CHEMICAL ENGINEERING TRANSACTIONS
VOL. 94, 2022
A publication of AIDIC
The Italian Association of Chemical Engineering
Online at www.cetjournal.it
Guest Editors: Petar S. Varbanov, Yee Van Fan, Jiří J. Klemeš, Sandro Nižetić

ISBN 978-88-95608-93-8; **ISSN** 2283-9216
DOI: 10.3303/CET2294162

Unified Characterisation and Property Estimation Framework for Composition Reconstruction of Biomass Pyrolysis Oil and Petroleum Fractions

Qiong Pan[a], Xiaolei Fan[b], Jie Li[a,*]

[a]Centre for Process Integration, Department of Chemical Engineering, School of Engineering, The University of Manchester, Manchester M13 9PL, UK
[b]Department of Chemical Engineering, School of Engineerwing, The University of Manchester, Manchester M13 9PL, UK
jie.li-2@manchester.ac.uk

Non-fossil biomass gradually becomes a promising raw material in energy consumption structure for achieving carbon neutrality. Most of existing process modelling works mainly focus on fossil streams, challenges in molecular composition determination and digitalization restrain the application of existing works to biomass-derived materials. In this work, a molecular level molecular composition reconstruction framework is proposed, the framework covers representation of molecules and transformation between fraction information and mixture bulk properties. Molecular fingerprint method is introduced for structure description and a data-driven pure-component estimation method is integrated in the framework. Statistical method is implemented for parameter reduction in model optimisation. The accuracy and potential application of the methodology is evaluated by composition reconstruction of a diesel and a bio-oil sample, deviation between most of the typical measured and predicted properties are within 1 %. In addition, detailed molecular information is retained using the molecular fingerprint method, which makes it easier to integrate the proposed framework with existing frameworks.

1. Introduction

Emerging environmental pressure and responding protection policy are prompting refineries transform toward intelligent and digitalized plants. On the one hand, fossil feedstock quality is deteriorating as heteroatom doped heavy oil dominants (Pinho et al., 2017), in the meantime sustainable biomass derived oil as an alternative energy source gradually gets more attention (Wang et al., 2015). On the other hand, car manufacturers' strategy of switching to emissions-free cars results in a decline of fossil fuel demand. Predictably, with vehicle fuel consumption expected to wane, dominate routine of petrochemical industry would be crude-to-chemicals complexes by the 2020s (Tullo, 2019). Existing refining technology roadmap should be upgraded to achieving carbon neutrality by the 2050s. The 'molecular management' concept, which is 'targeting the right molecules to be at the right place, at the right time and at the right price', is highly fitted with objectives of modern refineries (Wu, 2010). Process modelling and optimisation are effective ways of achieving the object. However, one prerequisite is in-depth understanding and digitalization of feedstock molecular composition.

Progresses in analytical techniques enable a better understanding of the complex mixtures, as well as the computer science advancement boosted the modelling of chemical processes. Though chromatography or spectrometry techniques can identify most of molecules in light oil, structural information or quantification result of heavy oil are difficult to obtain. In order to address the limitation of experimental approach, computer-aided molecular composition interpretation of petroleum steams based on available information is a practical way (Klein et al., 2005). Various molecular level framework has been developed to digitalize molecular composition at different period. Generally, developed molecular composition modelling methods consist of representation of molecular structure information using string/symbol and determination the abundance of representatives (Glazov et al., 2021). Highlighted framework including stochastic reconstruction (SR) (Neurock et al., 1990), reconstruct by entropy maximization (REM) (Hudebine and Verstraete, 2004), Structural-Oriented Lumping

Paper Received: 15 April 2022; Revised: 29 May 2022; Accepted: 09 June 2022
Please cite this article as: Pan Q., Fan X., Li J., 2022, Unified Characterisation and Property Estimation Framework for Composition Reconstruction of Biomass Pyrolysis Oil and Petroleum Fractions, Chemical Engineering Transactions, 94, 973-978 DOI:10.3303/CET2294162

(SOL) (Quann and Jaffe, 1992) and molecular-type homologous series (MTHS) (Peng, 1999) method. Technically, a composition modelling framework mainly comprises composition representation and fraction-property transformation, while analytical composition information, pure-component and mixture property estimation, model optimisation are indispensable for a unified framework.

Molecules are the basic unit of feedstock composition, property correlation, reaction kinetics. Therefore, information consistency is the governing principle in developing a unified framework (Bojkovic, 2021). However, accuracy of existing framework is limited due to the uneven development of each segment, which could be explained by the Cannikin Law that a bucket's capacity is determined by its shortest stave. One example is property estimation, most widely used method in above frameworks are simple correlation-based approaches which is not accurate enough. Though the Group Contribution (GC) methods are more accurate, but the 'groups' are usually incompatible with the structure attributes of the most of molecular composition reconstruction frameworks. For example, naphthalene's structure consists of a fused pair of benzene rings, it can be represented using a string 'A6 = 1, A4 =1' by the SOL method, which means one 6-member aromatic ring, attached with a 4-carbon aromatic ring increment (Quann and Jaffe, 1992). But in GC method, naphthalene is represented as a string 'aCH = 8, aC = 2', which means eight carbon atoms are connected with one hydrogen, while the other two carbon atoms without hydrogen connected (Marrero and Gani, 2002). This kind of information inconsistency reduces modular accuracy which further perturbs overall modelling results. Molecules are lumped to some degree in existing composition reconstruction frameworks, which makes the explored reaction mechanism difficult to be integrated, and even harder to predict the products. To overcome these challenges, in this work, a framework that is compatible between composition representation and property prediction, modular designed is developed, the unified information consistency framework would be flexible and easily adapted to different streams.

2. A novel framework for composition reconstruction of fuel streams

The proposed composition reconstruction framework comprises of two main sections, a) qualitative determination of molecules in streams, which includes structural characteristics representation and property estimation. b) quantitative determination of molecules in streams, which is transformation of the bulk properties into molecular composition by minimizing the differences between experimental properties and predicted properties implementing optimisation methods.

2.1 Molecule representation and property estimation

The simplified molecular-input line-entry system (SMILES) method was implemented to represent molecular structures explicitly (Weininger, 1988). Molecular characteristics such as molecular graph, bond connection information was extracted subsequently. Basic characteristics such as molecular weight, atom number or atom ratio (e.g. H/C), number of aromatic rings, number of aliphatic carbon cycles, etc., were directly calculated from molecular structure. Property estimation was implemented by an Artificial Neural Network (ANN) model, the ANN model was developed by fragmenting molecules into basic units according to the 'graph theory' using a novel molecular connectivity matrix transformation method, then model was trained on top of structural units and experimental property database. Evaluated using the largest normal boiling point property database (Alshehri et al., 2021), the developed ANN model has a better performance as compared with the GC method. Figure 1 shows the process of generating homologous series molecules of methylbenzene.

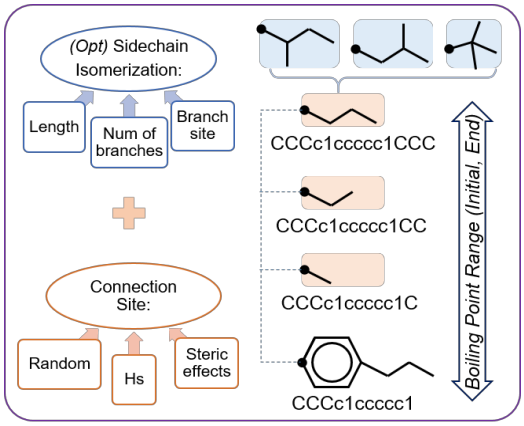

Figure 1: Proposed molecular representation method

Universal structure attributes were viewed as core structures, sidechains with different length were connected to core structures to form intact molecules, connecting sites were randomly selected according to both the available bond position and steric effects. It should be noted, only one sidechain was considered in this work due to the lack of carbon atom environment information from NMR analysis.
Physical properties such as normal boiling points of assembled molecules were estimated by the developed ANN model. Since normal boiling points increases over length of sidechains, normal boiling point range was used to constrain the upper and lower bound of sidechain length. Other molecules of various homologous series were generated similarly. The size of determined molecules is undoubtedly far below the sizes of the actual samples and is considered as a representative model.

2.2 Transformation between bulk properties and molecular composition

There are two findings that greatly reduce uncertainty of molecular composition reconstruction of oil fractions, the first is the content distribution of molecules in one homologous series subject to statistical functions. Another one is mixture properties can be calculated from pure-component properties based on mixing rules, note mixing can be nonlinear. Figure 2 shows a flowchart of molecular composition reconstruction process. Properties of each molecule were predicted by calling the trained ANN models. The Gamma probability density function (PDF) written as Eq.(1) and Eq.(2), were used to mathematically quantify fraction distribution of homologous series (Ren et al., 2019).

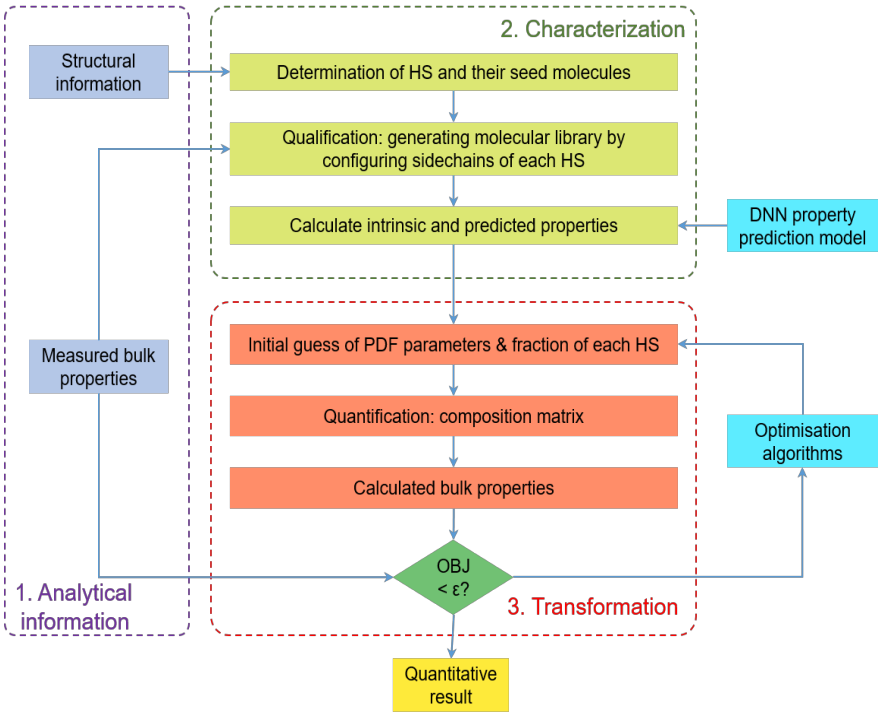

Figure 2: Flowchart of molecular composition reconstruction process

$$f(x) = \frac{(x-\gamma)^{\alpha-1}e^{-(x-\gamma)/\beta}}{\Gamma(\alpha)\beta^{\alpha}} \tag{1}$$

$$\Gamma(\alpha) = \int_0^{\infty} t^{z-1}e^{-t}dt \tag{2}$$

Where α, β and γ are shape, scale and location factors. Bulk properties were calculated from the pure-component properties based on the mixing rules. Experimentally available bulk properties were used for the optimization of the quantification. The objective function (Eq 3) was minimisation of differences between the calculated properties and experimental properties. Weight factors were used to convert the multi-objective optimisation to one-objective optimisation.

$$Obj = \sum_{k} abs\left(w_k \times \frac{C_k^{msd} - C_k^{pred}}{C_k^{msd}}\right) \quad (3)$$

Where superscripts *msd* and *pred* denote the measured and the predicted. Subscripts stands for the measurable properties usually comprised of a) temperature points of a distillation profile, b) measurably distinguished homologous series (PIONA), c) elemental content (in petroleum fractions and biomass, generally CHNOS are considered, while other elements like heteroatoms usually excluded because of their low content). d) other properties available such as specific gravity, molecular weight. w_i is the weighting factor of each item, items with higher priority were given larger weighting factor. Molar fraction $x_{i,j}^m$ was taken as the basis, mass fraction $x_{i,j}^w$ and volume fraction $x_{i,j}^v$ were calculated as follows:

$$x_{i,j}^w = \frac{x_{i,j}^m / MW_{i,j}}{\sum_{ii} \sum_{ij} x_{ii,jj}^m / MW_{ii,jj}} \quad (4)$$

$$x_{i,j}^v = \frac{x_{i,j}^w / SG_{i,j}}{\sum_{ii} \sum_{ij} x_{ii,jj}^w / SG_{ii,jj}} \quad (5)$$

$$x_j^w = \sum_{i} x_{i,j}^w \quad (6)$$

$$\sum_{j} x_j^w = 1 \quad (7)$$

$SG_{i,j}$, $MW_{i,j}$ stand for specific gravity and molecular weight of the molecule j of core structure i in the molecular repository.

A genetic algorithm is selected for an optimal solution searching, trade-off between minimizing the objective function and the computation power should be considered. Weight factors, variable space and parameters of the optimiser are defined heuristically.

3. Case study

To illustrate the capability and accuracy of the proposed framework, molecular composition of petroleum fraction and biomass derived oil are reconstructed. Universal molecular structure attributes in most samples are used as core/seed structures. Then molecule library of the sample to be modelled is generated by assembling core structures with sidechains following the proposed procedure. Based on reported molecular structure information obtained using GC, NMR and MS techniques, 37 seed molecules cover above oxygenates and hydrocarbons are used as core/seed structures, homologous series of each core were generated within the boiling point range of 323.15 -873.15 K. Consequently, 952 molecules were generated to represent the bio-oil sample. Similarly, 217 molecules within the boiling point range of 473-623 K were generated for the diesel sample. Molecular compositions are optimised, then mixture bulk properties of bio-oil sample and diesel sample are calculated and further compared with experimentally measured properties as presented in Table 1 and Table 2.

Table 1: Experimental and predicted bulk properties of the bio-oil sample

Properties		Measured	Predicted	ARE (%)
Element (wt.%)	C	0.72	0.72	0.01
	H	0.09	0.09	0.21
	O	0.20	0.20	0.00
Specific gravity		1.21	1.05	13.23
TBP Curve (ºC)	10	96.59	96.34	0.26
	30	115.82	119.29	3.00
	50	207.79	206.39	0.67
	70	292.24	292.73	0.17
	90	426.86	427.05	0.05

The predicted elemental contents and distillation results are in good agreements with results measured experimentally. Most of the absolute relative error (ARE) are within 1 %, the predicted specific gravities of the two samples have a greater deviation, one possible reason is the size of database is limited used in ANN

property estimation model training, although we used the largest database reported in literature. In addition, non-linear mixing behaviour of the specific gravity should not be neglected, but addressing the problem requires a lot of experimental data which is challenging at present.

Table 2: Experimental and predicted bulk properties of the diesel sample

Properties		Measured	Predicted	ARE (%)
Distillation Curve	10	267.96	266.41	0.58
(ASTM D86, ºC, v%)	30	275.98	276.51	0.19
	50	287.15	286.51	0.22
	70	299.49	299.82	0.11
	90	313.81	313.03	0.25
	100	325.58	328.90	1.02
Element (wt.%)	C	0.8715	0.8715	0.00
	H	0.1270	0.1270	0.04

One significant advantage of using SMILE representation is that detailed molecular information are retained. In Figure 3 the Sankey diagram displays molecules flow path between different level. No matter based on structure characteristic such as number of carbon atoms, or chemical properties like temperature cut, each molecule can be grouped into one specific category. The interconnection between different categories indicates that although previous frameworks have their own characteristics, but they can also be intrinsically connected. One potential application of this feature is that reported kinetic parameters can be used as reference.

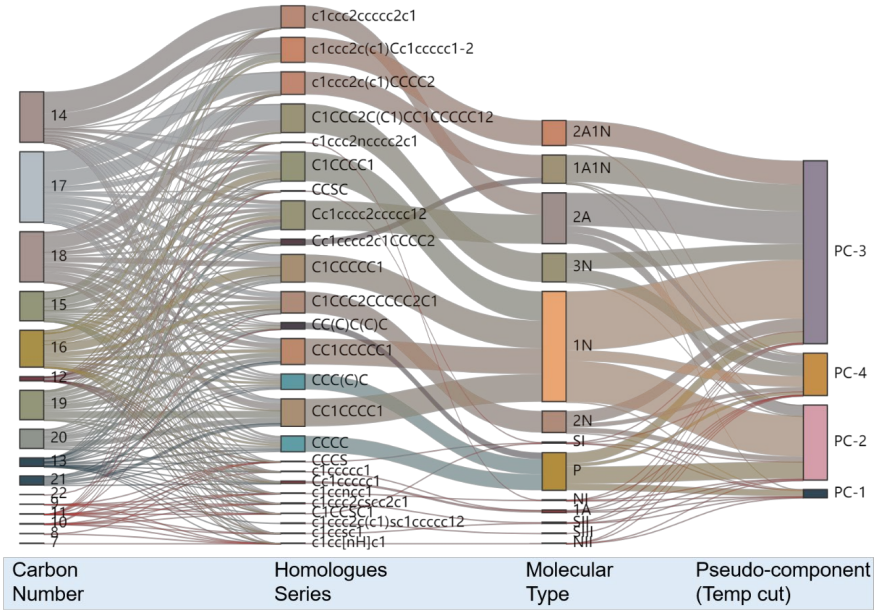

Figure 3: Information flow of molecules between different categories

4. Conclusions

A molecular composition reconstruction framework integrating molecular representation and bulk property transformation, has been proposed for composition modelling of bio-oil and diesel samples. One of the main advantages of this method over others is the detailed structure representation method that enables the introduction of data-driven ANN property prediction models, modular designed framework promotes the model adaptability. The proposed methodology was evaluated on two different types of oil, 217 and 952 were reconstructed to represent the diesel and the bio-oil streams. Most of the predicted properties are within 1 % ARE compared with experimental measured properties. Accurate result of the case studies indicates the proposed unified framework composition is promising in digitalisation of refinery streams. Molecular structure level representation makes composition information possibly connected with existing methods such as MTHS and Lumping method.

Acknowledgments

Qiong Pan would like to acknowledge the financial support by UoM-CSC joint studentship (No. 202006440006).

References

Alshehri A.S., Tula A.K., You F., Gani R., 2021. Next generation pure component property estimation models: With and without machine learning techniques. AIChE Journal, e17469.

Bojkovic A., Dijkmans T., Dao Thi H., Djokic M., Van Geem K.M., 2021. Molecular Reconstruction of Hydrocarbons and Sulfur-Containing Compounds in Atmospheric and Vacuum Gas Oils. Energy & Fuels, 35(7), 5777-5788.

Glazov N., Dik P., Zagoruiko A., 2021. Effect of experimental data accuracy on stochastic reconstruction of complex hydrocarbon mixture. Catalysis Today, 378, 202-210.

Hudebine D., Verstraete J.J., 2004. Molecular reconstruction of LCO gas oils from overall petroleum analyses. Chemical Engineering Science, 59 (22-23), 4755-4763.

Klein M.T., Hou G., Bertolacini R., Broadbelt L.J., Kumar A., 2005. Molecular modeling in heavy hydrocarbon conversions. CRC Press, Florida, US.

Marrero J.R. Gani R., 2002. Group-contribution-based estimation of octanol/water partition coefficient and aqueous solubility. Industrial & Engineering Chemistry Research, 41(25), 6623-6633.

Neurock M., Libanati C., Nigam A., Klein M.T., 1990. Monte Carlo simulation of complex reaction systems: Molecular structure and reactivity in modelling heavy oils. Chemical Engineering Science, 45 (8), 2083-2088.

Peng B., 1999. Molecular modelling of petroleum processes, PhD Thesis, University of Manchester, Manchester, UK.

Pinho A., Almeida M.B., Mendes F.L., Casavechia L.C., Talmadge M.S., Kinchin C.M., Chum H.L., 2017. Fast pyrolysis oil from pinewood chips co-processing with vacuum gas oil in an FCC unit for second generation fuel production. Fuel, 188, 462-473.

Quann R.J., Jaffe S. B., 1992. Structure-oriented lumping: describing the chemistry of complex hydrocarbon mixtures. Industrial & Engineering Chemistry Research, 31(11), 2483-2497.

Ren Y., Liao Z., Sun J., Jiang B., Wang J., Yang Y., Wu Q., 2019. Molecular reconstruction: Recent progress toward composition modelling of petroleum fractions. Chemical Engineering Journal, 357, 761-775.

Tullo A.H., 2019. Why the future of oil is in chemicals, not fuels? Chemical & Engineering News. <https://cen.acs.org/business/petrochemicals/future-oil-chemicals-fuels/97/i8>, accessed 20.03.2022.

Weininger D., 1988. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. Journal of Chemical Information and Computer Sciences, 28 (1), 31-36.

Wang M., Zhao S., Chung K.H., Xu C., Shi Q., 2015. Approach for selective separation of thiophenic and sulfidic sulfur compounds from petroleum by methylation/demethylation. Analytical Chemistry, 87 (2), 1083-1088.

Wu Y., 2010. Molecular management for refining operations, PhD Thesis. The University of Manchester, Manchester, UK.