with available Abraham Solute Parameters.

Compound	EACN	E	S	A	B	V	EACN Ref.
Branched and cyclic alkanes							
Cyclohexane	2.4	0.305	0.10	0.00	0.00	0.8454	[13]
Cyclohexane	1.7	0.305	0.10	0.00	0.00	0.8454	[13]
Cyclohexane	1.8	0.305	0.10	0.00	0.00	0.8454	[14]
Cyclohexane	2.5	0.305	0.10	0.00	0.00	0.8454	[15]
Methylcyclohexane	3.5	0.244	0.06	0.00	0.00	0.9863	[13]
Methylcyclohexane	2.8	0.244	0.06	0.00	0.00	0.9863	[13]
1,4-Dimethylcyclohexane	4.6	0.191	0.17	0.00	0.00	1.1272	[13]
1,4-Dimethylcyclohexane	4.5	0.191	0.17	0.00	0.00	1.1272	[13]
Ethylcyclohexane	4.5	0.263	0.10	0.00	0.00	1.1272	[13]
Ethylcyclohexane	4.5	0.263	0.10	0.00	0.00	1.1272	[13]
Ethylcyclohexane	3.7	0.263	0.10	0.00	0.00	1.1272	[16]
Cyclooctane	4.1	0.409	0.10	0.00	0.00	1.1272	[6]
1,2-Dimethylcyclohexane	3.9	0.320	0.23	0.00	0.00	1.1272	[16]
1,2-Dimethylcyclohexane	2.6	0.320	0.23	0.00	0.00	1.1272	[13]
Propylcyclohexane	5.8	0.257	0.14	0.00	0.00	1.2681	[13]
Propylcyclohexane	6.3	0.257	0.14	0.00	0.00	1.2681	[13]
Propylcyclohexane	5.5	0.257	0.14	0.00	0.00	1.2681	[16]
Isopropylcyclohexane	5.6	0.283	0.07	0.00	0.00	1.2681	[13]
Isopropylcyclohexane	5.7	0.283	0.07	0.00	0.00	1.2681	[13]
Isopropylcyclohexane	4.5	0.283	0.07	0.00	0.00	1.2681	[13]
Butylcyclohexane	7.2	0.255	0.14	0.00	0.00	1.4090	[13]
Butylcyclohexane	7.9	0.255	0.14	0.00	0.00	1.4090	[17]
Butylcyclohexane	6.9	0.255	0.14	0.00	0.00	1.4090	[16]
Cyclodecane	5.6	0.474	0.10	0.00	0.00	1.4090	[18]
cis-Decalin	5.3	0.544	0.25	0.00	0.00	1.3004	[18]
Myrcane	10.0	0.000	0.00	0.00	0.00	1.5176	[5,19]
Myrcane	11.2	0.000	0.00	0.00	0.00	1.5176	[6,20]
Myrcane	10.1	0.000	0.00	0.00	0.00	1.5176	[13]
Pinane	4.3	0.421	0.12	0.00	0.13	1.3004	[6,20]
Pinane	3.9	0.421	0.12	0.00	0.13	1.3004	[6,20]
p-Menthan	6.3	0.270	0.07	0.00	0.00	1.4090	[6,20]
p-Menthan	6.7	0.270	0.07	0.00	0.00	1.4090	[6,20]
p-Menthan	4.6	0.270	0.07	0.00	0.00	1.4090	[6,20]
Decylcyclohexane	14.4	0.243	0.23	0.00	0.00	2.2544	[21]
Dodecylcyclohexane	17.5	0.300	0.23	0.00	0.00	2.5362	[16]
Isododecane	11.7	0.000	0.00	0.00	0.00	1.7994	[7]
Hemisqualene	14.8	0.000	0.00	0.00	0.00	2.2221	[7]
Squalane	24.5	-	-	-	-	-	[21]
Squalene	13.8	-	-	-	-	-	[22]
Halogenated alkanes							
1-Bromo-2-methylpropane	-3.4	0.340	0.37	0.00	0.12	0.8472	[18]
1-Chlorooctane	1.0	0.191	0.40	0.00	0.09	1.3582	[23]
1-Chlorodecane	3.5	0.185	0.40	0.00	0.09	1.6400	[18]
1,10-Dichlorodecane	6.3	0.366	-	0.00	-	1.7624	[24]
1-Chlorododecane	5.6	0.181	0.40	0.00	0.10	1.9218	[18]
1-Chlorododecane	5.8	0.181	0.40	0.00	0.10	1.9218	[23]
1-Chlorotetradecane	8.0	0.176	0.41	0.00	0.10	2.2036	[18]
1-Chlorotetradecane	7.3	0.176	0.41	0.00	0.10	2.2036	[23]
1-Chlorohexadecane	9.8	0.173	0.42	0.00	0.10	2.4854	[18]
1-Chlorohexadecane	9.0	0.173	0.42	0.00	0.10	2.4854	[23]

Table 1. Cont.

Compound	EACN	E	S	A	B	V	EACN Ref.
Alkenes, terpenes, alkynes and aromatics							
Cyclohexene	-1.2	0.395	0.28	0.00	0.09	0.8024	[13]
1,3-Cyclohexadiene	-3.1	0.515	0.38	0.00	0.12	0.7594	[13]
1,4-Cyclohexadiene	-4.1	0.501	0.46	0.00	0.16	0.7594	[13]
1-Methyl-1-cyclohexene	0.8	0.391	0.18	0.00	0.10	0.9433	[13]
1-Methyl-1-cyclohexene	-0.8	0.391	0.18	0.00	0.10	0.9433	[13]
4-Methyl-1-cyclohexene	0.6	0.347	0.22	0.00	0.10	0.9433	[13]
4-Methyl-1-cyclohexene	-0.5	0.347	0.22	0.00	0.10	0.9433	[13]
3-Methyl-1-cyclohexene	0.4	0.360	0.20	0.00	0.10	0.9433	[13]
3-Methyl-1-cyclohexene	-1.4	0.360	0.20	0.00	0.10	0.9433	[13]
2,5-Norbornadiene	-3.2	0.495	0.32	0.00	0.11	0.7919	[13]
1-Octene	3.9	0.094	0.08	0.00	0.07	1.1928	[18]
cis-Cyclooctene	1.5	0.460	0.24	0.00	0.10	1.0842	[18]
1-Octyne	-1.8	0.155	0.22	0.09	0.10	1.1498	[18]
p-Xylene	-2.4	0.613	0.52	0.00	0.16	0.9982	[18]
1-Decene	5.5	0.093	0.08	0.00	0.07	1.4746	[18]
1-Decyne	0.1	0.143	0.22	0.09	0.10	1.4316	[18]
Butylbenzene	0.4	0.600	0.51	0.00	0.15	1.2800	[18]
Phenyl-1-butyne	-3.3	-	-	-	-	-	[18]
alpha-Pinene	3.6	0.438	0.20	0.00	0.14	1.2574	[6,20]
alpha-Pinene	3.4	0.438	0.20	0.00	0.14	1.2574	[6,20]
p-Menth-2-ene	3.1	0.350	0.12	0.00	0.07	1.3660	[19]
p-Menth-2-ene	3.6	0.350	0.12	0.00	0.07	1.3660	[6,20]
Delta-3-carene	2.9	0.492	0.22	0.00	0.14	1.2574	[6,20]
Delta-3-carene	2.0	0.492	0.22	0.00	0.14	1.2574	[6,20]
beta-Pinene	2.3	0.515	0.19	0.00	0.15	1.2574	[6,20]
beta-Pinene	2.0	0.515	0.19	0.00	0.15	1.2574	[6,20]
Limonene	2.0	0.501	0.31	0.00	0.23	1.3230	[6,20]
Limonene	1.6	0.501	0.31	0.00	0.23	1.3230	[6,20]
gamma-Terpinene	1.9	0.522	0.29	0.00	0.22	1.3230	[6,20]
gamma-Terpinene	1.4	0.522	0.29	0.00	0.22	1.3230	[6,20]
alpha-Terpinene	1.5	0.526	0.25	0.00	0.23	1.3230	[6,20]
alpha-Terpinene	0.8	0.526	0.25	0.00	0.23	1.3230	[6,20]
Terpinolene	1.3	0.590	0.31	0.00	0.20	1.3230	[6,20]
Terpinolene	0.1	0.590	0.31	0.00	0.20	1.3230	[6,20]
p-Cymene	-0.3	0.607	0.49	0.00	0.19	1.2800	[6,20]
p-Cymene	-1.3	0.607	0.49	0.00	0.19	1.2800	[6,20]
1-Dodecene	8.1	0.089	0.08	0.00	0.07	1.7564	[18]
1-Dodecyne	2.0	0.133	0.22	0.09	0.10	1.7134	[18]
1-Tetradecyne	3.9	-	-	-	-	-	[18]
Octylbenzene	4.0	0.579	0.48	0.00	0.15	1.8436	[16]
2,6,10-Trimethylundecane-2,6-diene	10.3	-	-	-	-	-	[6,20]
Longifolene	6.6	0.757	0.20	0.00	0.22	1.8533	[20]
Longifolene	7.3	0.757	0.20	0.00	0.22	1.8533	[6,20]
Caryophyllene	5.7	0.720	0.15	0.00	0.25	1.9189	[6,20]
Caryophyllene	6.2	0.720	0.15	0.00	0.25	1.9189	[6,20]
Decylbenzene	6.0	0.579	0.47	0.00	0.15	2.1254	[25]
1-Octadecene	14.4	0.079	0.08	0.00	0.07	2.6018	[18]
Dodecylbenzene	7.8	0.571	0.47	0.00	0.15	2.4072	[25]

Table 1. Cont.

Compound	EACN	E	S	A	B	V	EACN Ref.
Ethers, esters, nitriles and ketones							
Diisopropyl ether	2.2	−0.063	0.17	0.00	0.57	1.0127	[26]
Dibutyl ether	2.4	0.000	0.25	0.00	0.45	1.2945	[22]
Dibutyl ether	3.3	0.000	0.25	0.00	0.45	1.2945	[25]
Dibutyl ether	3.2	0.000	0.25	0.00	0.45	1.2945	[27]
2-Octanone	−3.4	0.108	0.68	0.00	0.51	1.2515	[22]
Octanenitrile	−1.7	0.162	0.90	0.00	0.36	1.2500	[22]
Dipentyl ether	4.2	0.000	0.25	0.00	0.45	1.5763	[22]
C3-O-C4-O-C3	1.9	−	−	−	−	−	[27]
C4-O-C2-O-C4	1.7	−	0.51	0.00	−	1.6350	[27]
2-Decanone	−2.1	0.108	0.68	0.00	0.51	1.5333	[22]
Decanenitrile	−0.6	0.156	0.90	0.00	0.36	1.5320	[22]
2-Undecanone	−1.3	0.101	0.68	0.00	0.51	1.6742	[22]
Ethyl decanoate	1.8	0.013	0.58	0.00	0.45	1.8738	[17]
Ethyl decanoate	2.3	0.013	0.58	0.00	0.45	1.8738	[20]
Ethyl decanoate	2.2	0.013	0.58	0.00	0.45	1.8738	[18]
Dihexyl ether	6.2	0.000	0.25	0.00	0.45	1.8581	[22]
2-Dodecanone	−0.6	0.103	0.68	0.00	0.51	1.8151	[22]
Dodecanenitrile	0.3	0.132	0.90	0.00	0.36	1.8132	[22]
Ethyl dodecanoate	3.8	0.002	0.58	0.00	0.45	2.1556	[6]
Decyl butyrate	5.0	−	−	−	−	−	[6]
Hexyl octanoate	6.2	0.002	0.56	0.00	0.45	2.1556	[6]
Diheptyl ether	8.0	−	−	−	−	−	[22]
Ethyl myristate	5.2	0.000	0.58	0.00	0.45	2.4374	[6]
Butyl dodecanoate	7.2	−	−	−	−	−	[6]
Octyl octanoate	8.1	−0.010	0.06	0.00	0.45	2.4374	[6]
Diethyl ether	10.3	0.000	0.25	0.00	0.45	2.4217	[22]
Myristyl propionate	6.8	−	−	−	−	−	[6]
Isopropyl myristate	7.2	−0.062	0.53	0.00	0.45	2.5783	[6]
Isopropyl myristate	7.3	−0.062	0.53	0.00	0.45	2.5783	[25]
Ethyl palmitate	6.8	0.000	0.58	0.00	0.45	2.7192	[6]
Hexyl dodecanoate	9.3	−	−	−	−	−	[20]
Ethyl oleate	7.3	−	−	−	−	−	[6]
Ethyl oleate	7.1	−	−	−	−	−	[28]
Bis(2-ethylhexyl) adipate	9.7	−0.010	1.10	0.00	1.13	3.3572	[6]
Tricaprilin	12.2	−	−	−	−	−	[6]
Tricaprin	13.8	−	−	−	−	−	[6]
Tricaprin	13.0	−	−	−	−	−	[25]
Trilaurin	15.7	−	−	−	−	−	[5]
Trimyristin	18.5	−	−	−	−	−	[5]
Tripalmitin	21.2	−	−	−	−	−	[5]
Tristearin	23.9	−0.040	1.25	0.00	1.28	8.3631	[5]
Triolein	21.2	−	−	−	−	−	[5]
Fragrances, acrylates and miscellaneous							
Menthone	−1.5	0.322	0.61	0.00	0.62	1.4247	[21]
Eucalyptol	−1.6	0.380	0.33	0.00	0.76	1.3591	[21]
Rose oxide	−1.7	−	−	−	−	−	[21]
D-Carvone	−3.1	0.674	0.86	0.00	0.57	1.3390	[21]
Hexyl methacrylate	0.4	0.154	0.49	0.00	0.45	1.5490	[29]
Hexyl methacrylate	−0.2	0.154	0.49	0.00	0.45	1.5490	[29]
Hexyl methacrylate	−0.1	0.154	0.49	0.00	0.45	1.5490	[29]
Hexyl methacrylate	0.8	0.154	0.49	0.00	0.45	1.5490	[29]

Table 1. Cont.

Compound	EACN	E	S	A	B	V	EACN Ref.
Fragrances, acrylates and miscellaneous							
Hexyl methacrylate	0.7	0.154	0.49	0.00	0.45	1.5490	[29]
Hexyl methacrylate	-0.2	0.154	0.49	0.00	0.45	1.5490	[29]
Hexyl methacrylate	1.5	0.154	0.49	0.00	0.45	1.5490	[29]
Hexyl methacrylate	0.2	0.154	0.49	0.00	0.45	1.5490	[29]
Hexyl methacrylate	-0.4	0.154	0.49	0.00	0.45	1.5490	[29]
Menthyl acetate	-0.1	0.243	0.65	0.00	0.54	1.7652	[6,20]
Citronellyl acetate	-0.2	0.198	0.59	0.00	0.64	1.8308	[21,22]
Geranyl acetate	-0.6	0.368	0.65	0.00	0.68	1.7878	[21,22]
Linalyl acetate	-0.9	0.331	0.65	0.00	0.65	1.7878	[21,22]
alpha-Damascone	-1.3	-	-	-	-	-	[21]
Methyl dihydrojasmonate	-1.7	0.340	1.53	0.00	0.97	1.9218	[21]
beta-Ionone	-1.9	0.892	0.78	0.00	0.76	1.7614	[17]
Ethylene brassylate	-1.1	-	-	-	-	-	[21]
Methyl cedryl ether	3.5	-	-	-	-	-	[21]
Ambrettolide	1.0	-	-	-	-	-	[21]

3. Results

The standard Abraham solute descriptor-based model is represented by Equation (6), where c is the intercept, E , S , A , B , and V are the Abraham solute descriptors, and e , s , a , b , v are their coefficients obtained by linear regression:

$$\text{EACN} = c + e \cdot E + s \cdot S + a \cdot A + b \cdot B + v \cdot V \quad (6)$$

When using all the data, we found a moderate correlation ($R^2 = 0.61$) between S and B . Removing S -parameter outliers bis(2-ethylhexyl) adipate, tristearin, and methyl dihydrojasmonate ($S > 1$) resulted in a dataset where all pairwise correlations had coefficients of determination lower than 0.50 and the coefficients of determination of each parameter against all others was lower than 0.80. Removing S -parameter outliers provides a greater reliability to the model, but may limit its application to a smaller chemical space.

Performing linear regression with leave-out-out (LOO) cross-validation showed that dodecylcyclohexane, decylcyclohexane, and octyl octanoate were clear outliers. Removing the three outliers, and again using LOO linear regression, we found that EACN can be estimated with similar accuracy (LOO measures) using Abraham solute parameters ($N = 80$, $R^2 = 0.92$, $\text{RMSE} = 1.16$, $\text{MAE} = 0.90$, $p < 2.2 \times 10^{-16}$) as compared to previous models [5–8], at least within the chemical space represented in the study:

$$\text{EACN} = -2.16 - 2.08 \cdot E - 9.51 \cdot S - 50.91 \cdot A - 5.41 \cdot B + 6.83 \cdot V \quad (7)$$

Analyzing the EACN estimates for each compound type, we see that the model performs the best for alkenes, terpenes, alkynes, and aromatics ($N = 31$, $\text{ME} = 0.37$, $\text{MAE} = 0.57$, $\text{RMSE} = 0.76$). Good performance is seen for the four other types: branched and cyclic alkanes ($N = 16$, $\text{ME} = -0.81$, $\text{MAE} = 0.85$, $\text{RMSE} = 1.11$); halogenated alkanes ($N = 6$, $\text{ME} = 0.99$, $\text{MAE} = 0.99$, $\text{RMSE} = 1.16$); fragrances, acrylates, and miscellaneous ($N = 9$, $\text{ME} = -0.16$, $\text{MAE} = 0.94$, $\text{RMSE} = 1.18$); and ethers, esters, nitriles, and ketones ($N = 18$, $\text{ME} = -0.17$, $\text{MAE} = 1.16$, $\text{RMSE} = 1.32$)—with consistent over-prediction for halogenated alkanes and consistent under-prediction for branched and cyclic alkanes, see Figure 1.

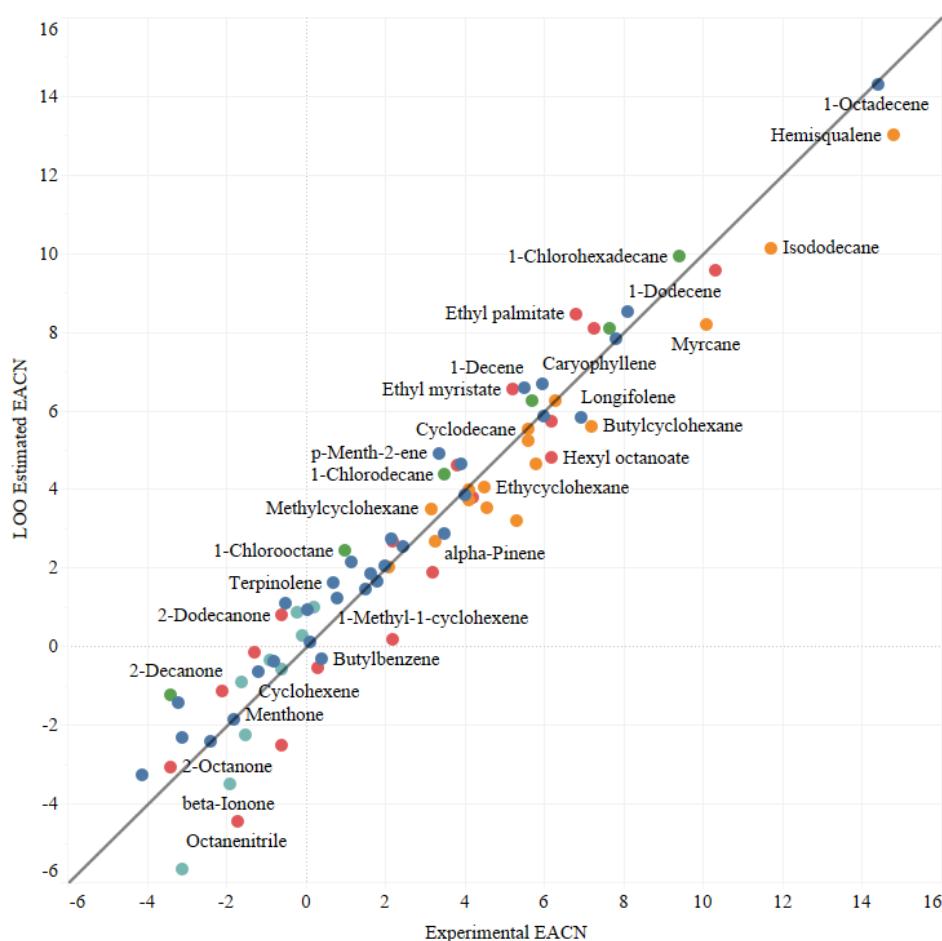


Figure 1. Estimated vs. Measured EACN values colored by type: alkenes, terpenes, alkynes, and aromatics (dark blue); branched and cyclic alkanes (orange); ethers, esters, nitriles, and ketones (red); fragrances, acrylates, and miscellaneous (light blue); and halogenated alkanes (green).

4. Discussion

We have demonstrated that EACN can be estimated using the standard Abraham solute parameter model, see Equation (7). The first four parameters all have negative coefficients where E is the solute excess molar refractivity—the measured liquid or gas molar refraction at 20 °C minus that of a hypothetical alkane of identical volume-in units of $(\text{cm}^3/\text{mol})/10$, S is the solute dipolarity/polarizability, A and B are the overall or summation hydrogen bond acidity and basicity, and V is the McGowan characteristic volume in units of $(\text{cm}^3/\text{mol})/100$. These results align with previous results [5–8], using different parameter systems, but showing similar accuracy and that EACN has contributions from shape (size and branching) [5–8], polarity/polarizability [5–8], and hydrogen bond basicity [6–8]. Our addition of hydrogen bond acidity, represented by the A descriptor, leads to superior estimation of EACN values for alkynes something not seen in previous models.

We began with a dataset of $N = 86$ oils with both measured EACN values and measured Abraham solute descriptors. During modeling we removed six compounds: bis(2-ethylhexyl) adipate, tristearin, methyl dihydrojasmonate, decylcyclohexane, dodecylcyclohexane, and octyl octanoate.

The first three compounds were removed because they had large S -values which resulted in an artificially high collinearity with the B parameter. The second set of three compounds were removed as outliers from a first LOO cross-validation analysis. Even so, Equation (7)-predicted EACN values for these compounds are generally of the right order, see the predicted values in the open dataset [11]. The utility of Equation (7) can also be seen by using it to predict EACN values of compounds that have measured EACN values but

that do not have measured Abraham solute parameters. Using predicted Abraham solute parameters [9], we predicted the EACN values of several of these compounds without measured Abraham solute parameters from Table 1, see Table 2. For the compounds listed in Table 2, Equation (7) performs relatively well, with statistics similar to those found for the estimated EACN results above, specifically: N = 11, ME = 0.17, MAE = 1.13, RMSE = 1.61.

Table 2. Equation (7)-predicted EACN values using predicted Abraham solute parameters.

Compound	EACN	E	S	A	B	V	Predicted EACN
1-Tetradecyne	3.9	0.150	0.24	0.05	0.12	1.9952	5.7
2,6,10-Trimethylundecane-2,6-diene	10.3	0.350	0.23	0.00	0.34	2.1361	7.7
Decyl butyrate	5.0	0.000	0.56	0.00	0.55	2.1556	4.3
Butyl dodecanoate	7.2	0.010	0.56	0.00	0.52	2.4374	6.3
Myristyl propionate	6.8	0.050	0.56	0.00	0.53	2.5783	7.2
Diheptyl ether	8.0	0.000	0.18	0.00	0.46	2.1399	8.3
Hexyl dodecanoate	9.3	0.010	0.56	0.00	0.52	2.7192	8.3
Ethyl oleate	7.1	0.130	0.72	0.00	0.68	2.9580	7.2
Ethyl oleate	7.3	0.130	0.72	0.00	0.68	2.9580	7.2
Methyl cedryl ether	3.5	0.650	0.23	0.00	0.22	2.0959	7.4
Ambrettolide	1.0	0.540	0.68	0.00	0.76	2.2858	1.8
alpha-Damascone	-1.3	0.680	0.71	0.00	0.54	1.7614	-1.2

The most recent paper by Delforce et al. [8] notes that the measured EACN of 2.2 for diisopropyl ether reported previously [26] is an outlier for their model. Our model estimates the EACN value of diisopropyl ether to be 0.2 which is in line with their newly measured EACN value of 0.6.

Our approach provides a useful tool for estimating equivalent alkane carbon numbers as Abraham solute parameters are available for a significant number of compounds [9,10]. While a general model is presented, models for specific families of compounds can be easily created using our open dataset [11]. We also note that the estimated EACN values of individual hydrocarbons from Equation (7) will allow estimation of EACN values of heavy hydrocarbon mixtures, $EACN_{mix}$, using the mathematical expression proposed by Cayias et al. [30] and Cash et al. [31]:

$$EACN_{mix} = \sum_{i=1}^N x_i EACN_i \quad (8)$$

where x_i and $EACN_i$ denote the mole fraction and numerical EACN value of the individual hydrocarbon component i , respectively.

Future research directions include measuring the EACN values of more diverse compounds, especially those with known non-zero A-parameter values.

Author Contributions: Conceptualization, W.-K.C.; methodology, W.E.A.J., A.S.I.D.L. and H.M.; data collection and curation, A.S.I.D.L. and H.M.; writing—original draft preparation, W.E.A.J., W.-K.C., A.S.I.D.L. and H.M.; writing—review and editing, W.E.A.J., W.-K.C., A.S.I.D.L. and H.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data used in this study is available from figshare under a CC0 license [11].

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Salager, J.L.; Morgan, J.C.; Schechter, R.S.; Wade, W.H.; Spe-Aime, M.; Vasquez, E. Optimum formulation of surfactant/water/oil systems for minimum interfacial tension or phase behavior. *Soc. Pet. Eng. AIME J.* **1979**, *19*, 107–115. [[CrossRef](#)]
2. Ghayour, A.; Acosta, E. Characterizing the oil-like and surfactant-like behavior of polar oils. *Langmuir* **2019**, *35*, 15038–15050. [[CrossRef](#)] [[PubMed](#)]

3. Virin Kittithammavong, V.; Charoensaeng, A.; Khaodhiar, S. A Normalized HLD (HLD_N) tool for optimal salt-concentration prediction of microemulsions. *Appl. Sci.* **2021**, *11*, 9151. [[CrossRef](#)]
4. Aubry, J.-M.; Ontiveros, J.F.; Salager, J.L.; Nardello-Rataj, V. Use of the normalized hydrophilic-lipophilic-deviation (HLD_N) equation for determining the equivalent alkane carbon number (EACN) of oils and the preferred alkane carbon number (PACN) of nonionic surfactants by the fish-tail method (FTM). *Adv. Colloid Interface Sci.* **2020**, *276*, 102099. [[CrossRef](#)]
5. Bouton, F.; Durand, M.; Nardello-Rataj, V.; Borosy, A.P.; Quellet, C.; Aubry, J.M. A QSPR model for the prediction of the “fish-tail” temperature of C_iE4/water/polar hydrocarbon oil systems. *Langmuir* **2010**, *26*, 7962–7970. [[CrossRef](#)]
6. Lukowicz, T.; Illous, E.; Nardello-Rataj, V.; Aubry, J.M. Prediction of the equivalent alkane carbon number (EACN) of aprotic polar oils with COSMO-RS sigma-moments. *Colloids Surf. A Physicochem. Eng. Asp.* **2018**, *536*, 53–59. [[CrossRef](#)]
7. Lukowicz, T.; Benazzouz, A.; Nardello-Rataj, V.; Aubry, J.-M. Rationalization and Prediction of the Equivalent Alkane Carbon Number (EACN) of Polar Hydrocarbon Oils with COSMO-RS σ -Moments. *Langmuir* **2015**, *31*, 11220–11226. [[CrossRef](#)]
8. Delforce, L.; Duprat, F.; Ploix, J.-L.; Ontiveros, J.F.; Goussard, V.; Nardello-Rataj, V.; Aubry, J.-M. Fast Prediction of the Equivalent Alkane Carbon Number Using Graph Machines and Neural Networks. *ACS Omega* **2022**, *in press*.
9. Ulrich, N.; Endo, S.; Brown, T.N.; Watanabe, N.; Bronner, G.; Abraham, M.H.; Goss, K.-U. UFZ-LSER Database v 3.2.1 [Internet], Leipzig, Germany, Helmholtz Centre for Environmental Research-UFZ. 2017. Available online: <http://www.ufz.de/lserd> (accessed on 31 December 2021).
10. Chung, Y.; Vermeire, F.H.; Wu, H.; Walker, P.J.; Abraham, M.H.; Green, W.H. Group Contribution and Machine Learning Approaches to Predict Abraham Solute Parameters, Solvation Free Energy, and Solvation Enthalpy. *J. Chem. Inf. Modeling* **2022**, *62*, 433–446. [[CrossRef](#)]
11. Acree, W.E.; Chong, W.; Lang, A.S.I.D.; Mozafari, H. Dataset: EACN and Abraham Parameter Values. Figshare. *Dataset* **2022**. [[CrossRef](#)]
12. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. 2022. Available online: <https://www.R-project.org/> (accessed on 31 December 2021).
13. Bouton, F. Influence of Terpenes and Terpenoids on the Phase Behavior of Micro-and Macro-Emulsions. Ph.D. Thesis, University of Lille, Lille, France, 2010.
14. Kahlweit, M.; Strey, R.; Busse, G. Effect of alcohols on the phase behavior of microemulsions. *J. Phys. Chem.* **1991**, *95*, 5344–5352. [[CrossRef](#)]
15. Burauer, S.; Sottmann, T.; Strey, R. Nonionic microemulsions with cyclic oils—Oil penetration, efficiency and monomeric solubility. *Tenside Surfactants Deterg.* **2000**, *37*, 8–16.
16. Queste, S.; Salager, J.L.; Strey, R.; Aubry, J.M. The EACN scale for oil classification revisited thanks to fish diagrams. *J. Colloid Interface Sci.* **2007**, *312*, 98–107. [[CrossRef](#)] [[PubMed](#)]
17. Lukowicz, T. Synergistic Solubilisation of Fragrances in Binary Surfactant Systems and Prediction of Their EACN Value with COSMO-RS. Ph.D. Thesis, Université Lille, Lille, France, 2015.
18. Tchakalova, V.; Testard, F.; Wong, K.; Parker, A.; Benczédi, D.; Zemb, T. Solubilization and interfacial curvature in microemulsions: I. Interfacial expansion and co-extraction of oil. *Colloids Surf. A Physicochem. Eng. Asp.* **2008**, *331*, 31–39. [[CrossRef](#)]
19. Bouton, F.; Durand, M.; Nardello-Rataj, V.; Serry, M.; Aubry, J.M. Classification of terpene oils using the fish diagrams and the Equivalent Alkane Carbon (EACN) scale. *Colloids Surf. A Physicochem. Eng. Asp.* **2009**, *338*, 142–147. [[CrossRef](#)]
20. Ontiveros, J.F.; Pierlot, C.; Catté, M.; Molinier, V.; Pizzino, A.; Salager, J.L.; Aubry, J.M. Classification of ester oils according to their Equivalent Alkane Carbon Number (EACN) and asymmetry of fish diagrams of C10E4/ester oil/water systems. *J. Colloid Interface Sci.* **2013**, *403*, 67–76. [[CrossRef](#)]
21. Gradzielski, M.; Langevin, D.; Sottmann, T.; Strey, R. Droplet microemulsions at the emulsification boundary: The influence of the surfactant structure on the elastic constants of the amphiphilic film. *J. Chem. Phys.* **1997**, *106*, 8232–8238. [[CrossRef](#)]
22. Tchakalova, V.; Testard, F.; Wong, K.; Parker, A.; Benczédi, D.; Zemb, T. Solubilization and interfacial curvature in microemulsions: II. Surfactant efficiency and PIT. *Colloids Surf. A Physicochem. Eng. Asp.* **2008**, *331*, 40–47. [[CrossRef](#)]
23. Israelachvili, J.N.; Mitchell, D.J.; Ninham, B.W. Theory of self-assembly of hydrocarbon amphiphiles into micelles and bilayers. *J. Chem. Soc. Faraday Trans. 2 Mol. Chem. Phys.* **1976**, *72*, 1525–1568. [[CrossRef](#)]
24. Deen, G.R.; Pedersen, J.S. Phase behavior and microstructure of C12E5 nonionic microemulsions with chlorinated oils. *Langmuir* **2008**, *24*, 3111–3117. [[CrossRef](#)]
25. Engelskirchen, S.; Elsner, N.; Sottmann, T.; Strey, R. Triacylglycerol microemulsions stabilized by alkyl ethoxylate surfactants—A basic study: Phase behavior, interfacial tension and microstructure. *J. Colloid Interface Sci.* **2007**, *312*, 114–121. [[CrossRef](#)] [[PubMed](#)]
26. Wormuth, K.R.; Kaler, E.W. Microemulsifying polar oils. *J. Phys. Chem.* **1989**, *93*, 4855–4861. [[CrossRef](#)]
27. Tanford, C. *The Hydrophobic Effect: Formation of Micelles and Biological Membranes*, 2nd ed.; J. Wiley: Hoboken, NJ, USA, 1980.
28. Kahlweit, M.; Busse, G.; Faulhaber, B.; Eibl, H. Preparing nontoxic microemulsions. *Langmuir* **1995**, *11*, 4185–4187. [[CrossRef](#)]
29. van Os, N.M.; Haak, J.R.; Rupert, L.A.M. *Physico-Chemical Properties of Selected Anionic, Cationic and Nonionic Surfactants*; Elsevier: Amsterdam, The Netherlands, 2012.
30. Cayias, J.L.; Schechter, R.S.; Wade, W.H. Modeling crude oils for low interfacial tension. *Soc. Pet. Eng. J.* **1976**, *16*, 351–357. [[CrossRef](#)]
31. Cash, L.; Cayias, J.L.; Fournier, G.; Macallister, D.; Schares, T.; Schechter, R.S.; Wade, W.H. The application of low interfacial tension scaling rules to binary hydrocarbon mixtures. *J. Colloid Interface Sci.* **1976**, *59*, 39–44. [[CrossRef](#)]