energies

MDPI

Review

A Comparative Assessment of Biodiesel Cetane Number Predictive Correlations Based on Fatty Acid Composition

Evangelos G. Giakoumis * and Christos K. Sarakatsanis

Internal Combustion Engines Laboratory, School of Mechanical Engineering,
National Technical University of Athens, 15780 Athens, Greece; chris.sarakatsanis@gmail.com
* Correspondence: vgiakms@central.ntua.gr; Tel.: +30-210-772-1360

Received: 30 December 2018; Accepted: 24 January 2019; Published: 29 January 2019

Abstract: Sixteen biodiesel cetane number (CN) predictive models developed since the early 1980s have been gathered and compared in order to assess their predictive capability, strengths and shortcomings. All are based on the fatty acid (FA) composition and/or the various metrics derived directly from it, namely, the degree of unsaturation, molecular weight, number of double bonds and chain length. The models were evaluated against a broad set of experimental data from the literature comprising 50 series of measured CNs and FA compositions. It was found that models based purely on compositional structure manifest the best predictive capability in the form of coefficient of determination R^2. On the other hand, more complex models incorporating the effects of molecular weight, degree of unsaturation and chain length, although reliable in their predictions, exhibit lower accuracy. Average and maximum errors from each model's predictions were also computed and assessed.

Keywords: biodiesel; cetane number; chain length; coefficient of determination; degree of unsaturation; fatty acid composition

1. Introduction

The level of research carried out in recent decades regarding the use of biodiesel in engines has been intense. One unique aspect of biodiesels (and their parent oils/fats) is the fact that they are produced from a variety of vegetable or animal feedstock [1–5]. These possess different compositional structures in the form of fatty acids (FA), as summarized in Table 1 for the most influential ones. Past research has shown that a relatively high degree of variability is observed regarding the structural form of common biodiesels depending on the originating oil [6]. Since the composition of each oil/fat in the fatty acids varies, the physical and chemical properties of biodiesel differ too. Perhaps the most prominent example here is the cetane number (cold-flow properties as well) [3,5,7].

The dimensionless cetane number is one of the most influential fuel properties. It is highly responsible for ignition delay; thus it determines, to a large extent, the proportion between premixed and diffusion combustion in a diesel engine [8]. In this regard, it affects the heat release profile and is also responsible for the emission of pollutants and combustion noise [3,5,9–13]. In light of the above, it is not surprising that the CN has been widely researched and reported in the literature, with the published values differing quite a lot. Figure 1 is helpful here, illustrating the large variability in the reported CN values [6]. These values (average for each feedstock) range from lower than that of the respective automotive diesel fuel up to much higher. As discussed earlier, this variability in the measured CN values can be, at least in part, attributed to the different fatty acid methyl ester (FAME) composition, with computational and experimental errors and uncertainties playing a non-negligible

role too. To make things even worse, sometimes large variations have been reported in the literature, even when biodiesels from the same originating feedstock have been studied. This is reflected in the sometimes large standard deviation of the collected data, as depicted in Figure 2. Knothe [10] argued that one major factor affecting the variability of the results is the fact that CN is actually a 'lumped' quantity that expresses various phenomena such as spray formation, vaporization, mixing etc.

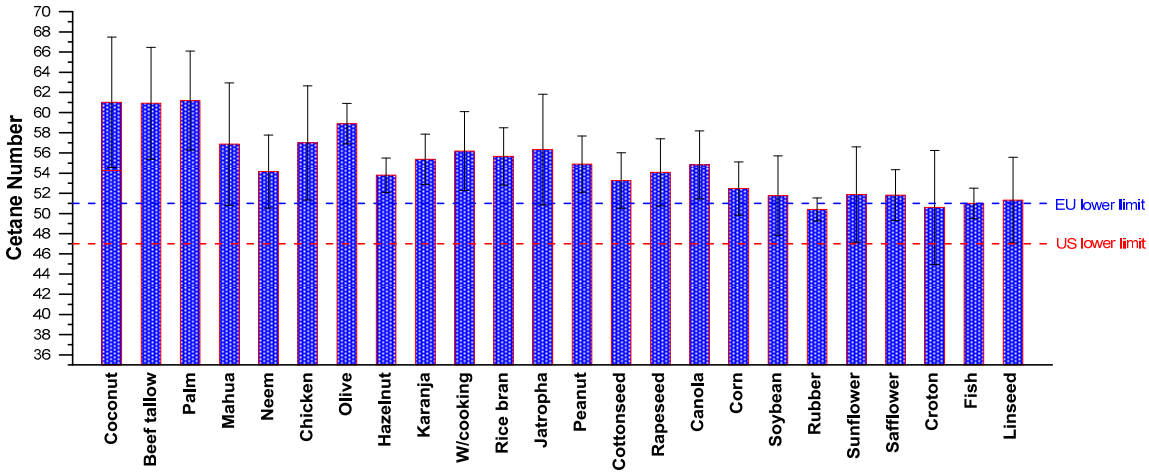

Figure 1. Average cetane numbers of biodiesels from various feedstocks; the EU and US lower limits correspond to automotive applications (reprinted from [6] with permission from Elsevier).

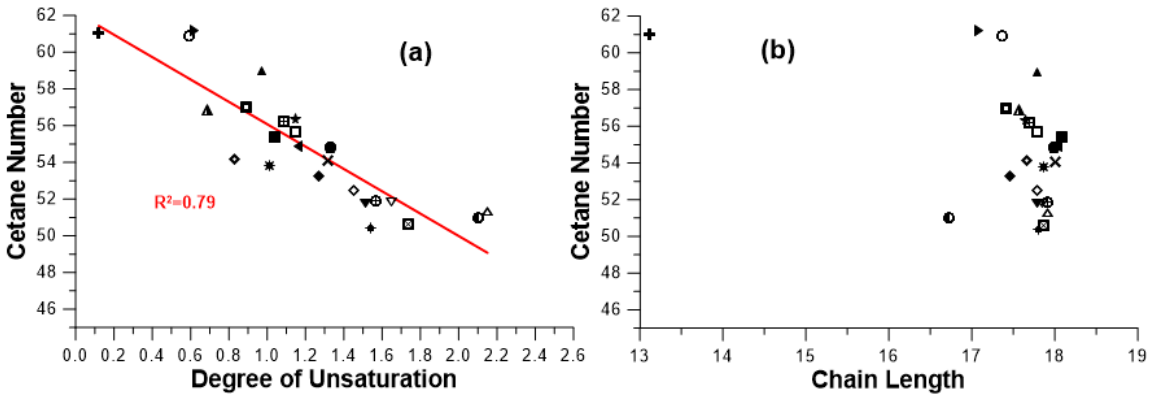

Figure 2. Correlation between degree of unsaturation (**a**) and chain length (**b**) with biodiesel average cetane number from 24 feedstocks (sub-figure (**a**) reprinted from [6] with permission from Elsevier).

Owing to its significance in engine performance and emissions, as well as to the fact that its experimental determination is time consuming, costly and scientifically challenging, it is not surprising that several CN predictive models have been developed in the past. Biodiesel CN has been correlated in these models with various other metrics or properties. Klopfenstein [14] reported one of the earliest correlations with respect to the number of carbon atoms and the number of double bonds. Similarly, Ramirez-Verduzco et al. [15] correlated CN with the molecular weight and the number of double bonds, and Pinzi et al. [16] with the degree of unsaturation and chain length. A quadratic correlation with the number of carbon atoms in the original fatty acid and the number of double bonds was statistically selected as the most suitable by Lapuerta et al. [17].

Other researchers preferred correlations of the CN directly with the fatty acid composition (without intermediate metrics such as the degree of unsaturation or the chain length). Bamgboye and Hansen [18] reported the first correlation of this kind with respect to the FA composition applying multiple linear regression analysis (MLR). Piloto-Rodriguez et al. [19] applied both MLR and artificial

neural networks (ANN). The latter approach was the one followed by Ramadhas et al. [20], while the former (MLR) was chosen by Gopinath et al. [21] as well as by the present research group [22].

One common feature in all these models is the fact that at the time each one was introduced, it was claimed to be superior over the previous ones with regards to its predictive capability and the associated relative errors in its predictions. It seems reasonable, and at the same time interesting for the scientific community, to try and evaluate all models on an objective basis and identify the truly exceptional ones. Please note that for the comparative evaluation, only models based solely on the FA composition (directly, or indirectly through the degree of unsaturation or the chain length) are taken into account, and only for methyl esters (ethyl/propyl/butyl esters are not included in the analysis). This means that models that correlate CN with other physical or chemical properties such as saponification value or density (e.g. [11,21,23,24]) will not be considered. For the latter category of models, which is equally wide, Refs. [21,23] are a good starting point for a review of the available literature and corresponding equations. In the same regard, models based on artificial neural networks are not considered either, as the respective "algorithms" are obviously not available.

The paper is organized as follows: Section 2 will provide some fundamentals on the CN and its measuring procedure. Section 3 will review and discuss the predictive models for biodiesel CN based on the FA composition that are used in the comparative analysis. Then, in Section 4, which is the core of this work, a detailed comparison of all models will be performed to assess their predictive capability, identify trends and inefficiencies and reach some interesting overall conclusions.

2. Cetane Number Fundamentals

The cetane number of a fuel (typical values between 35 and 65) is an indicator of its ignition and combustion quality. Its name derives from cetane or n-hexadecane ($C_{16}H_{34}$), which is a straight (without branching) saturated hydrocarbon. The cetane's ease of ignition is very high; hence, it has been assigned a CN of 100. On the other hand, α-methylnaphthalene or AMN ($C_{11}H_{10}$, a polycyclic aromatic hydrocarbon) was assigned a CN of 0, owing to its poor auto-ignition quality. Because of storage stability and availability issues with AMN, the reference fuel for the lower end of the cetane number scale was changed in 1962 to 2,2,4,4,6,8,8-heptamethylnonane (HMN) (also hexadecane $C_{16}H_{34}$, yet highly branched), which was assigned a CN of 15 [10,25].

The cetane number of a fuel is influenced by its physical and chemical characteristics. Physical attributes in particular, such as density, viscosity and surface tension, affect the time needed for the fuel to vaporize and mix with the surrounding air. The CN also depends on the molecular structure of its constituent hydrocarbons. In general, and as past research has established [10–12,23], long straight-chain hydrocarbons (minimum branching) possess high CN values, and are therefore ideal for fuels in diesel engines. Double bonds, i.e., the unsaturation level, on the other hand, lead to lower CN values. In this regard, alkenes possess lower CNs than alkanes. With regards to biodiesel, methyl esters from highly-saturated feedstock, such as palm, coconut and tallow, exhibit the highest CN values, i.e., in the order of 60, as demonstrated in Figure 2 (but worse cold-flow properties). A quite strong dependence between CN and molecular weight (chain length too) has been confirmed in the literature.

The universally accepted method for measuring CN is described in ISO 5165:2017 ("Petroleum products—Determination of the ignition quality of diesel fuels—Cetane engine method") and ASTM D 613 ("Standard Test Method for Cetane Number of Diesel Fuel Oil"). Both employ a standard CFR single-cylinder, four-stroke, indirect injection diesel engine of variable compression ratio. The procedure uses fuels of known cetane number as references. Using the fuel under test, the engine compression ratio is varied so as to achieve the same ignition delay as those with two 'bracketing' reference fuels, thereby permitting the interpolation of the CN [25]. European standard DIN 16906:2017 (formerly 51773:2010) "Liquid petroleum products—Determination of the ignition quality of diesel fuels—BASF engine method" employs a different engine of fixed compression ratio, where during the test, the amount of induced air is altered.

The shortcomings of these methods are: the high fuel cost for pure compounds that are very expensive, time, sophisticated laboratory facilities, and expertise for accurate measurements. In light of the above, an alternative CN test method has been developed in recent years employing a constant volume combustion chamber (CVCC), named 'ignition quality tester' (IQT). In this method, a sample of fuel is injected into a heated CVCC at approximately 575 °C and 21.37 bar (310 psi). Initially, the pressure decreases due to cooling from fuel evaporation, but then increases as combustion commences, ultimately achieving the initial pressure at the pressure recovery point. The ignition delay is then determined as the time between the beginning of the injector needle lift and the chamber pressure recovery point. A correlation has also been developed between the measured ignition delay and the cetane number. This method has been approved by ASTM as Standard D 6890 ("Standard test method for determination of ignition delay and derived cetane number (DCN) of diesel fuel oils by combustion in a constant volume chamber"). The main advantages of the IQT Care the shorter test time (approx. 20 min compared to few hours for the CFR procedure), smaller volume samples of the test fuel (approx. one tenth of the CFR), and lower reproducibility errors [25]. The CN values determined using the IQT are known as DCNs, i.e., derived CNs.

An alternative to the cetane number is the cetane index (CI). CI is calculated from the fuel's density and (10, 50 and 90%) distillation temperatures, known as the four-variable equation. ISO 4264:2018 ("Petroleum products - Calculation of cetane index of middle-distillate fuels by the four-variable equation") and ASTM D4737 ("Standard test method for calculated cetane index by four variable equation") describe the relevant computational procedure. For hydrocarbon fuels, CI is an easily obtainable substitute for CN, saving refineries a lot of work and rendering CN-testing engines unnecessary; the same can be applied to biodiesel when a test engine is not available to determine CN directly, or when an insufficient amount of a sample is available for an engine rating.

The current work only deals with CN and not CI values.

3. Review of Biodiesel CN Predictions Based on the FA Composition

The available literature on predictive correlations for biodiesel properties based on their fatty acid composition is quite broad. Table 1 provides information on the most important fatty acids found in the FAME mixture that will be discussed in the text. In this work, we have gathered all available methyl ester CN models published in the last 30 years based on the following criteria:

(a) The models are based on the fatty acid composition (directly, or indirectly through the chain length or the molecular weight),
(b) The model equations are explicitly provided by the authors (hence, no artificial neural network algorithms have been considered).

The models that were found can be classified into three categories:

- Models that are based directly on the biodiesel fatty acid composition (termed in this review 'compositional' models), i.e., without employing 'intermediate' metrics in the calculations such as the degree of unsaturation or the chain length. For this kind of modeling, an equation of the following type is formulated

$$CN = A + \sum_i a_i w_i \quad (1)$$

with A and a_i being independent variables to be derived from a multiple linear regression analysis by the least squares method. Index i denotes the individual fatty acid in the biodiesel mixture, and w_i is the weight percentage of each FA. The commonly employed acids, as shown in the last column of Table 1, are lauric (index 1), myristic (2), palmitic (3), stearic (4), palmitoleic (5), oleic (6), linoleic (7) and linolenic (8); this means that not all FAs in the biodiesel mixture are taken into consideration but (usually) only the above-mentioned ones (or even fewer), which also happen to be the most common. Obviously, when a ME consists primarily of FAs not represented in Equation (1), the prediction may be seriously affected.

- Models that are based on one or two (as in most cases) metrics of the FAME to predict its CN. As discussed in the previous section, the first is always the degree of unsaturation (DU), and the second the chain length (CL) or the molecular weight (MW). Such models are usually linear or quadratic. The average biodiesel CL and MW needed in these models are calculated as the weighted average based on each fatty acid's percentage in the biodiesel mixture. For the biodiesel average degree of unsaturation, the calculation is further based on the commonly-employed approach of accounting unsaturated FAs of the form Cxx:y with a weight percentage of y (e.g., palmitoleic C16:1 or oleic C18:1 with a weight percentage of 1, linoleic C18:2 with a percentage of 2, and linolenic C18:3 with 3—see column 'Number of double bonds' in Table 1). In contrast to the previous category, here by default, all FAs present in the biodiesel mixture are considered for the calculation of DU and CL/MW.
- Models that use each neat FAME's CN as the basis, applying then a mixing rule for the biodiesel 'mixture'.

$$CN = CN_{FAME} = \sum_i w_i CN_i \tag{2}$$

where CN_i is the predicted (or measured) cetane number of each neat FAME (methyl oleate, methyl palmitate etc), and w_i its weight percentage in the biodiesel blend. In this category too, all FAs present in the biodiesel mixture are taken into account. In Equation (2), the universally accepted assumption is applied that the CN of a blend is a linear combination of the CNs of its components.

In the next subsections, the gathered models will be discussed in more detail for each one of the categories mentioned above, before their comprehensive evaluation in Section 4.

3.1. Compositional Models

In 2008, Bamgboye and Hansen [18] published the first compositional-type CN model based on a large sample of experimental data from various sources; the relation reached was

$$CN = 61.1 + 0.088x_2 + 0.133x_3 + 0.152x_4 - 0.101x_5 - 0.039x_6 - 0.243x_7 - 0.395x_8 \tag{3}$$

Notice that the effect of lauric acid (index 1 in Table 1) is not taken into account. A closer look at the above relationship reveals the positive effect of saturated acids ($x_{2\div4}$) on the CN, as well as the negative effect of the unsaturated ones ($x_{5\div8}$). A coefficient of determination in the order of 88.3% was reported regarding the model's predictive capability against a set of nine experimental data series from the literature (different from the ones used to build Equation (3)).

Following a similar approach, Gopinath et al. [21] developed in 2009 a relation based on 68 experimental values of CN and fatty acid composition, and the MLR analysis provided the following equation:

$$CN = 62.2 + 0.017x_1 + 0.074x_2 + 0.115x_3 + 0.177x_4 - 0.103x_6 - 0.279x_7 - 0.366x_8 \tag{4}$$

The effect of lauric acid (index 1) is now included in the relation. In contrast, the effect of (the usually low-percentage) C16:1 palmitoleic acid (index 5) is not taken into account. Notice in Equation (4) that the weight of the saturated acids' ($x_{1\div4}$) regression coefficients increases with increasing molecular weight (i.e., with increasing index i). Similarly, the weight of the unsaturated acids' coefficients ($x_{6\div8}$) decreases with increasing degree of unsaturation. Thus, the theoretical basis of Equation (4) is unequivocally sound. A maximum error of 8% was reported when comparing the model's predictions with a data set of eleven experimental values.

Table 1. Common fatty acids found in biodiesel mixtures.

Cxx:y	Common Name	Formal Name	CAS	Chemical Formula	Carbon Atoms	Molecular Weight	Number of Double Bonds	Index for Compo-Sitional Models (Section 3.1)
8:0	Caprylic	Octanoic	124-07-2	$C_8H_{16}O_2$	8	144.21	0	—
10:0	Capric	Decanoic	334-48-5	$C_{10}H_{20}O_2$	10	172.27	0	—
12:0	Lauric	Dodecanoic	143-07-7	$C_{12}H_{24}O_2$	12	200.32	0	1
14:0	Myristic	Tetradecanoic	544-63-8	$C_{14}H_{28}O_2$	14	228.37	0	2
16:0	Palmitic	Hexadecanoic	57-10-3	$C_{16}H_{32}O_2$	16	256.42	0	3
18:0	Stearic	Octadecanoic	57-11-4	$C_{18}H_{36}O_2$	18	284.48	0	4
16:1	Palmitoleic	*cis*-9 hexadecenoic	373-49-9	$C_{16}H_{30}O_2$	16	254.41	1	5
18:1	Oleic	*cis*-9 octadecenoic	112-80-1	$C_{18}H_{34}O_2$	18	282.47	1	6
18:2	Linoleic	*cis*-9, *cis*-12 octadecadienoic	60-33-3	$C_{18}H_{32}O_2$	18	280.45	2	7
18:3	Linolenic	*cis*-9, *cis*-12, *cis*-15 octadecatrienoic	463-40-1	$C_{18}H_{30}O_2$	18	278.43	3	8
20:0	Arachidic	Eicosanoic	506-30-9	$C_{20}H_{40}O_2$	20	312.54	0	—
20:1	Gondoic	*cis*-11 Eicosenoic	5561-99-9	$C_{20}H_{38}O_2$	20	310.52	1	9
22:0	Behenic	Docosanoic	112-85-6	$C_{22}H_{44}O_2$	22	340.59	0	—
22:1	Erucic	*cis*-13 docosenoic	112-86-7	$C_{22}H_{42}O_2$	22	338.58	1	10
24:0	Lignoceric	Tetracosanoic	557-59-5	$C_{24}H_{48}O_2$	24	368.63	0	—

In 2013, Piloto-Rodriguez et al. [19] expanded on previous analyses by also including the effect of gondoic and erucic mono-unsaturated fatty acids in the derived correlation (both are usually of low percentage weight, with the exception of certain 'extreme' or exotic feedstocks such as jojoba). The sum of other FAs found in the biodiesel mixture ('res') was also taken into account. The derived equation, which is much more complex than the previous two, was based on a set of 50 experimental values of measured FA compositions and CNs, and reads as:

$$\begin{aligned} CN = 56.16 + 0.07x_1 + 0.1x_2 + 0.15x_3 + 0.23x_4 - 0.05x_5 - 0.03x_6 \\ -0.19x_7 - 0.31x_8 + 0.08x_9 + 0.18x_{10} - 0.1x_{res} \end{aligned} \quad (5)$$

In this case too, the always positive magnitude of the saturated acids' coefficients ($x_{1\div4}$) increases with increasing molecular weight. The 'res' contribution, which is expected to consist mostly of saturated short-chain (C8:0 and C10:0) and long-chain (C20:0 and C22:0) FAs, is, however, negative. The contribution of the unsaturated acids ($x_{5\div10}$) is mixed, with the major unsaturated acids contributing negatively and the two minor ones (C20:1 and C22:1) positively. A rather high degree of accuracy of the model's predictive capability was reported (R^2 in the order of 91.46%), and was claimed to be approximately five to six percentage points higher than the previous two models.

The fourth model following the 'compositional' philosophy and applying MLR analysis was developed by Giakoumis and Sarakatsanis in 2018 [22], and uses the following equation:

$$\begin{aligned} CN = 55.87 + 0.0747x_1 + 0.098x_2 + 0.164x_3 + 0.176x_4 \\ -0.050x_5 + 0.001x_6 - 0.140x_7 - 0.273x_8 \end{aligned} \quad (6)$$

The data set used for Equation (6) contained 45 different sets of CN and FA composition values from various feedstocks (incl. also neat FAMEs); the degree of correlation R^2 was approx. 90%, i.e., a quite a high value, despite the wide variability of the selected data set. The following comments can be made with respect to Equation (6):

- The first (constant) term is of the same order of magnitude as past similar research—see Equations (3)–(5);
- saturated acids contribute to an increase in the CN number;
- the magnitude of the saturated acids' coefficients increases with increasing carbon number 0.0747 (C12:0) to 0.098 (C14:0) to 0.164 (C16:0) to 0.176 (C18:0); and
- unsaturated acids contribute to a decrease in the predicted CN.

Thus, it is believed that the derived equation has a sound theoretical basis.

3.2. Models Based on the Average Degree of Unsaturation and Chain Length

In 2012, Hoekman et al. [12] performed a large-scale statistical analysis of the fatty acid composition and physical/chemical properties for biodiesels derived from twelve sources (camelina, canola, coconut, corn, jatropha, palm, rapeseed, safflower, soybean, sunflower, tallow and yellow grease). Among other things, correlations for CN were derived based on the degree of unsaturation or the chain length (but not both). For the former, the developed relation was

$$CN = 62.876 - 6.6684 \cdot DU \quad (7)$$

As expected, an increasing unsaturation contributes to a decrease in the predicted CN; the coefficient of determination from Equation (7) was reported as 80.49% against the experimental data its derivation was based on. No statistical correlation could be established between CN and chain length.

Around the same time, Giakoumis [6] performed an even broader survey, covering now 26 different feedstocks, with data from hundreds of literature sources. The feedstocks in alphabetical order were: beef tallow, canola, castor, chicken, coconut, corn, cottonseed, croton, fish, hazelnut,

jatropha, karanja, lard, linseed, mahua, neem, olive, palm, peanut, rapeseed, rice bran, rubber seed, safflower, soybean, sunflower and waste cooking oil. A similar to Equation (7) relation was reached (castor excluded) with an R^2 value of 79%, illustrated in the lower sub-diagram of Figure 2.

$$CN = 62.32 - 6.13 \cdot DU \tag{8}$$

Since the CN depends not only on the degree of unsaturation, but also on the molecular weight (or the chain length; see upper sub-diagram of Figure 2), it is not surprising that both Equations (7) and (8) did not present a very high degree of correlation with the experimental data they were based on.

Apart from the above two research groups, many more have investigated the effects of unsaturation on CN (as well as on other FAME properties), e.g., Ramos et al. [26]. Nonetheless, the latter investigations concerned a much smaller scale of examined feedstocks; hence, these correlations will not be included in the current comparative evaluation, as, in essence, they are 'covered' by Equations (7) and (8).

The rest of the models of this category attempted a more comprehensive approach, incorporating both DU and CL/MW effects. The obvious challenge here is to capture the effect of chain length (or molecular weight), as, following Figure 4, the effect of CL on CN, although increasing, does not exhibit a clear trend.

The simplest relation taking into account the degree of unsaturation (negative effect) and the average biodiesel carbon number (positive effect) was developed by Chang and Liu [27]

$$CN = 2.005 - 20.077 \cdot DU + 4.201 \cdot CL \tag{9}$$

Very little information is provided as to how this equation was developed. The authors mention that Equation (9) was validated with 31 experimental data points, manifesting an average relative deviation of 7.43%.

The next two models of this category were even more complex. In 2011, Pinzi et al. [16] published a second-order equation model based on both the degree of unsaturation (found to have the biggest contribution) and the chain length:

$$\begin{aligned} CN \;=\; & 46.6632 - 1.7357 \cdot CL \;+\; 12.3976 \cdot DU \;+\; 0.243275 \cdot CL^2 \\ & -2.64964 \cdot CL \cdot DU \;+\; 5.65655 \cdot DU^2 \end{aligned} \tag{10}$$

Equation (10) seems purely 'mathematical', without possessing a clear theoretical basis (notice the negative sign of the first-order CL coefficient and the positive sign of the first and second-order DU ones). Moreover, based on the rather obvious linear relation between CN and DU from Figure 2, one might reach the conclusion that a second-order DU term is redundant. The degree of unsaturation and the chain length of the biodiesel are computed based on its FA composition (weighted averages), and then the CN is derived from Equation (10); a very high R^2 value of 95.15% was reported.

Another similar approach was followed by Mishra et al. [28] in 2016, who developed the following relation based on 42 measured data sets of CN and FA composition:

$$CN = 63.41 - 0.073 \cdot DU + 0.035 \cdot SCSF - 3.26 \times 10^{-4} \cdot DU \cdot SCSF \tag{11}$$

In this equation, apart from the degree of unsaturation, the straight-chain saturation factor (SCSF) is taken into account as well. The latter metric is calculated as

$$SCSF = \frac{1}{100} \sum_{sat} MW_i \cdot w_i \tag{12}$$

with MW_i as the molecular weight of the individual saturated methyl ester. Essentially, parameter DU is employed to take into account the effect of unsaturated fatty acids (hence the negative sign

in Equation (11)), and SCSF the saturated ones (hence the positive sign in Equation (11)). Contrary to Equation (10), Equation (11) possesses a sounder theoretical basis. As is the usual case, it was claimed that the model was superior to others developed earlier, based on a relatively small sample of experimental data.

3.3. Models Based on the Individual Neat FAME's CN

In 1982, Klopfenstein [14] developed a model based on a series of experimental data to predict the cetane number for neat FAMEs (saturated esters with 8-24 carbon atoms plus esters of palmitoleic, oleic, linoleic and linolenic acids). The equation derived was:

$$CN_i = 58.1 + 2.8 \cdot \left(\frac{n_i - 8}{2}\right) - 15.9 \cdot db_i \tag{13}$$

where n_i is the number of carbon atoms of the corresponding FA. Using Equation (13), the cetane number for each pure FAME can be assessed; then, applying Equation (2) the CN of the biodiesel is evaluated.

The same researcher studied saturated neat methyl/ethyl/propyl/butyl esters and reached the following equation for methyl ester ones [29] based on measurements according to ASTM D 613 conducted on a Waukesha variable compression ratio engine:

$$CN_{i_sat} = -24.48 + 8.431 \cdot n_i - 0.1299 \cdot n_i^2 \tag{14}$$

Equation (14) was developed based on measurements for esters of the saturated fatty acids from C8 to C18. In the original citation [29], Equation (14) had a positive constant term sign, which, however, produces unreasonably high CNs. Reproducing the original procedure used by Klopfenstein to develop his equation from a figure he provided with experimental data, a negative constant term was derived, as is also confirmed in [11].

Freedman and Bagby [11] studied saturated neat FAMEs too and developed a similar to Equation (14) relation for esters of the saturated fatty acids C6, C10, C12:0, C14, C16 and C18. Again, measurements according to ASTM D 613 were used (notice the negative sign of the constant term)

$$CN_{i_sat} = -57.26 + 14.892 \cdot n_i - 0.4149 \cdot n_i^2 \tag{15}$$

Lapuerta et al. [17] studied the CN of various neat FAMEs too; they gathered 52 values from the literature for esters of the FAs C6:0, C6:2, C7:0, C8:0, C10:0, C12:0, C14:0, C16:0, C16:1, 18:0, C18:1, C18:2, C18:3, C20:0 and C22:1 acids, and developed correlations to predict them and inter-relate them with the iodine number. Initially, two equations were derived: one based on the CFR and another on the IQT data:

$$CN_i = -52.974 + \left(13.767 - 1.202 \cdot db_i + 0.152 \cdot db_i^2\right) \cdot n_i - 0.351 \cdot n_i^2 \text{ (CFR)}$$
$$CN_i = -14.092 + \left(6.704 - 2.013 \cdot db_i + 0.236 \cdot db_i^2\right) \cdot n_i - 0.038 \cdot n_i^2 \text{ (IQT)}$$

Their predictions were found to be quite close when saturated esters were being tested, but differences were noticed for unsaturated FAMEs. Furthermore, a new correlation was formed based on all available data (most of which were of the IQT type)

$$CN_i = -21.157 + \left(7.965 - 1.785 \cdot db_i + 0.235 \cdot db_i^2\right) \cdot n_i - 0.099 \cdot n_i^2 \tag{16}$$

The value of R^2 was reported at 91.8% and again the constant term is negative. As is the case with all models of this category, Equation (2) is used to calculate the FAME's cetane number from the individual pure FAMEs weight percentage and CN. An interesting aspect of the particular analysis

was that it was also extended to neat ethyl, propyl and butyl esters. In general, small differences were observed between the FA methyl ester, FA ethyl ester, FA propyl ester and FA butyl ester cetane numbers, when the number of carbon atoms was higher than 14.

In 2012, Ramirez-Verduzco et al. [15] developed another relation for each neat FAME's CN (density, higher heating value and kinematic viscosity as well) based on the FAME's molecular weight and the number of double bonds (again with a negative constant term; note that Equation (17) has a similar structure with Equation (9)):

$$CN_i = -7.8 + 0.302 \cdot MW_i - 20 \cdot db_i \tag{17}$$

Another research group that based their analysis on pure FAMEs' CN was Tong et al. [30]; in this case, however, only one property was employed, namely, the number of carbon atoms. The sample used was not detailed enough (it seems that for most of the CNs, only one experimental value was used from the literature). Unlike the previous correlations of this category, Tong et al. [30] derived two polynomial equations from the collected data regarding the CN of pure FAMEs: one for saturated and one for unsaturated FAMEs (incorporating in this way DU effects).

$$\begin{gathered} CN_i = -107.71 + 31.126 \cdot n_i - 2.042 \cdot n_i^2 + 0.0499 \cdot n_i^3 \text{ (saturated FAMEs)} \\ CN_i = 109.000 - 9.292 \cdot n_i + 0.354 \cdot n_i^2 \text{ (unsaturated FAMEs)} \end{gathered} \tag{18}$$

It should be pointed out that for the unsaturated FAMEs relation, only neat FAMEs containing one double bond were used for Equation (18) (i.e., C16:1, C18:1 and C20:1). Having CN values in hand for each pure FAME, one would expect the application of Equation (2) for the calculation of the biodiesel CN. Instead, the differently-structured Equation (19) was proposed by the authors; it was developed from a multiple linear regression analysis based on 40 data series of fatty acid composition and CN.

$$CN = 1.068 \sum_i CN_i \cdot w_i - 6.747 \tag{19}$$

It was claimed that this equation manifests a coefficient of determination R^2 = 90.58% against the data it was based on, as well as a very small relative error in the order of 3% (at the maximum). Obviously, for neat FAMEs, Equation (19) produces different results from Equation (18).

Since all the models mentioned in this subsection are based on experimental neat FAMEs' CN values, another 'model' was formulated for the purposes of this study, following the logic suggested by Clements [31] and Knothe [10]. Instead of using equations to predict each pure FAME's CN, the actual experimental (average) values are used, applying then the mixing rule Equation (2) for the whole biodiesel mixture. CN values for pure FAMEs, following either the CFR or the IQT procedure, were gathered from Refs [7,10,11,13,25,29,32–42], and are reproduced in Table 2; Reference [25], in particular, is an invaluable citation for many kinds of pure compounds' cetane numbers including (methyl) esters.

Experimental CN values are provided in Figure 3 (together with the respective standard deviation) in comparison to the predicted ones from all models of this subsection. The well-established trends of increasing CN with increasing chain length (in a non-linear manner however [29,43]), and decreasing the number of double bonds, is again made evident. As noticed, a fairly high variability in the predicted neat FAMEs' CN values is observed (in the experimental too), which will later be reflected into different predictions from the models (Section 4). The model of Klopfenstein [14] seems to overestimate the CN of methyl decanoate and underestimate CN values for long-chain FAMEs (C18 and above); that of Tong et al. [30] assigns the same values to methyl esters of C18:1, C18:2 and C18:3, and, most probably, overestimates the CN for methyl lignocerate. On the other hand, the model whose predictions lie most closely to the experimental average values is undoubtedly that of Lapuerta et al. [17]; this holds true for both the saturated and unsaturated FAMEs. Thus, it can be rather safely said that Equation (16) is the one that provides the best possible prediction of pure FAME CNs.

Table 2. Experimental CN values for pure FAMEs (CNs in bold correspond to the CFR method and in italics to the IQT method).

FAME	Chemical Formula	CAS	Molecular Weight	Experimental CN Values	Average CN	Standard Deviation
Octanoate	$C_9H_{18}O_2$	111-11-5	158.24	**33.6** [29]; *39.8* [7]; 34 [32]	35.8	3.47
Decanoate	$C_{11}H_{22}O_2$	110-42-9	186.29	**47.2** [29]; **47.9** [11]; *52.7* [25]; *50.7* [25]; *51.6* [7]; *54.1* [25]; 52.1 [33]; *51.6* [7]	51.0	2.35
Laurate	$C_{13}H_{26}O_2$	111-82-0	214.35	**61.4** [29]; **60.8** [11]; **61.2** [13]; 54 [34]; 70 [35]; *66.7* [7]; *66.3* [25]; *66.7* [25]; 60.4 [42]	63.1	4.82
Myristate	$C_{15}H_{30}O_2$	124-10-7	242.40	**66.2** [29]; **73.5** [11]; *72* [35]; *75.8* [36]	71.9	4.09
Palmitate	$C_{17}H_{34}O_2$	112-39-0	270.45	**74.5** [29]; **74.3** [13]; 91 [37]; *85.9* [7]; 80 [35]; **74.3** [11]; 86 [38]; 88 [42]	81.8	6.83
Stearate	$C_{19}H_{38}O_2$	112-61-8	298.51	**86.9** [13]; **75.6** [11]; 100 [37]; 101 [38];81 [35]; 87 [29]; *95.6* [25]	89.6	9.64
Arachidate	$C_{21}H_{42}O_2$	1120-28-1	326.56	100 [37]	100	-
Palmitoleate	$C_{17}H_{32}O_2$	1120-25-8	268.44	51 [38]; *56.6* [7]	53.8	3.96
Oleate	$C_{19}H_{36}O_2$	112-62-9	296.49	80 [37]; 55 [39]; **56** [13]; *59.8* [25]; *56.6* [7]; *59.3* [7]; 71 [35]; 53 [32]; 59 [38]	61.1	8.76
Ricinoleate	$C_{19}H_{36}O_3$	141-24-2	312.49	*37.4* [7]	37.4	-
Linoleate	$C_{19}H_{34}O_2$	112-63-0	294.48	**41.7** [13]; *43.9* [25]; *38.2* [7]; 38 [38]; 42 [39]; 43 [32]	41.1	2.48
Linolenate	$C_{19}H_{32}O_2$	301-00-8	292.46	*22.7* [10]; **45.9** [13]; 23 [39]; *37* [25]; *29.2* [41]	31.6	9.91
Gondoate	$C_{21}H_{40}O_2$	2390-09-2	324.54	*73.2* [40]	73.2	-
Erucate	$C_{23}H_{44}O_2$	1120-34-9	352.60	*74.2* [25]	74.2	-

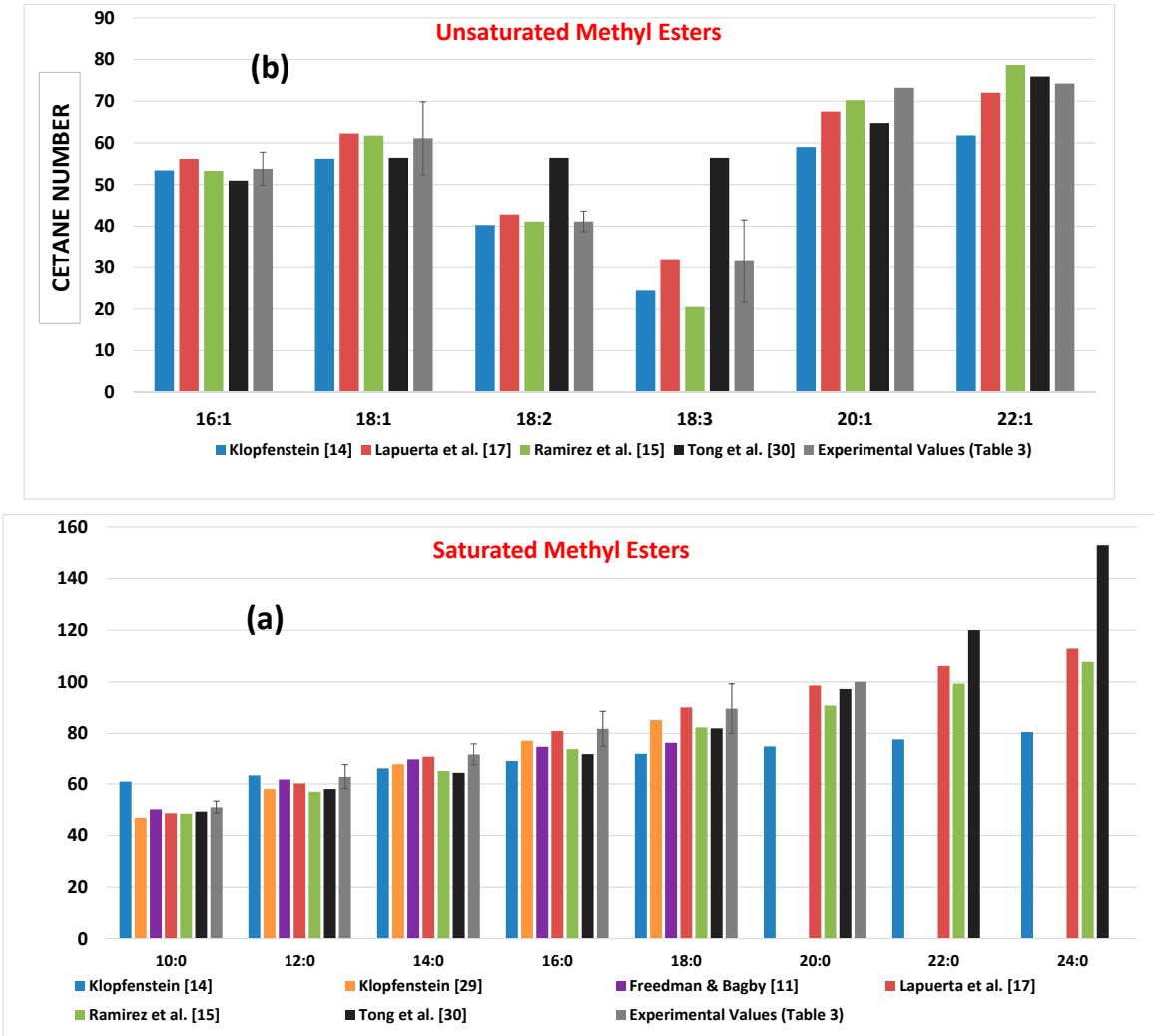

Figure 3. Predicted cetane numbers of pure FAMEs from the equations of Section 3.3, compared to average experimental values from the literature, showing also standard deviations for the experimental values: (**a**) saturated methyl esters; (**b**) unsaturated methyl esters.

In essence, the models of this section suffer from the (sometimes) significant variability in the experimental neat methyl esters CN values they rely upon. Yet, they are perhaps the only ones that have been calibrated based on 'true' experimental measurements.

Concluding this section, Table 3 summarizes basic information together with the developed equations for each biodiesel CN model found in the literature. Note in Table 2 that for the models of the current subsection, all researchers use the FA carbon number ni in their equations, even when referring to the corresponding FAME.

Unlike all the other relations mentioned in Section 3, Equation (19) from the model of Tong et al. [30] could not be confirmed by the current research group using the data provided in the original paper. Therefore, this model will be excluded from the comparative analysis of Section 4. Obviously, Equations (14) and (15) that deal with saturated FAMEs only will not be included in the comparative analysis as well.

Table 3. Summary of methyl esters CN predictive correlations based on the FA composition (R^2 and standard error values refer to the degree of correlation of the developed equation with the data its derivation was based on, as provided in the original citation).

	Research Group	Ref.	Year	Equation	Comments
Compositional Models (Section 3.1) *(Index: 1: lauric; 2: myristic; 3: palmitic; 4: stearic; 5: palmitoleic; 6: oleic; 7: linoleic; 8: linolenic; 9: gondoic; 10: erucic)*					
1	Bamgboye and Hansen	[18]	2008	$CN = 61.1 + 0.088x_2 + 0.133x_3 + 0.152x_4 - 0.101x_5 - 0.039x_6 - 0.243x_7 - 0.395x_8$	
2	Gopinath et al.	[21]	2009	$CN = 62.2 + 0.017x_1 + 0.074x_2 + 0.115x_3 + 0.177x_4 - 0.103x_6 - 0.279x_7 - 0.366x_8$	R^2 = 95.3%, St. dev.=2.27
3	Piloto and Rodriguez et al.	[19]	2013	$CN = 56.16 + 0.07x_1 + 0.1x_2 + 0.15x_3 + 0.23x_4 - 0.05x_5 - 0.03x_6 - 0.19x_7 - 0.31x_8 + 0.08x_9 + 0.18x_{10} - 0.1x_{res}$	R^2 = 91.1%, St. error=4.6
4	Giakoumis and Sarakatsanis	[22]	2018	$CN = 55.87 + 0.0747x_1 + 0.098x_2 + 0.164x_3 + 0.176x_4 - 0.050x_5 + 0.001x_6 - 0.140x_7 - 0.273x_8$	R^2 = 89.6%, St. error=3.04
Models based on the Average Methyl Ester Degree of Unsaturation and Chain Length (Section 3.2*)*					
5	Hoekman et al.	[12]	2012	$CN = 62.876 - 6.6684 \cdot DU$	Based on a large survey of experimental values for 12 vegetable and animal feedstocks; R^2 = 80.5%
6	Giakoumis	[6]	2013	$CN = 62.32 - 6.13 \cdot DU$	Based on a large survey of experimental values for 26 vegetable and animal feedstocks; R^2 = 79%
7	Pinzi et al.	[16]	2011	$CN = 46.6632 - 1.7357 \cdot CL + 12.3976 \cdot DU + 0.243275 \cdot CL^2 - 2.64964 \cdot CL \cdot DU + 5.65655 \cdot DU^2$	R^2 = 95.2%
8	Chang and Liu	[27]	2010	$CN = 2.005 - 20.077 \cdot DU + 4.201 \cdot CL$	
9	Mishra et al.	[28]	2016	$CN = 63.41 - 0.073 \cdot DU + 0.035 \cdot SCSF - 3.26 \times 10^{-4} \cdot DU \cdot SCSF$ $SCSF = \frac{1}{100} \sum_{sat} MW_i \cdot w_i$	

Table 3. *Cont.*

	Research Group	Ref.	Year	Equation	Comments
				Models based on the Neat FAME's Cetane Number applying a Mixing Rule for the Whole FAME (Section 3.3*)* $CN = \sum_i CN_i w_i$	
10	Klopfenstein	[14]	1982	$CN_i = 58.1 + 2.8 \cdot \left(\frac{n_i - 8}{2}\right) - 15.9 \cdot db_i$	
11	Klopfenstein	[29]	1985	$CN_{i_sat} = -24.48 + 8.431 \cdot n_i - 0.1299 \cdot n_i^2$	Saturated FAMEs C8 to C18 only (missing negative sign of constant term in the original citation)
12	Freedman and Bagby	[11]	1990	$CN_{i_sat} = -57.26 + 14.892 \cdot n_i - 0.4149 \cdot n_i^2$	Saturated FAMEs C6 to C18 only
13	Lapuerta et al.	[17]	2009	$CN_i = -21.157 + \left(7.965 - 1.785 \cdot db_i + 0.235 \cdot db_i^2\right) \cdot n_i - 0.099 \cdot n_i^2$	Developed also equations for ethyl/propyl and butyl esters, as well as separate equations for methyl esters based on the CFR or IQT data; R^2 = 91.8%
14	Ramirez-Verduzco et al.	[15]	2012	$CN_i = -7.8 + 0.302 \cdot MW_i - 20 \cdot db_i$	Developed similar relations for density, higher heating value and kinematic viscosity; MW_i corresponds to FAME molecular weight
15	Tong et al.	[30]	2011	$CN_i = -107.71 + 31.126 \cdot n_i - 2.042 \cdot n_i^2 + 0.0499 \cdot n_i^3$ (saturated FAMEs) $CN_i = 109.000 - 9.292 \cdot n_i + 0.354 \cdot n_i^2$ (unsaturated FAMEs) $CN = 1.068 \sum_i (CN_i \cdot w_i) - 6.747$	R^2 = 90.6%

4. Comparative Evaluation of All Models' Predictive Capability

A general comment should be made first regarding the procedure followed for the comparative evaluation of all models. Picking and choosing experimental data can easily manipulate the results and lead to whatever findings one wants to reach. In order to avoid such phenomena and provide a reliable and objective comparison:

(a) An extensive, and at the same time, quite broad in terms of DU, CL and CN values, set of experimental data was selected for the comparison of 50 series of FA-CN values in total. It is believed that the amount of experimental data, as well as its variability, is adequate to establish trends and reach some reliable conclusions. All values of CN refer to the cetane number, and not the cetane index.

(b) Data was chosen only from those sources where values for both CN and fatty acid composition were provided, having been measured using the universally-accepted methods (e.g., ASTM D7806 for fatty acid composition and D613 for CN) [18,19,26,44–58]. Of course, this does not guarantee that all reported values are undeniably correct, but it indicates an acceptable level of confidence.

(c) All selected data from the literature values was compared to the average values provided in [6], and it was confirmed that it was indeed within the 'acceptable' limits with regards to both FA composition and CN.

(d) Combinations of FAs-CNs that resulted in large errors in at least half of the models (from all three categories) were not considered in the analysis, particularly if these reported FA and/or CN data were outside the standard values reported in the two earlier extensive statistical analyses [6,12].

(e) Extreme cases of methyl esters rich in rare and unusual FAs, e.g., ricinoleic acid, were not included in the comparison, as they cannot be considered representative.

Table 4 (at the end of the text) provides the details for the 50 experimental data series, with percentage weight composition for all reported fatty acids, along with the measured CNs. Moreover, this table lists for each FAME the (calculated) degree of unsaturation and chain length needed for many of the models. The CNs in Table 4 range from 47 to 63.6, with the variability in FAs being rather broad as evidenced from the DU, which ranges from 0.20 (palm kernel oil) to 1.71 (camelina). For the whole data set, all values for CN, DU and CL are illustrated in Figure 4. Notice also in Table 4 the very low percentage weight of the myristic and palmitoleic acids, as well as of all acids with more than 18 carbon atoms.

Table 4. Experimental data for the evaluation of all models' predictive capability (fatty acid compositions and CNs have been reproduced with the accuracy provided in the original citation).

		Caprylic	Capric	Lauric	Myristic	Palmitic	Stearic	Palmitoleic	Oleic	Linoleic	Linolenic	Arachidic	Behenic	Gondoic	Erucic	Lignoceric		Degree of Unsaturation	Chain Length	Cetane Number
	FAME/Ref.	**8:0**	**10:0**	**12:0**	**14:0**	**16:0**	**18:0**	**16:1**	**18:1**	**18:2**	**18:3**	**20:0**	**22:0**	**20:1**	**22:1**	**24:0**	**Total (%)**			
1	SME [18]				0.06	10.64	3.88	0.14	32.38	46.36	5.53						98.99	1.42	17.60	**50.9**
2	PME [18]			0.29	0.87	43.08	4.31	0.12	40.55	9.64	0.32						99.18	0.61	16.94	**62.7**
3	TME [18]			0.08	1.3	23.88	17.88	0	45.25	2.85	0.53						91.77	0.53	15.98	**61.9**
4	COME [18]			0.02	0.32	22.05	2.17	0.13	16.13	55.72	0.25						96.79	1.28	16.96	**52.0**
5	PEME [18]					11.1	4.22		47.23	32.14	0.68						95.37	1.14	16.94	**54.2**
6	LME [18]			0.07	1	26.03	15		45.43	9.87	0.5						97.9	0.67	17.06	**63.6**
7	RME [18]				0.02	4.06	1.2	0.04	63.12	21.28	8.63						98.35	1.32	17.62	**52.8**
8	SFME [19]				0	6	2.9	0.1	17	74	0						100	1.65	17.88	**49.0**
9	WPME [19]				1	39	4.3	0.2	43.7	10.5	0.2			0.2			99.1	0.66	17.02	**60.4**
10	PEME [19]				0.1	8	1.8	0	53.3	28.4	0.3			2.4			94.3	1.13	16.86	**53.0**
11	JME [54]				0.1	14.2	7.1	0	43.2	34.9	0.2	0.2	0	0.1	0	0.1	100.1	1.14	17.74	**54.0**
12	JME [51]	0	0	0	0	14.2	6.9	1.4	43.1	34.4	0	0	0	0	0	0	100	1.13	17.69	**57.1**
13	PGME [51]					9.8	6.2	0	72.2	11.8	0	0					100	0.96	17.80	**55.1**
14	SME [53]					8.8	4.55	0.09	24.16	52.67	7.74	0.39	0.41	0.23	0.01	0.13	99.18	1.53	17.71	**47.7**
15	RME [44]				0	7.1	2.2	0	58.4	21.4	7.5						96.6	1.24	17.25	**54.0**
16	RME [52]	0	0	0	0.1	4.6	1.8	0.3	60.7	19.1	8.3	0.6	0.3	1.4	0.3	0.1	97.6	1.26	17.54	**52.2**
17	SME [45]					11.7	3.97		21.27	53.7	8.12	1.23					99.99	1.53	17.79	**51.3**
18	PEME [45]					17.2	2.7		40.5	36.6	0.5	0.9	1.5				99.9	1.15	17.72	**54.0**
19	CRME [45]					11.4	1.3		27.1	60.2	0	0					100	1.48	17.77	**55.4**
20	SFME [45]					4.9	2.3		32.6	59.4	0	0	0.5				99.7	1.51	17.87	**51.6**
21	RME [45]					5.2	1.4		66	18.9	5.6	1.9	1				100	1.21	17.97	**54.5**
22	PME [45]			0.5	1.6	49.8	2.9		38.6	6.6							100	0.52	16.91	**62.0**
23	PKME [45]	3.6	3.1	48	14.7	11.5	1.4	0	15.9	1.8							100	0.20	13.69	**62.1**
24	WFME [45]			1.6	1.5	27.3	4.9		36.1	25.7	1.9						99.0	0.93	17.12	**55.0**
25	SME [47]				0.1	11.3	4.6	0.2	23	52.8	7.7	0.3					100	1.52	17.77	**51.0**
26	SME [48]				0	10.56	4.74		22.51	52.39	8.22	0.36					98.78	1.52	17.58	**51.5**
27	SME [55]				0.09	10.54	3.89	0.13	23.18	48.92	8.24	1.71	1.88	0.71	0.18		99.47	1.47	17.82	**47.2**
28	CAME [55]				0.07	5.25	2.69	0.22	58.09	21.79	7.04	1.04	1.17	1.69	0.37		99.42	1.25	17.90	**55.0**
29	LME [55]			0.12	1.86	24.58	15.28	2.8	38.32	13.51	1.03	0.45	1.04	0.77	0.06		99.82	0.72	17.41	**63.6**
30	TME1 [55]			0.06	2.91	25.91	22.11	3.44	40.23	2.82	0.65	0.29	0.41	1.02	0.09		99.94	0.52	17.33	**62.9**
31	TME2 [55]			0.08	2.13	23.93	21.88	2.79	38.54	6.67	0.84	0.34	0.68	0.64	0.06		98.58	0.58	17.17	**61.7**
32	YGME [55]					11.53	13.36	0.18	60.67	12.64		0.41	0.81	0.21			99.81	0.86	17.78	**57.8**
33	SME [42]					11	4.13		25.12	53.37	6.35						99.97	1.51	17.77	**54.0**
34	Camelina [56]				0.1	6.8	2.7		19.7	19.6	32.6	1.5	0.2	12.4	2.3		97.9	1.71	17.86	**52.8**
35	Coriander [57]					5.3	3.1	0.3	77.1	13							98.8	1.03	17.67	**53.3**
36	Macadamia [58]				0.58	8.25	3.55	15.39	61.09	1.86		2.94		2.55			96.21	0.83	16.93	**57.5**
37	SME [50]					11.3	4.5		23.4	52.1	7.2						98.5	1.49	17.50	**51.8**
38	RSME [46]		0			19.64	5.47		27.82	35.17	11.89						99.99	1.34	17.61	**52.5**
39	PME [26]		0	0.1	0.7	36.7	6.6	0.1	46.1	8.6	0.3	0.4	0.1	0.2		0.1	100	0.65	17.25	**61.0**
40	OME [26]					11.6	3.1	1	75	7.8	0.6	0.3	0.1			0.5	100	0.93	17.79	**57.0**
41	PEME [26]				0.1	8	1.8		53.3	28.4	0.3	0.9	3	2.4		1.8	100	1.13	18.13	**53.0**
42	RME [26]					4.9	1.6		33	20.4	7.9			9.3	23		100.1	1.30	19.03	**55.0**

Table 4. *Cont.*

		Caprylic	Capric	Lauric	Myristic	Palmitic	Stearic	Palmitoleic	Oleic	Linoleic	Linolenic	Arachidic	Behenic	Gondoic	Erucic	Lignoceric		Degree of Unsaturation	Chain Length	Cetane Number
	FAME/Ref.	**8:0**	**10:0**	**12:0**	**14:0**	**16:0**	**18:0**	**16:1**	**18:1**	**18:2**	**18:3**	**20:0**	**22:0**	**20:1**	**22:1**	**24:0**	**Total (%)**			
43	SME [26]					11.3	3.6	0.1	24.9	53	6.1	0.3		0.3	0.3	0.1	100	1.50	17.80	**49.0**
44	SFME [26]					6.2	3.7	0.1	25.2	63.1	0.2	0.3	0.7	0.2	0.1	0.2	100	1.52	17.93	**50.0**
45	GME [26]				0.1	6.9	4	0.1	19	69.1	0.3						99.8	1.58	17.83	**48.0**
46	HO SFME [26]					4.6	3.4	0.1	62.8	27.5	0.1	0.3	0.7			0.3	99.8	1.18	17.92	**53.0**
47	AME [26]					10.4	2.9	0.5	77.1	7.6	0.8	0.3	0.1			0.2	99.9	0.95	17.79	**57.0**
48	CRME [26]					6.5	1.4	0.6	65.6	25.2	0.1	0.1		0.1	0.1	0.1	99.8	1.17	17.84	**53.0**
49	SME [49]				0	9	4		23	51	13						100	1.64	17.82	**49.0**
50	SFME [49]				0	9	7		10	74							100	1.58	17.82	**47.0**
	Avg. Values	0.90	0.52	3.64	1.01	14.05	5.25	0.93	40.75	31.01	4.05	0.62	0.70	1.75	1.79	0.28	99.00	1.15	17.49	**77.98**
	-	**8:0**	**10:0**	**12:0**	**14:0**	**16:0**	**18:0**	**16:1**	**18:1**	**18:2**	**18:3**	**20:0**	**22:0**	**20:1**	**22:1**	**24:0**	**Total (%)**	**DU**	**CL**	**CN**

(AME: almond; CAME: canola; COME: cottonseed; CRME: corn; EMME: egushi mellon; GME: grape; JME: jatropha; LME: lard; OME: olive; PEME: peanut; PGME: pongamia; PKME: palm kernel; PME: palm; RME: rapeseed; RSME: rubber seed; SFME: sunflower; SME: soybean; TME: tallow; WFME: waste frying; WPME: waste palm; YGME: yellow grease).

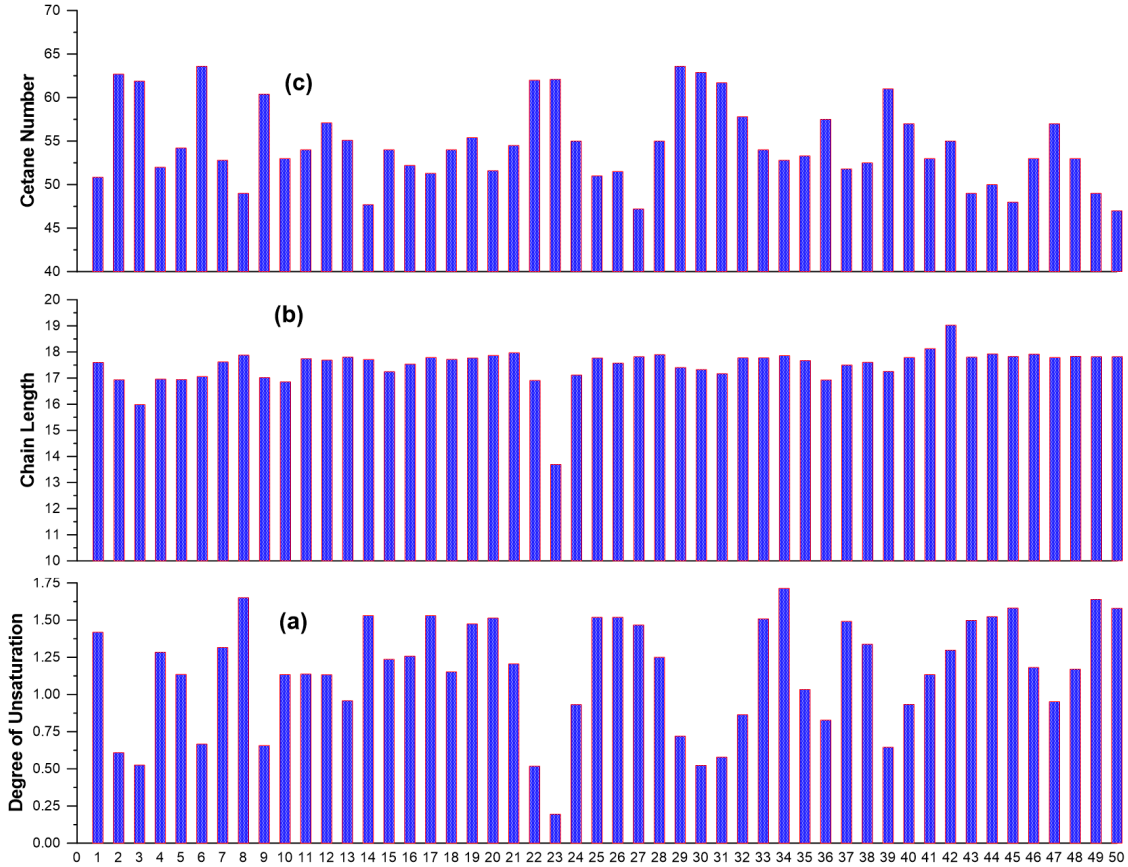

Figure 4. Degree of unsaturation (**a**), chain length (**b**) and cetane number (**c**) for the 50 methyl esters in the data set (CN values are experimental and are also provided in Table 4; DU and CL represent weighted averages based on the experimentally obtained FA compositions. For the DU, the calculation is further based on the usually employed approach of accounting unsaturated FAs of the form Cxx:y with a weight percentage of y).

Applying the equations provided in Section 3 for all models, Table 5 (at the end of the text) provides the predicted CN values together with the absolute error (%) for each model and for each one of the 50 experimental data lines. The average error from each model is indicated in the last row of Table 5, and further illustrated in Figure 5. The lowest average errors are reported by the models by Mishra et al. [28] and Giakoumis and Sarakatsanis [22] at 2.97%. The two models that are based on the degree of unsaturation (Equations (7) and (8)) both manifest rather low average errors (4.05 and 4.29% respectively). The average errors were very low from the models by Klopfenstein [14], at 4.34%, and Bamgboye and Hansen [18], at 4.78%. On the other hand, the highest average errors were observed from the models by Pinzi et al. [16] (9.27%) and by Lapuerta et al. [17] (approx. 8%). Regarding the maximum error from the predictions, the model that 'excels' here is again the one by Mishra et al. [28] at 10.21%, followed by the two DU-only models, at 12.48 [12] and 12.97% [6] respectively.

Table 5. Comparison between experimental and computed CN values, and absolute errors for all models.

	FAME	Experimental CN	Bamgboye, Hansen [18]		Piloto et al. [19]		Gopinath et al. [21]		Giakoumis Sarakatsanis [22]		Hoekman et al. [12]		Giakoumis [6]		Pinzi et al. [16]		Chang and Liu [27]		Mishra et al. [28]		Klopfen-stein [14]		Lapuerta et al. [17]		Ramirez et al. [15]		'Experi-Mental'	
			Pred. Value	Error (%)	Pred. Value	Error (%)	Pred. Value	Error (%)	Pred. Value	Error (%)	Pred. Value	Error (%)	Pred. Value	Error (%)	Pred. Value	Error (%)	Pred. Value	Error (%)	Pred. Value	Error (%)	Pred. Value	Error (%)	Pred. Value	Error (%)	Pred. Value	Error (%)	Pred. Value	Error (%)
1	SME [18]	**50.9**	48.4	4.9	47.2	7.3	45.8	9.9	50.3	1.0	53.4	5.1	53.6	5.5	54.3	6.8	47.5	6.7	52.6	3.4	48.5	4.6	54.0	6.2	51.4	1.0	54.3	6.3
2	PME [18]	**62.7**	63.5	1.3	60.6	3.4	61.0	2.7	62.4	0.5	58.8	6.2	58.6	6.6	69.4	10.6	60.9	2.8	61.0	2.8	60.5	3.4	69.1	10.2	65.2	4.1	68.9	9.0
3	TME [18]	**61.9**	64.4	4.1	61.9	0.0	62.6	1.1	62.6	1.1	59.4	4.1	59.1	4.5	66.9	8.1	58.6	5.3	61.7	0.3	57.1	7.8	66.0	6.6	62.5	0.9	65.6	5.7
4	COME [18]	**52.0**	50.1	3.6	48.8	6.1	47.8	8.0	52.0	0.1	54.3	4.4	54.4	4.7	54.8	5.3	47.5	8.7	53.6	3.0	48.7	6.3	54.1	4.0	51.3	1.4	53.2	2.2
5	PEME [18]	**54.2**	53.3	1.7	51.1	5.8	50.1	7.5	53.8	0.8	55.3	2.0	55.4	2.1	57.5	6.1	50.4	7.0	55.0	1.5	50.4	7.0	56.2	3.6	54.2	0.0	55.3	2.0
6	LME [18]	**63.6**	62.6	1.6	60.2	5.3	60.3	5.2	61.4	3.6	58.4	8.1	58.2	8.4	68.5	7.7	60.3	5.2	60.1	5.5	59.2	6.9	68.0	6.9	64.5	1.4	67.6	5.9
7	RME [18]	**52.8**	50.8	3.8	48.4	8.3	47.3	10.4	51.5	2.6	54.1	2.5	54.3	2.8	56.3	6.6	49.6	6.1	53.7	1.7	49.9	5.6	55.5	5.2	53.5	1.4	56.6	6.7
8	SFME [19]	**49.0**	43.7	10.8	43.2	11.9	41.0	16.3	47.0	4.2	51.9	5.8	52.2	6.5	51.1	4.2	44.0	10.3	50.9	3.9	45.7	6.8	49.8	1.6	47.8	2.4	48.4	1.3
9	WPME [19]	**60.4**	62.7	3.8	59.7	1.2	60.0	0.6	61.6	2.0	58.5	3.2	58.3	3.5	68.5	13.5	60.3	0.2	60.3	0.2	59.9	0.9	68.2	12.8	64.6	6.9	67.7	10.8
10	PEME [19]	**53.0**	53.3	0.7	50.5	4.8	49.9	5.8	53.5	0.9	55.3	4.4	55.4	4.5	57.2	8.0	50.1	5.6	55.1	3.9	49.8	6.0	55.2	4.2	53.8	1.5	52.6	0.7
11	JME [54]	**54.0**	53.8	0.3	51.9	3.8	50.8	5.9	54.6	1.0	55.3	2.4	55.4	2.5	60.4	11.9	53.7	0.5	55.0	1.8	53.7	0.5	60.1	11.3	57.8	7.1	59.1	8.6
12	JME [51]	**57.1**	53.9	5.7	52.0	9.0	51.0	10.7	54.6	4.6	55.3	3.1	55.4	3.0	60.3	5.6	53.6	6.2	55.0	3.6	53.6	6.0	60.0	5.2	57.7	1.0	59.0	3.2
13	PGME [51]	**55.1**	57.7	4.7	54.6	0.8	53.7	2.5	57.0	3.3	56.5	2.5	56.4	2.4	64.8	17.5	57.6	4.5	56.6	2.7	56.6	2.7	63.5	15.3	61.8	12.1	62.5	11.9
14	SME [53]	**47.7**	46.2	3.2	45.3	4.9	44.0	7.8	48.6	1.9	52.7	10.4	52.9	11.0	52.6	10.3	45.7	4.2	51.6	8.2	47.0	1.5	52.3	9.6	49.5	3.8	52.5	9.2
15	RME [44]	**54.0**	51.9	3.8	49.6	8.2	48.7	9.9	52.4	3.0	54.6	1.2	54.7	1.4	56.6	4.7	49.6	8.1	54.2	0.4	49.8	7.8	55.6	3.0	53.5	1.0	56.5	4.4
16	RME [52]	**52.2**	51.7	1.0	49.2	5.8	48.4	7.2	52.1	0.3	54.5	4.4	54.6	4.6	57.1	9.4	50.4	3.4	54.1	3.6	50.3	3.6	56.2	7.8	54.3	4.0	56.0	6.9
17	SME [45]	**51.3**	46.2	10.0	45.6	11.2	44.1	14.0	48.8	5.2	52.7	2.7	52.9	3.2	52.9	3.0	46.0	10.3	51.5	0.5	47.5	7.5	53.1	3.4	49.9	2.7	54.0	5.0
18	PEME [45]	**54.0**	53.1	1.6	51.0	5.6	50.1	7.2	53.9	0.1	55.2	2.2	55.3	2.3	60.0	11.1	53.3	1.3	54.8	1.6	53.3	1.2	59.9	10.9	57.4	6.3	57.5	6.0
19	CRME [45]	**55.4**	47.1	14.9	45.9	17.1	44.2	20.3	49.6	11.8	53.0	4.3	53.3	3.8	53.8	2.9	47.1	15.1	52.2	5.8	48.3	12.8	53.0	4.3	51.0	8.0	51.8	7.0
20	SFME [45]	**51.6**	46.4	10.1	45.1	12.6	43.2	16.2	48.8	5.7	52.8	2.3	53.0	2.8	53.4	3.4	46.7	9.6	52.0	0.9	47.7	7.6	52.3	1.3	50.6	2.0	50.4	2.4
21	RME [45]	**54.5**	52.6	3.4	50.0	8.2	48.9	10.2	52.9	3.1	54.8	0.6	54.9	0.8	59.8	9.7	53.3	2.2	54.5	0.0	52.9	3.0	59.4	8.9	57.4	5.3	58.7	7.1
22	PME [45]	**62.0**	65.2	5.2	62.1	0.1	62.8	1.2	63.9	2.9	59.4	4.2	59.1	4.6	71.6	15.5	62.6	1.0	62.3	0.5	62.3	0.5	71.2	14.8	67.1	8.2	71.1	12.8
23	PKME [45]	**62.1**	63.1	1.6	61.5	0.9	63.5	2.3	62.8	1.1	61.6	0.8	61.1	1.6	64.1	3.2	55.6	10.4	67.3	8.4	63.0	1.4	63.4	2.0	60.1	3.2	64.8	4.2
24	WFME [45]	**55.0**	57.2	4.0	55.1	0.2	54.8	0.4	57.4	4.2	56.7	3.0	56.6	2.9	62.4	13.5	55.2	0.4	57.0	3.7	55.6	1.0	62.6	13.8	59.3	7.9	62.5	12.0
25	SME [47]	**51.0**	46.5	8.8	45.8	10.1	44.4	12.9	49.1	4.0	52.7	3.4	53.0	3.9	53.0	3.9	46.2	9.5	51.7	1.3	47.6	6.6	53.1	4.2	50.1	1.8	54.0	5.5
26	SME [48]	**51.5**	46.4	10.0	45.7	11.3	44.3	14.0	48.9	5.4	52.7	2.4	53.0	2.9	52.4	1.8	45.3	12.0	51.7	0.3	46.8	9.2	52.2	1.4	49.2	4.5	53.2	3.2
27	SME [55]	**47.2**	47.0	0.3	46.0	2.5	45.1	4.5	49.2	4.1	53.1	12.5	53.3	13.0	54.1	14.5	47.4	0.4	52.0	10.2	48.3	2.3	54.4	15.4	51.3	8.7	53.0	11.0
28	CAME [55]	**55.0**	51.8	5.7	49.4	10.3	48.6	11.6	52.3	5.2	54.5	0.8	54.7	0.6	58.6	6.5	52.1	5.3	54.1	1.6	51.8	5.8	58.2	5.9	56.1	2.0	56.6	2.8
29	LME [55]	**63.6**	61.4	3.5	59.2	6.8	59.8	6.0	60.5	5.1	58.1	8.7	57.9	9.0	68.8	8.2	60.7	4.6	59.6	6.4	59.7	6.1	68.7	8.0	65.0	2.2	66.7	4.7
30	TME1 [55]	**62.9**	65.3	3.8	63.2	0.5	64.1	2.0	63.6	1.1	59.4	5.6	59.1	6.0	73.6	17.1	64.3	2.2	62.2	1.1	62.8	0.2	72.9	15.8	68.7	9.2	71.4	11.9
31	TME2 [55]	**61.7**	64.1	3.8	62.1	0.6	62.8	1.9	62.6	1.4	59.0	4.4	58.8	4.7	71.3	15.6	62.5	1.3	61.4	0.4	61.1	1.0	70.9	14.8	66.8	8.3	69.4	11.1
32	YGME [55]	**57.8**	59.2	2.4	56.7	2.0	56.1	2.9	58.4	1.0	57.1	1.2	57.0	1.3	66.9	15.8	59.3	2.7	57.6	0.3	58.0	0.3	66.1	14.3	63.6	10.0	64.2	9.9
33	SME [42]	**54.0**	46.7	13.5	45.9	15.0	44.4	17.8	49.2	9.7	52.8	2.2	53.1	1.7	53.2	1.5	46.4	14.1	51.8	4.1	47.8	11.5	53.1	1.6	50.3	6.9	53.6	0.8
34	Camelina [56]	**52.8**	44.0	16.6	42.7	19.2	44.0	16.6	45.8	15.2	51.4	2.6	51.8	1.9	50.0	5.3	42.6	19.3	50.2	4.9	43.7	17.3	50.8	3.9	46.3	12.3	49.8	6.1
35	Coriander [57]	**53.3**	56.1	5.2	52.9	0.8	51.8	2.8	55.5	4.0	56.0	5.0	56.0	5.0	62.4	17.1	55.5	4.1	55.9	4.9	54.6	2.5	60.8	14.1	59.6	11.8	59.7	10.7
36	Macadamia [58]	**57.5**	58.4	1.6	55.3	3.8	57.0	0.9	56.9	1.0	57.4	0.2	57.2	0.4	64.0	11.3	56.5	1.7	57.7	0.4	55.7	3.2	62.4	8.5	60.5	5.3	59.6	3.6
37	SME [50]	**51.8**	46.9	9.5	46.1	11.1	44.7	13.7	49.3	5.1	52.9	2.2	53.2	2.7	52.7	1.8	45.6	12.0	51.9	0.2	47.0	9.3	52.4	1.1	49.4	4.6	53.1	2.4
38	RSME [46]	**52.5**	50.2	4.4	49.2	6.4	48.4	7.8	51.9	1.1	54.0	2.8	54.1	3.1	55.8	6.3	49.1	6.5	53.0	1.0	50.3	4.3	57.0	8.5	53.1	1.1	59.1	11.2
39	PME [26]	**61.0**	63.0	3.3	60.1	1.4	60.4	1.0	61.9	1.4	58.6	4.0	58.4	4.3	70.0	14.7	61.5	0.9	60.4	1.0	60.8	0.3	69.4	13.7	65.9	8.0	68.8	11.3
40	OME [26]	**57.0**	58.0	1.7	54.6	4.2	54.0	5.3	57.1	0.2	56.6	0.6	56.6	0.7	65.3	14.5	58.0	1.7	56.8	0.4	57.0	0.1	63.4	11.1	62.2	9.1	62.5	8.7
41	PEME [26]	**53.0**	53.3	0.7	50.0	5.6	49.9	5.8	53.5	0.9	55.3	4.4	55.4	4.5	62.0	17.0	55.4	4.5	55.0	3.8	54.3	2.4	59.3	11.9	59.5	12.3	53.5	1.0
42	RME [26]	**55.0**	52.6	4.3	53.2	3.3	51.1	7.2	52.0	5.8	54.2	1.4	54.4	1.2	61.9	12.5	55.9	1.6	53.8	2.2	52.9	3.7	60.0	9.2	60.0	9.0	55.5	0.9
43	SME [26]	**49.0**	46.9	4.3	46.0	6.1	44.6	9.1	49.3	0.6	52.9	7.9	53.1	8.4	53.4	9.1	46.7	4.7	51.9	5.9	48.0	2.1	53.3	8.7	50.6	3.3	53.5	8.4
44	SFME [26]	**50.0**	46.1	7.8	45.1	9.9	43.3	13.4	48.7	2.7	52.7	5.4	53.0	6.0	53.4	6.7	46.7	6.6	51.8	3.6	47.8	4.5	52.4	4.8	50.6	1.3	50.3	0.5
45	GME [26]	**48.0**	45.0	6.3	44.3	7.7	42.4	11.7	48.0	0.1	52.3	9.0	52.6	9.6	51.9	8.1	44.9	6.5	51.4	7.0	46.4	3.4	50.8	5.9	48.7	1.5	49.5	3.1
46	HO SFME [26]	**53.0**	53.0	0.1	50.4	4.9	49.2	7.3	53.4	0.8	55.0	3.8	55.1	3.9	60.1	13.4	53.6	1.1	54.7	3.2	53.1	0.2	58.8	10.9	57.6	8.8	56.9	6.8
47	AME [26]	**57.0**	57.7	1.2	54.4	4.6	53.6	6.0	56.9	0.3	56.5	0.8	56.5	0.9	64.8	13.7	57.6	1.1	56.6	0.7	56.6	0.7	63.2	10.9	61.8	8.5	62.3	8.6
48	CRME [26]	**53.0**	53.4	0.7	50.6	4.4	49.4	6.8	53.7	1.2	55.1	3.9	55.1	4.0	60.0	13.3	53.4	0.8	54.8	3.4	53.2	0.3	58.7	10.8	57.5	8.5	57.6	7.9
49	SME [49]	**49.0**	44.5	9.2	44.0	10.2	42.6	13.1	47.4	3.4	51.9	6.0	52.3	6.7	51.1	4.3	43.9	10.3	50.8	3.6	45.8	6.6	51.2	4.4	47.8	2.5	53.3	8.1
50	SFME [49]	**47.0**	45.0	4.3	44.8	4.8	42.8	8.9	48.2	2.5	52.3	11.4	52.6	12.0	52.1	10.8	45.1	3.9	51.1	8.8	46.7	0.6	51.5	9.5	49.0	4.3	50.2	6.3
	Average Error	**-**		**4.78**		**6.20**		**7.89**		**2.97**		**4.05**		**4.29**		**9.27**		**5.49**		**2.97**		**4.34**		**8.04**		**5.19**		**6.35**

(AME: almond; CAME: canola; COME: cottonseed; CRME: corn; EMME: egushi mellon; GME: grape; JME: jatropha; LME: lard; OME: olive; PEME: peanut; PGME: pongamia; PKME: palm kernel; PME: palm; RME: rapeseed; RSME: rubber seed; SFME: sunflower; SME: soybean; TME: tallow; WFME: waste frying; WPME: waste palm; YGME: Yellow grease).

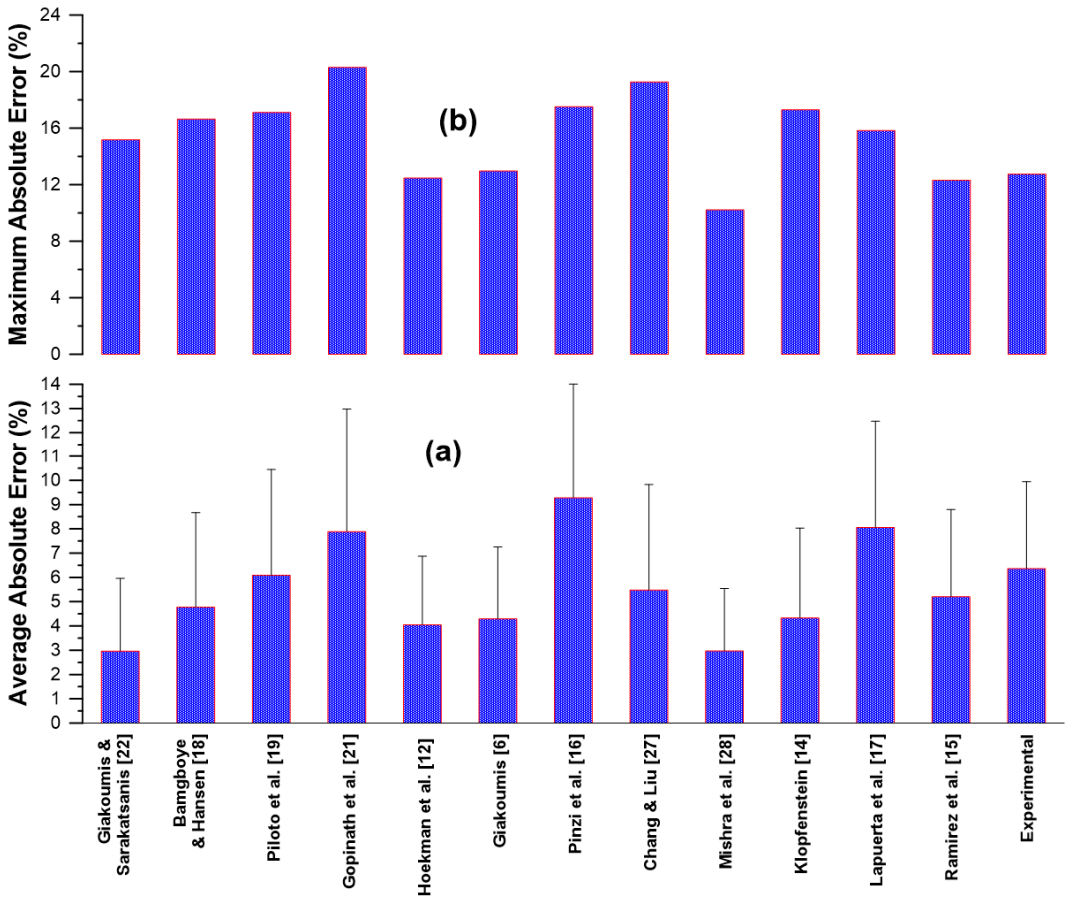

Figure 5. Comparative illustration of average (**a**) and maximum (**b**) absolute errors from all models' predictions based on the results of Table 5.

A more detailed picture of each model's errors is provided in Figure 6; it is further analyzed in Table 6, which provides the instances of higher than 5% and 10% absolute errors from all models. As evidenced in Table 6, the model by Mishra et al. [28] is best here (only one instance with higher than 10% error and nine occurrences with higher than 5%); very close second was the model by Giakoumis and Sarakatsanis [22] (two and 10 occurrences respectively). On the other hand, all mixing rule models manifest at least 20 out of 50 predictions with higher than 5% error. Lastly, it is important to note that there was no instance of higher than 20% error in any of the models.

Table 6. Errors from all models (based on the data from Table 5).

Model	Higher than 5% (Number of Occurrences Out of 50)	Higher than 10% (Number of Occurrences Out of 50)	Model Type
Mishra et al. [28]	9	1	Average DU/CL/MW
Giakoumis and Sarakatsanis [22]	10	2	Compositional
Hoekman et al. [12]	14	3	Average DU
Giakoumis [6]	14	3	Average DU
Bamgboye and Hansen [18]	16	5	Compositional
Klopfenstein [14]	20	3	Mixing Rule
Ramirez-Verduzco et al. [15]	23	5	Mixing Rule
Chang and Liu [27]	24	9	Average DU/CL/MW
Piloto-Rodriguez et al. [19]	27	11	Compositional
Experimental	32	10	Mixing Rule
Lapuerta et al. [17]	34	18	Mixing Rule
Gopinath et al. [21]	36	16	Compositional
Pinzi et al. [16]	39	22	Average DU/CL/MW

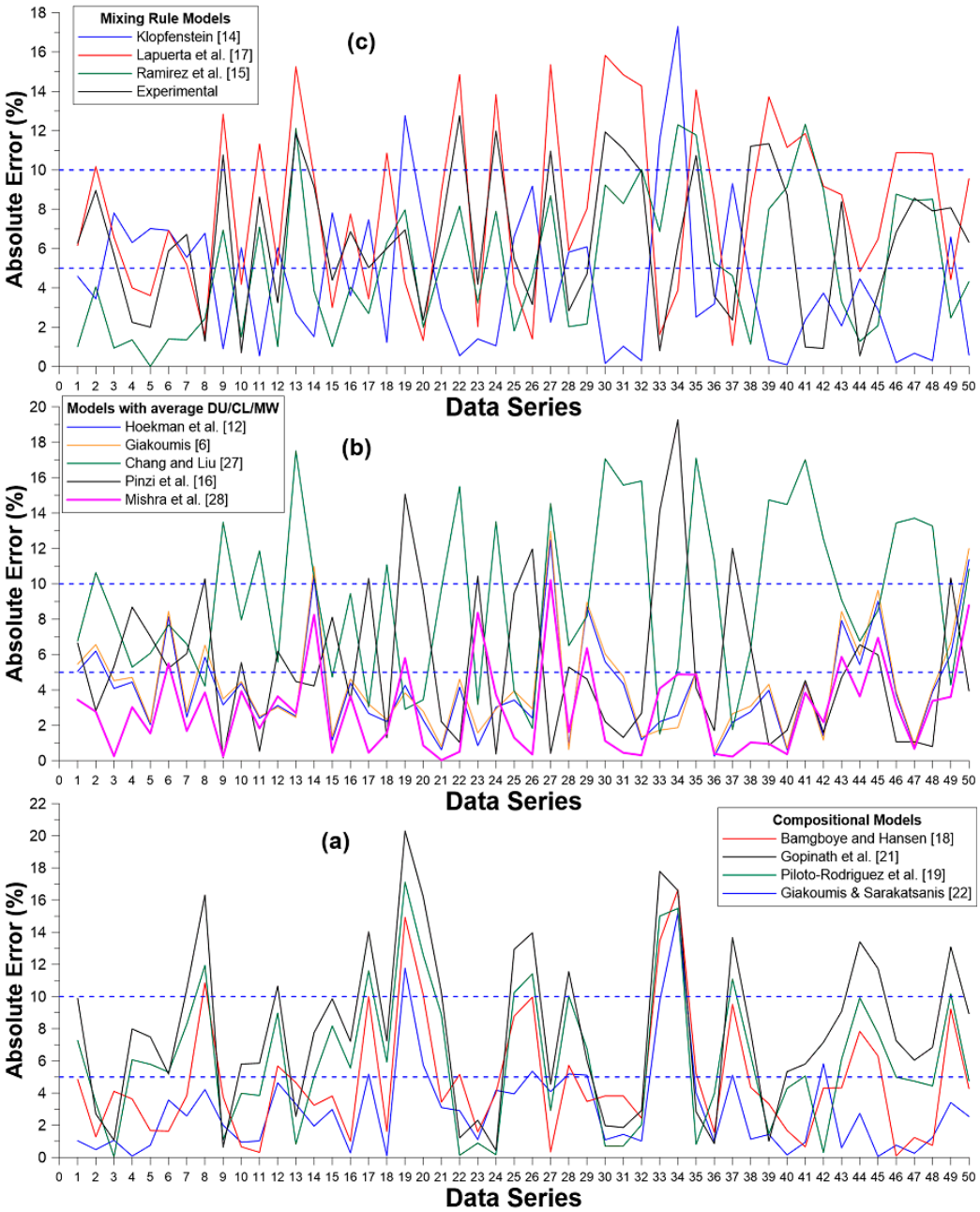

Figure 6. Absolute errors for all models ((**a**): compositional; (**b**) models with average DU/CL; and (**c**) mixing rule models) and for all 50 experimental data lines.

A slightly different picture of the models' predictive capabilities against the data set of experimental values is demonstrated in Figures 7–9. These figures illustrate the degree of accuracy achieved by each model against the experimental data to which it was tested. More specifically, Figure 7 focuses on the four compositional models, Figure 8 on the five models taking into account average values for the degree of unsaturation and chain length/molecular weight, and Figure 9 on the four models that are based on the neat FAMEs' CN values for their predictions (including the 'experimental' one). A closer look at the graphs in Figures 7–9 reveals that two discrete groups of CN values can be identified. One is located between CN 40 and 56, and another above CN 60. The former corresponds to vegetable oil-derived biodiesels, and the latter mostly to animal-derived ones (palm is included here too). The following findings can be reached from the results depicted in Figures 7–9 and Table 7.

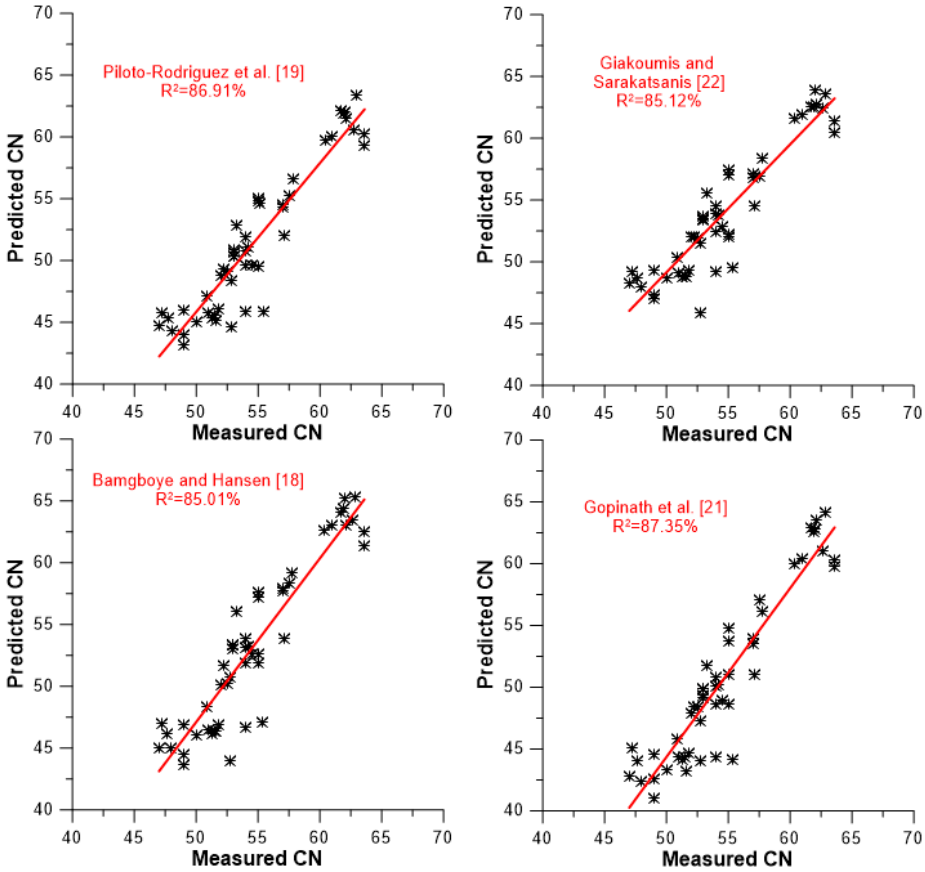

Figure 7. Comparison of predicted (equations from Section 3.1) vs. experimental CN values for the four compositional models (experimental data from the third column of Table 5).

All models present fairly satisfactory accuracy, ranging from 77.5% (Chang and Liu [27] and Ramirez-Verduzco et al. [15]) up to 87.35% for the model by Gopinath et al. [21]. No correlation seems to be markedly superior compared to the others (this was a usual claim by many research groups when introducing their own predictive model).

The four compositional models seem to possess cumulatively the best predicting qualities (Figure 7 and Table 7). Moreover, their R^2 values are fairly consistent, ranging from 85.01 to 87.35%, i.e., within only 2.34 percentage points. Obviously, omitting (or adding) a few experimental data lines might change the differences between them, even in terms of their order (see discussion at the end of the section). However, this would not change the fact that they seem to be more accurate in their predictions than the models of the other two categories.

Interestingly, the two models that only take the degree of unsaturation into consideration (lower two sub-graphs of Figure 8) manifest very satisfactory predictions, with identical R^2 values of 82.74% (recall from the analysis of Section 3.2 that the respective equations were relatively similar, and that the degree of unsaturation has a strong and clearly linear effect on CN).

Despite their sounder theoretical basis and the much more complex equations involved, the models that incorporate the effects of both DU and molecular weight or chain length were not proven to be notably superior in their predictions. Another interesting feature of the models of this category (Section 3.2) is that they present the greatest variability in their predictions, ranging from 77.53 to 82.74%, as indicated in Figure 8.

With respect to the models that predict the individual FAME's CN and then apply a mixing rule for the biodiesel (Section 3.3), the model by Lapuerta et al. [17] is best (R^2 = 82.69%), followed by the 'experimental' one (81.90%). Recall from the discussion of Figure 3 that the model by Lapuerta et al. [17] was identified as the one whose predictions were closest to the experimental values, hence the fairly

similar coefficients of determination. Nonetheless, as Table 6 and Figure 6 indicate, the model by Lapuerta et al. [17] suffers from several high errors in its predictions (much more than the other models; the same holds true for the model by Pinzi et al. [16] and Gopinath et al. [21]). The Klopfenstein model [14] (R^2 = 81.22%) is close to the best two of this category, whereas that of Ramirez-Verduzco [15] demonstrates a lower predictive level (R^2 = 77.53%). The R^2 values of this category are within 5.16 percentage points.

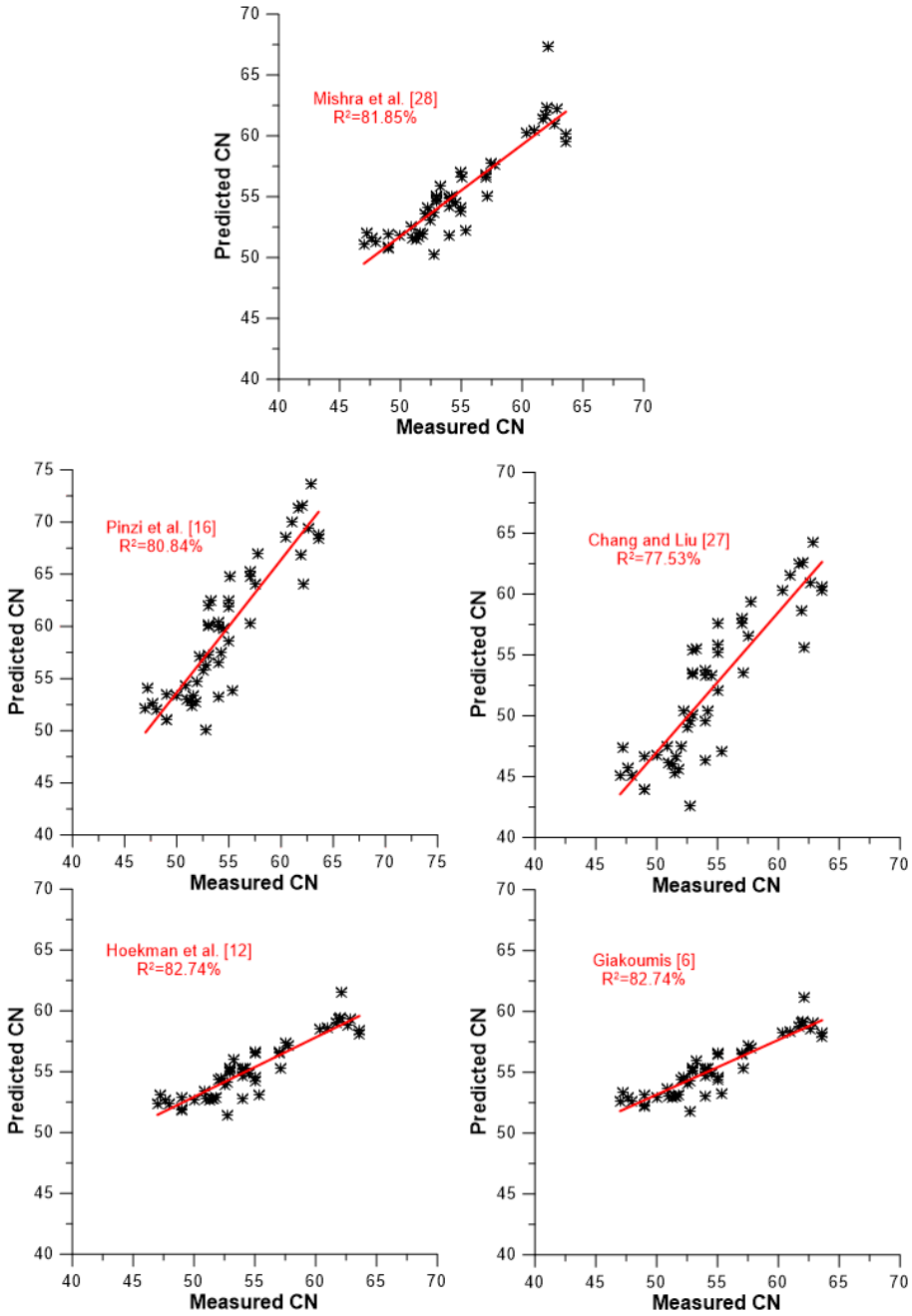

Figure 8. Comparison of predicted (equations from Section 3.2) vs. experimental CN values for the five models based on the average biodiesel degree of unsaturation and chain length (experimental data from the third column of Table 5).

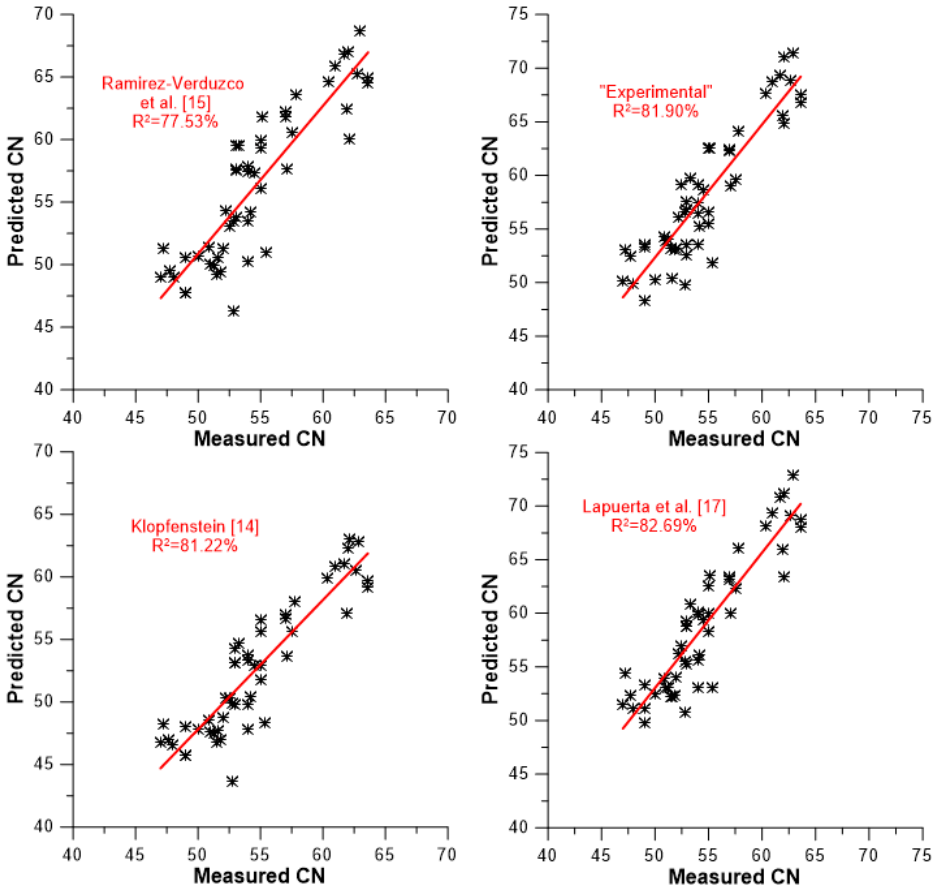

Figure 9. Comparison of predicted (equations from Section 3.3) vs. experimental CN values for the four models based on the neat FAMEs' cetane numbers and applying the mixing rule for the biodiesel CN (experimental data from the third column of Table 5).

Table 7. Coefficients of determination of all models (based on the 50 data lines from Table 5 and Figures 7–9).

Model	Equation	R^2 (%)	Model Type
Gopinath et al. [21]	4	87.35	Compositional
Piloto-Rodriguez et al. [19]	5	86.91	Compositional
Giakoumis and Sarakatsanis [22]	6	85.12	Compositional
Bamgboye and Hansen [18]	3	85.01	Compositional
Giakoumis [6]	8	82.74	Avg. DU
Hoekman et al. [12]	7	82.74	Avg. DU
Lapuerta et al. [17]	16	82.69	Mixing Rule
Experimental	—	81.90	Mixing Rule
Mishra et al. [28]	11	81.85	Avg. DU/CL/MW
Klopfenstein [14]	13	81.22	Mixing Rule
Pinzi et al. [16]	10	80.84	Avg. DU/CL/MW
Chang and Liu [27]	9	77.53	Avg. DU/CL/MW
Ramirez-Verduzco et al. [15]	17	77.53	Mixing Rule

The next part of the investigation in the predictive capability of the CN models was to identify possible systematic faulty behavior in the models' predictions under specific conditions. The following were identified (although no conclusive results could be reached):

- In general, compositional models are expected to be less capable in predicting the cetane number for those (rather unusual cases of) methyl esters that are rich in fatty acids not included in their

MLR equations. Excluding the compositional models, the other equations incorporate, by default, all fatty acids' effects in their equations; hence, they are better equipped in this regard.

- Interestingly, all four compositional models proved inefficient in two cases, despite the fact that the only detected FAs were included in their equations. The first case concerned line 19 of Table 4. The ME being tested here is cottonseed, and its FA composition consists only of FAs included in the four compositional models equations. Moreover, the reported CN is within the acceptable limits, as the survey in [6] confirms. Yet, all four compositional models manifest high errors in their predictions, in the order of 11.8-20.3%, which is not the case for most of the other examined models (that is why the specific data line was not excluded from the analysis). The exact same observation can be made with reference to line 33 in Table 4 (soybean ME). For both incidences, no clear explanation can be provided, except if the experimental values are incorrect. All four models' CN predictions are in both cases between 44 and 50 and the measured values between 54 and 55.4.
- Three of the compositional models (that of Bamgboye and Hansen [18], Gopinath et al. [21], and Piloto-Rodriguez et al. [19]) exhibit most of their largest errors for high DUs (1.50 and above). This might indicate overestimation of the unsaturation coefficients in the respective equations. Interestingly, this trend is not observed for the fourth model of this category, namely, the one by Giakoumis and Sarakatsanis [22]. The models of the other two categories do not suffer from this problem (faulty predictions for MEs with high DUs), with the exception of the one by Chang and Liu [27].
- For low DUs, on the other hand, it is the model by Pinzi et al. [16], and to a lesser extent, the Lapuerta et al. [17] and the experimental ones, that produce their largest errors. This might indicate errors in the short-chain saturated FAMEs CN values.
- The model by Chang and Liu [27] seems to also suffer from wrong predictions when the chain length is very low; this might suggest underestimation of the CL constant in Equation (9).
- The models by Lapuerta et al. [17] and Pinzi et al. [16] were found to lead to several faulty predictions (18 and 22 respectively with absolute error higher than 10%) for biodiesels possessing high CN values (above 60).

Lastly, a sensitivity analysis was conducted regarding the effect that the length of the data series has on the models' prediction. Table 8 provides the R^2 values for three cases: a) when only the first 25 data lines from Table 4 are taken into consideration; b) when the data lines from 16 to 40 are considered; and c) when the data lines from 26 to 50 are accounted for. As evidenced, reducing the data sample affects the reached coefficients of determination. Nonetheless, the general trends are largely unaffected, e.g., the four compositional models are always the best, the models by Chang and Liu [27] and Ramirez-Verduzco et al. [15], the ones with the lowest R^2 etc. It is very interesting that, in general, all models seem to behave in a uniform way when fewer data are considered, i.e., they all improve or deteriorate in their predictions. In light of these arguments, it is believed that the complete set of 50 data lines from Tables 4 and 5, as summarized in Table 7 and Figures 7–9, is indeed representative of the models' predictive capability.

Table 8. Coefficients of determination for all models when fewer data lines from Table 4 are taken into account.

Model	Data Lines 1 to 25 in Table 4	Data Lines 16 to 40 in Table 4	Data Lines 26 to 50 in Table 4
		R^2 (%)	
Gopinath et al. [21]	87.19	83.62	87.72
Piloto-Rodriguez et al. [19]	87.73	84.05	86.26
Giakoumis and Sarakatsanis [22]	87.92	81.60	82.30
Bamgboye and Hansen [18]	86.51	80.61	83.43
Giakoumis [6]	84.50	78.86	81.15
Hoekman et al. [12]	84.50	78.86	81.15
Lapuerta et al. [17]	82.33	77.91	84.83
Experimental	81.42	77.03	82.57
Mishra et al. [28]	81.90	78.83	83.36
Klopfenstein [14]	84.32	80.52	79.13
Pinzi et al. [16]	83.33	77.77	81.25
Chang and Liu [27]	80.68	74.03	77.82
Ramirez-Verduzco et al. [15]	80.83	74.48	77.74

5. Summary and Conclusions

A detailed review of the existing literature on biodiesel CN predictive correlations based on the fatty acid composition was performed gathering, in total, sixteen models. In the end, 13 models were compared against a broad/representative data set of 50 experimental FA composition and CN values carefully selected from a variety of literature sources. It is believed that the amount of experimental data, as well as their variability, is adequate to establish trends and reach some reliable conclusions.

All models proved their satisfactory predictive capability, with better results reached by the four compositional models (R^2 ranging between 85 and 87.35%). The equations based only on the average degree of unsaturation were proven to be very effective (R^2 = 82.74%). Surprisingly, more detailed models, incorporating both the effect of the (average) degree of unsaturation and that of the average chain length, with more complex, e.g., quadratic, form, did not lead to better predictions (R^2 ranging from 77.53 to 81.85%). Models based on neat methyl esters CN prediction for their calculations inherently suffer from the sometimes wide variability of the respective experimental data they rely upon; the two best achieved R^2 values are in the order of 81.9% and 82.7% i.e., 4–5 percentage points lower than the values reached by the compositional models. Regarding the errors from the predictions, no modeling category exceled. One compositional and one model based on the average DU/CL were identified as the best here.

Overall, the models by Gopinath et al. [21] and Piloto-Rodriguez et al. [19] exceled in the R^2 values but suffered from many high-error predictions. That of Mishra et al. [28] was the one with the lowest overall errors but a rather moderate R^2 value. The model by Giakoumis and Sarakatsanis [22] was proven to be the most cohesive of all, exhibiting the third best R^2 value combined with the second lowest errors. That of Bamgboye and Hansen [18] was quite effective too, being fourth in terms of R^2 and fifth in terms of the lowest errors.

Based on the large set of experimental data, a few trends were also identified regarding possible inefficiencies of the various models at extreme DU, CL and CN values, although no conclusive results can be established.

It is believed that the results from this study can prove useful to researchers and institutions with respect to simulation and planning, particularly when taking into account the quite challenging experimental determination of cetane number.

Funding: This research received no external funding.

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

Nomenclature

ASTM	American Society for Testing and Materials
CAS	Chemical Abstracts Service
CFR	Cooperative fuel research
CI	Cetane index
CL	Chain length
CN	Cetane number
CVCC	Constant-volume combustion chamber
Cxx:y	Fatty acid with xx number of carbon atoms; y=1 for mono-unsaturated, ≥ 2 for poly-unsaturated
db	Number of double bonds
DU	Degree of unsaturation
FA	Fatty acid
FAME	Fatty acid methyl ester
IQT	Ignition quality tester
ISO	International Organization for Standardization
ME	Methyl ester
MLR	Multiple linear regression
MW	Molecular weight
n	Carbon number
w	weight percentage
Subscripts	
i	each fatty acid or fatty acid methyl ester

References

1. Hansen, A.C.; Kyritsis, D.C.; Lee, C.F. Characteristics of biofuels and renewable fuel standards. In *Biomass to Biofuels—Strategies for Global Industries*; Vertes, A.A., Qureshi, N., Blaschek, H.P., Yukawa, H., Eds.; Blackwell Publishing: Oxford, UK, 2009.
2. Agarwal, A.K. Biofuels (alcohols and biodiesel) applications as fuels in internal combustion engines. *Progr. Energy Combust. Sci.* **2007**, *32*, 233–271. [CrossRef]
3. Graboski, M.S.; McCormick, R.L. Combustion of fat and vegetable oil derived fuels in diesel engines. *Progr. Energy Combust. Sci.* **1998**, *24*, 125–164. [CrossRef]
4. Tziourtzioumis, D.; Stamatelos, A. Effects of a 70% biodiesel blend on the fuel injection system operation during steady-state and transient performance of a common rail diesel engine. *Energy Convers. Manag.* **2012**, *60*, 56–67. [CrossRef]
5. Giakoumis, E.G.; Rakopoulos, C.D.; Dimaratos, A.M.; Rakopoulos, D.C. Exhaust emissions of diesel engines operating under transient conditions with biodiesel fuel blends. *Progr. Energy Combust. Sci.* **2012**, *38*, 691–715. [CrossRef]
6. Giakoumis, e.g., A statistical investigation of biodiesel physical and chemical properties, and their correlation with the degree of unsaturation. *Renew. Energy* **2013**, *50*, 858–878. [CrossRef]
7. Knothe, G. "Designer" biodiesel: Optimizing fatty ester composition to improve fuel properties. *Energy Fuels* **2008**, *22*, 1358–1364. [CrossRef]
8. Heywood, J.B. *Internal Combustion Engine Fundamentals*; McGraw-Hill: New York, NY, USA, 1988.
9. Knothe, G. Dependence of biodiesel fuel properties on the structure of fatty acid alkyl esters. *Fuel Process. Technol.* **2005**, *86*, 1059–1070. [CrossRef]
10. Knothe, G. A comprehensive evaluation of the cetane numbers of fatty acid methyl esters. *Fuel* **2014**, *119*, 6–13. [CrossRef]
11. Freedman, B.; Bagby, M.O. Predicting cetane numbers of alcohols and methyl esters from their physical properties. *J. Am. Oil Chem. Soc.* **1990**, *67*, 565–571. [CrossRef]

12. Hoekman, S.K.; Broch, A.; Robbins, C.; Ceniceros, E.; Natarajan, M. Review of biodiesel composition, properties and specifications. *Renew. Sustain. Energy Rev.* **2012**, *16*, 143–169. [CrossRef]
13. Mc Cormick, R.L.; Grabowski, M.S.; Alleman, T.L.; Herring, A.M.; Tyson, K.S. Impact of biodiesel source material and chemical structure on emissions of criteria pollutants from a heavy-duty engine. *Environ. Sci. Technol.* **2001**, *35*, 1742–1747. [CrossRef]
14. Klopfenstein, W.E. Estimation of cetane index for esters of fatty acids. *J. Am. Oil Chem. Soc.* **1982**, *59*, 531–533. [CrossRef]
15. Ramirez-Verduzco, L.F.; Rodriguez-Rodriguez, J.E.; Del Rayo Jaramillo-Jacob, A. Predicting cetane number, kinematic viscosity, density and higher heating value of biodiesel from its fatty acid methyl ester composition. *Fuel* **2012**, *91*, 102–111. [CrossRef]
16. Pinzi, S.; Leiva, D.; Arzamendi, G.; Gandia, L.M.; Dorado, M.P. Multiple response optimization of vegetable oils fatty acid composition to improve biodiesel physical properties. *Biores. Technol.* **2011**, *102*, 7280–7288. [CrossRef]
17. Lapuerta, M.; Rodriguez-Fernandez, J.; Font de Mora, E. Correlation for the estimation of the cetane number of biodiesel fuels and implications on the iodine number. *Energy Policy* **2009**, *37*, 4337–4344. [CrossRef]
18. Bamgboye, A.I.; Hansen, A.C. Prediction of cetane number of biodiesel fuel from the fatty acid methyl ester (FAME) composition. *Int. Agrophys.* **2008**, *22*, 21–22.
19. Piloto-Rodriguez, R.; Sanchez-Borroto, Y.; Lapuerta, M.; Goyos-Perez, L.; Verhelst, S. Prediction of the cetane number of biodiesel using artificial neural networks and multiple linear regression. *Energy Convers. Manag.* **2013**, *65*, 255–261. [CrossRef]
20. Ramadhas, A.S.; Jayaraj, S.; Muraleedharan, C.; Padmakumari, K. Artificial neural networks used for the prediction of the cetane number of biodiesel. *Renew. Energy* **2006**, *31*, 2524–2533. [CrossRef]
21. Gopinath, A.; Puhan, S.; Nagarajan, G. Relating the cetane number of biodiesel fuels to their fatty acid composition: A critical study. *Proc. Inst. Mech. Eng. J. Automobile Eng.* **2009**, *223*, 565–583. [CrossRef]
22. Giakoumis, E.G.; Sarakatsanis, C.K. Estimation of biodiesel cetane number, density, kinematic viscosity and heating values from its fatty acid weight composition. *Fuel* **2018**, *222*, 574–585. [CrossRef]
23. Ladommatos, N.; Goacher, J. Equations for predicting the cetane number of diesel fuels from their physical properties. *Fuel* **1995**, *74*, 1083–1093. [CrossRef]
24. Demirbas, A. Relationships derived from physical properties of vegetable oil and biodiesel fuels. *Fuel* **2008**, *87*, 1743–1748. [CrossRef]
25. Yanowitz, J.; Ratcliff, M.A.; McCormick, R.L.; Taylor, J.D.; Murphy, M.J. *Compendium of Experimental Cetane Number Data*; NREL/TP-5400-67585; National Renewable Energy Laboratory: Denver, CO, USA, 2017.
26. Ramos, M.J.; Fernandez, C.M.; Casas, A.; Rodriguez, L.; Perez, A. Influence of fatty acid composition of raw materials on biodiesel properties. *Bioresour. Technol.* **2009**, *100*, 261–268. [CrossRef] [PubMed]
27. Chang, A.F.; Liu, Y.A. Integrated process modeling and product design of biodiesel manufacturing. *Ind. Eng. Chem. Res.* **2010**, *49*, 1197–1213. [CrossRef]
28. Mishra, S.; Anand, K.; Mehta, P.S. Predicting the cetane number of biodiesel fuels from their fatty acid methyl ester composition. *Energy Fuels* **2016**, *30*, 10425–10434. [CrossRef]
29. Klopfenstein, W.E. Effect of molecular weights of fatty acid esters on cetane numbers as diesel fuel. *J. Am. Oil Chem. Soc.* **1985**, *62*, 1029–1031. [CrossRef]
30. Tong, D.; Hu, C.; Jiang, K.; Li, Y. Cetane number prediction of biodiesel from the composition of the fatty acid methyl esters. *J. Am. Oil Chem. Soc.* **2011**, *88*, 415–423. [CrossRef]
31. Clements, L.D. Blending rules for formulating biodiesel fuel. In Proceedings of the 3rd Liquid Fuel Conference, Nashville, TN, USA, 15–16 September 1996; pp. 44–53.
32. Shay, E.G. Diesel fuel from vegetable oils: Status and opportunities. *Biomass Bioenergy* **1993**, *4*, 227–242. [CrossRef]
33. Baumgardner, M.E.; Sarathy, S.M.; Marchese, A.J. Autoignition characterization of primary reference fuels and n-heptane/n-butanol mixtures in a constant volume combustion device and homogeneous charge compression ignition engine. *Energy Fuels* **2013**, *27*, 7778–7789. [CrossRef]

34. Natarajan, M.; Frame, E.A.; Naegeli, D.W.; Asmus, T.; Clark, W.; Garbak, J.; Manuel, A.; González, D.; Liney, E.; Piel, W.; et al. *Oxygenates for Advanced Petroleum-Based Diesel Fuels: Part 1. Screening and Selection Methodology for the Oxygenates*; SAE Paper No. 2001-01-3631; SAE International: Warrendale, PA, USA, 2001.
35. Serdari, A.; Lois, E.; Stournas, S. Impact of esters of mono- and dicarboxylic acids on diesel fuel quality. *Ind. Eng. Chem. Res.* **1999**, *38*, 3543–3548. [CrossRef]
36. Dahmen, M.; Marquardt, W. A novel group contribution method for the prediction of the derived cetane number of oxygenated hydrocarbons. *Energy Fuels* **2015**, *29*, 5781–5801. [CrossRef]
37. Freedman, B.; Bagby, M.; Callahan, T.; Ryan, T. *Cetane Numbers of Fatty Esters, Fatty Alcohols and Triglycerides Determined in a Constant Volume Combustion Bomb*; SAE Paper No. 900343; SAE International: Warrendale, PA, USA, 1990.
38. Knothe, G.; Matheaus, A.C.; Ryan, T.W., III. Cetane numbers of branched and straight-chain fatty esters determined in an ignition quality tester. *Fuel* **2003**, *82*, 971–975. [CrossRef]
39. Knothe, G.; Bagby, M.O.; Ryan, T.W., III. *Cetane Numbers of Fatty Compounds: Influence of Compound Structure and of Various Potential Cetane Improvers*; SAE Paper No. 971681; SAE International: Warrendale, PA, USA, 1997.
40. Moser, B.R.; Knothe, G.; Cermack, S.C. Biodiesel from Meadowfoam (*Limnanthes alba* L.) seed oil: Oxidative stability and unusual fatty acid composition. *Energy Environ. Sci.* **2010**, *3*, 318–327. [CrossRef]
41. Knothe, G. Fuel properties of methyl esters of borage and black currant oils containing methyl γ-linolenate. *Eur. J. Lipid Sci. Technol.* **2013**, *115*, 901–908. [CrossRef]
42. Knothe, G.; Sharp, C.A.; Ryan, T.W. Exhaust emissions of biodiesel, petrodiesel, neat methyl esters, and alkanes in a new technology engine. *Energy Fuels* **2006**, *20*, 403–408. [CrossRef]
43. Sajjadi, B.; Raman, A.A.A.; Arandiyan, H. A comprehensive review on properties of edible and non-edible vegetable oil-based biodiesel: Composition, specifications and prediction models. *Renew. Sustain. Energy Rev.* **2016**, *63*, 62–92. [CrossRef]
44. Bachler, C.; Schober, S.; Mittelbach, M. Simulated distillation for biodiesel analysis. *Energy Fuels* **2010**, *24*, 2086–2090. [CrossRef]
45. Lin, B.-F.; Huang, J.-H.; Huang, D.-I. Experimental study of the effects of vegetable oil methyl ester on DI diesel engine performance characteristics and pollutant emissions. *Fuel* **2009**, *88*, 1779–1785. [CrossRef]
46. Bello, E.I.; Out, F.; Osasona, A. Cetane number of three vegetable oils, their biodiesels and blends with diesel fuel. *J. Petroleum. Technol. Altern. Fuels* **2012**, *3*, 52–57.
47. Ng, J.-H.; Ng, H.-K.; Gan, S. Engine-out characterisation using speed-load mapping and reduced test cycle for a light-duty diesel engine fuelled with biodiesel blends. *Fuel* **2011**, *90*, 2700–2709. [CrossRef]
48. Canakci, M. Combustion characteristics of a turbocharged DI compression ignition engine fueled with petroleum diesel fuels and biodiesel. *Bioresource Technol.* **2007**, *98*, 1167–1175. [CrossRef] [PubMed]
49. Bazooyar, B.; Ghorbani, A.; Shariati, A. Combustion performance and emissions of petrodiesel and biodiesels based on various vegetable oils in a semi industrial boiler. *Fuel* **2011**, *90*, 3078–3092. [CrossRef]
50. Szybist, J.P.; Boehman, A.L.; Taylor, J.D.; McCormick, R.L. Evaluation of formulation strategies to eliminate the biodiesel NOx effect. *Fuel Process Technol.* **2005**, *86*, 1109–1126. [CrossRef]
51. Sarin, R.; Sharma, M.; Sinharay, S.; Malhotra, R.K. Jatropha-Palm biodiesel blends: An optimum mix for Asia. *Fuel* **2007**, *86*, 1365–1371. [CrossRef]
52. Schoenborn, A.; Ladommatos, N.; Williams, J.; Allan, R.; Rogerson, J. The influence of molecular structure of fatty acid monoalkyl esters on diesel combustion. *Combust. Flame* **2009**, *156*, 1396–1412. [CrossRef]
53. Yehliu, K.; Boehman, A.L.; Armas, O. Emissions from different alternative diesel fuels operating with single and split fuel injection. *Fuel* **2010**, *89*, 423–437. [CrossRef]
54. Chen, Y.-H.; Chen, J.-H.; Luo, Y.-M.; Shang, N.-C.; Chang, C.-H.; Chang, C.-Y.; Chang, C.-H.; Chang, C.-Y.; Chiang, P.-C.; Shie, J.-L. modification of jatropha oil biodiesel by blending with other biodiesels or adding antioxidants. *Energy* **2011**, *36*, 4415–4421. [CrossRef]
55. Kinast, J.A. *Production of Biodiesels from Multiple Feedstocks and Properties of Biodiesels and Biodiesel/Diesel Blends*; Final Report, NREL/SR-510-31460; National Renewable Energy Laboratory: Denver, CO, USA, 2003.

56. Moser, B.R.; Vaughn, S.F. Evaluation of alkyl esters from camelina sativa oil as biodiesel and as blend components in ultra low-sulfur diesel fuel. *Bioresour. Technol.* **2010**, *101*, 646–653. [CrossRef]
57. Moser, B.R.; Vaughn, S.F. Coriander seed oil methyl esters as biodiesel fuel: Unique fatty acid composition and excellent oxidative stability. *Biomass Bioenergy* **2010**, *34*, 550–558. [CrossRef]
58. Azad, A.K.; Rasul, M.; Khan, M.M.; Sharma, S. Macadamia biodiesel as a sustainable and alternative transport fuel in Australia. *Energy Procedia* **2017**, *110*, 543–548. [CrossRef]

© 2019 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 (http://creativecommons.org/licenses/by/4.0/).