Natural Resources Canada | Ressources naturelles Canada

Advanced Data Processing Techniques in GC×GC-TOFMS for Biocrude Analysis

SMILEs and be SMARTs



Rafal Gieleciak*, Anton Alvarez-Majmutov, Sandeep Badoga, Jinwen Chen

Natural Resources Canada

CanmetENERGY Devon

CanmetENERGY
Leadership in ecoInnovation

16th Multidimensional Chromatography (MDC) Workshop
Liege, Belgium – February 3-5, 2025

16TH Multidimensional Chromatography Workshop

Canada

CanmetENERGY (part of Natural Resources Canada - NRCan)

Canada's leading research and technology organization in the field of clean energy

Responsible for developing the connection between energy efficiency and technology policies, programs, and R&D areas

- Buildings and Communities
- Industrial processes
- Renewable & Distributed Energy
- RETScreen International

Varennes

Downstream & Renewables
- Upstream & Environment

Devon

Advanced materials for:
- Energy end-use, production, and distribution (pipelines)
- Safety and Security

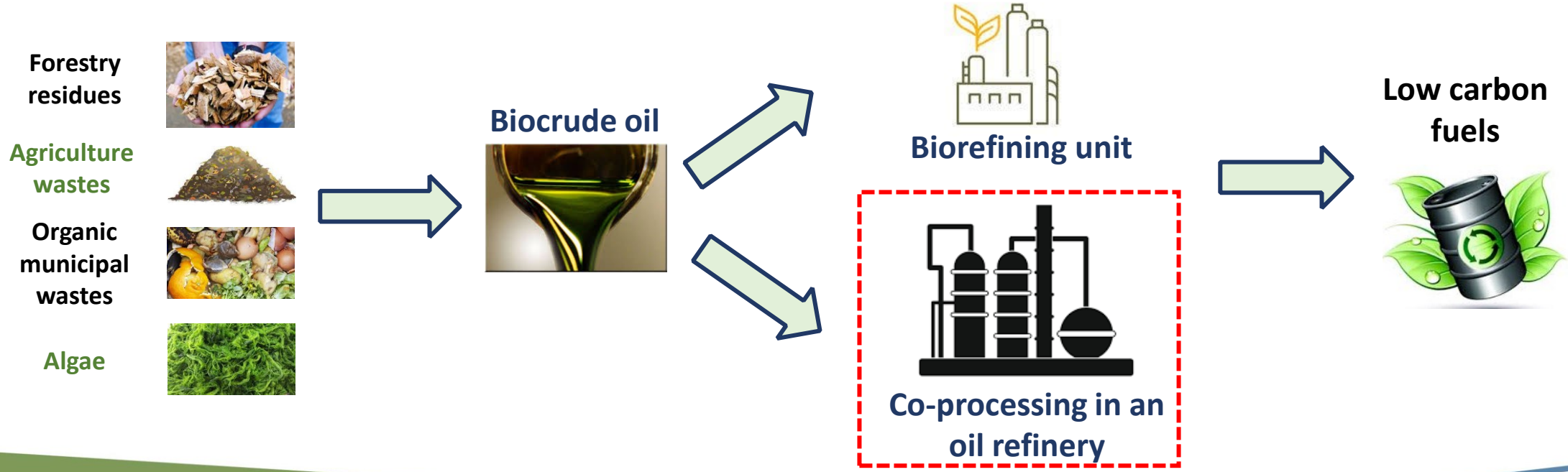
Calgary and Hamilton

- Clean fossil fuels
- Renewables
- Energy end-use

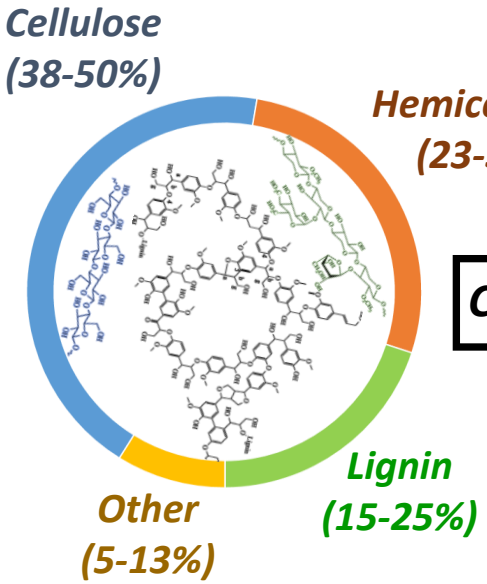
Ottawa

Integrating biocrude into petroleum refineries

- Bioenergy is expected to play a significant role in achieving Canada's net zero emissions target by 2050.
- Biofuel production emits less GHG than fossil fuel production on a life cycle basis.
- Stand-alone biorefineries require large capital investment.
- Co-processing biocrude with petroleum in refineries is attractive and practical.

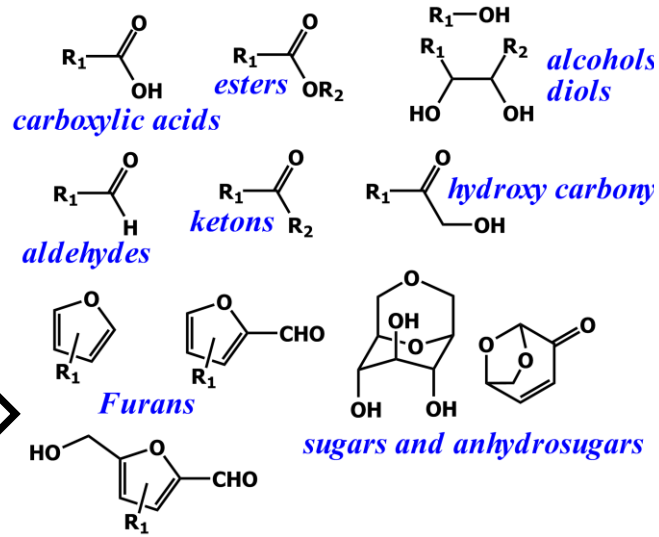


Biomass/biocrude composition



Biomass

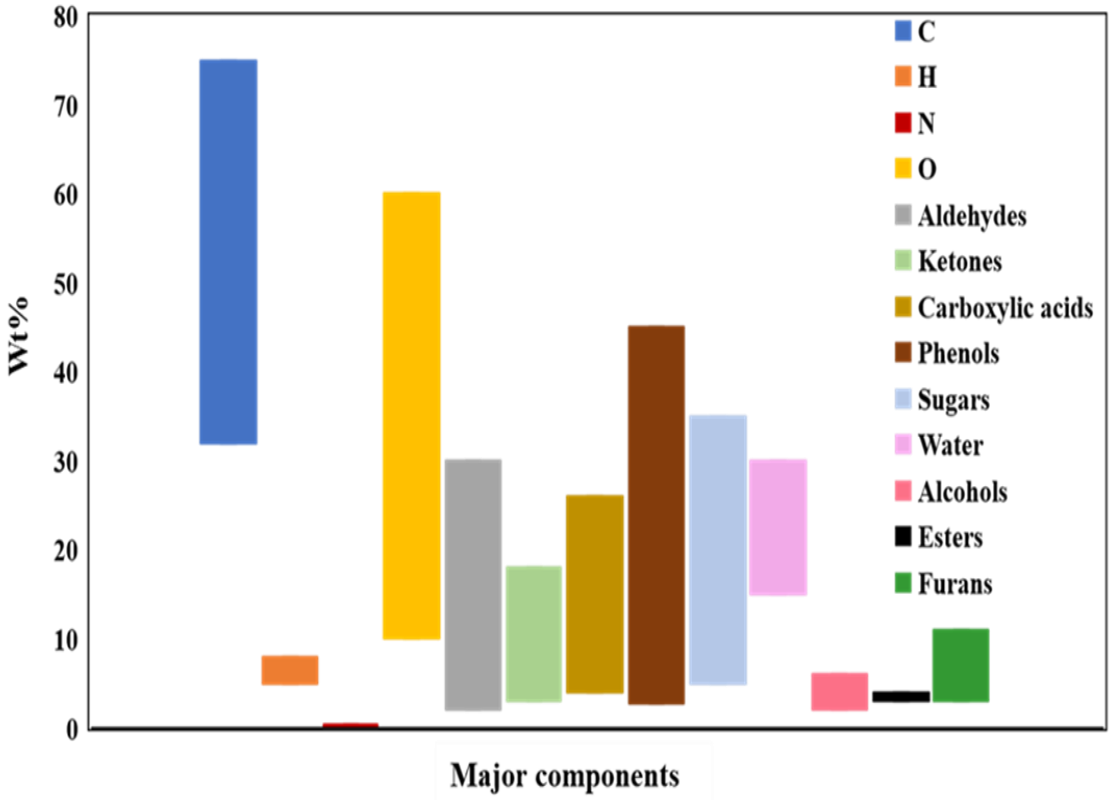
Conversion



+ many others

Biocrude

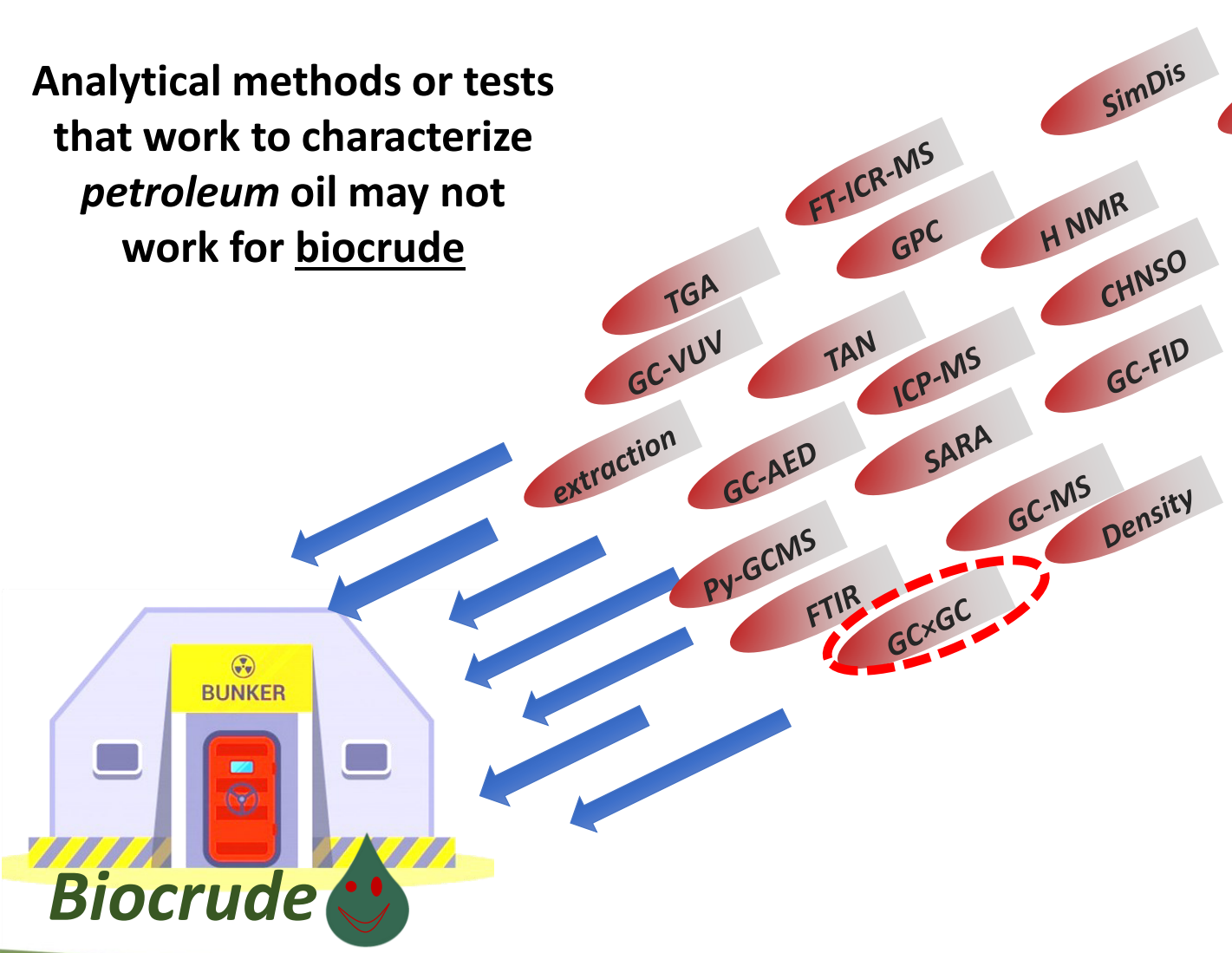
(quite a complex mixture)

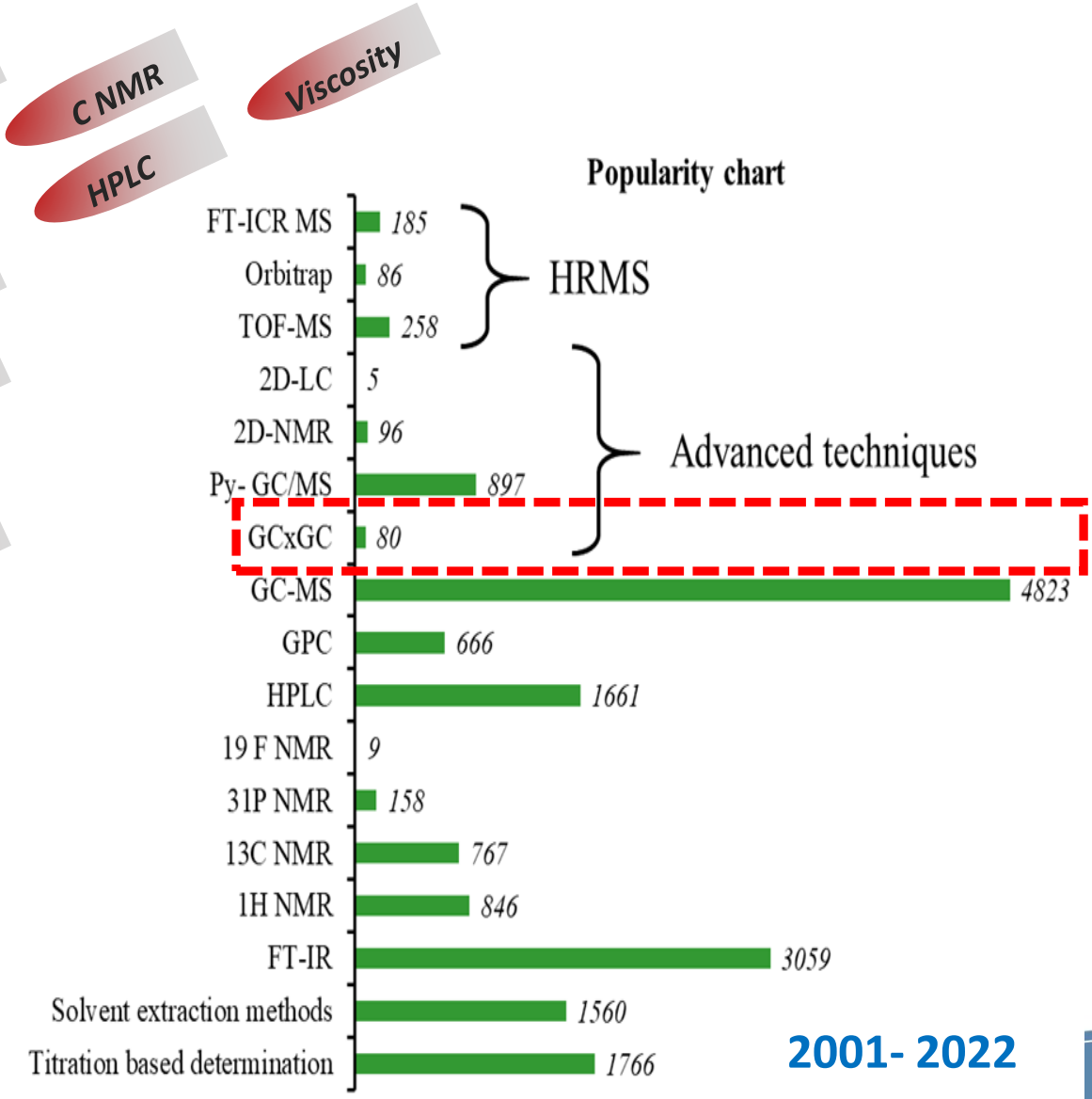


Chemical composition of biocrude derived from various types of biomass (including woody biomass and forestry residue)

Analytical methods or tests

Analytical methods or tests that work to characterize *petroleum* oil may not work for <u>biocrude</u>

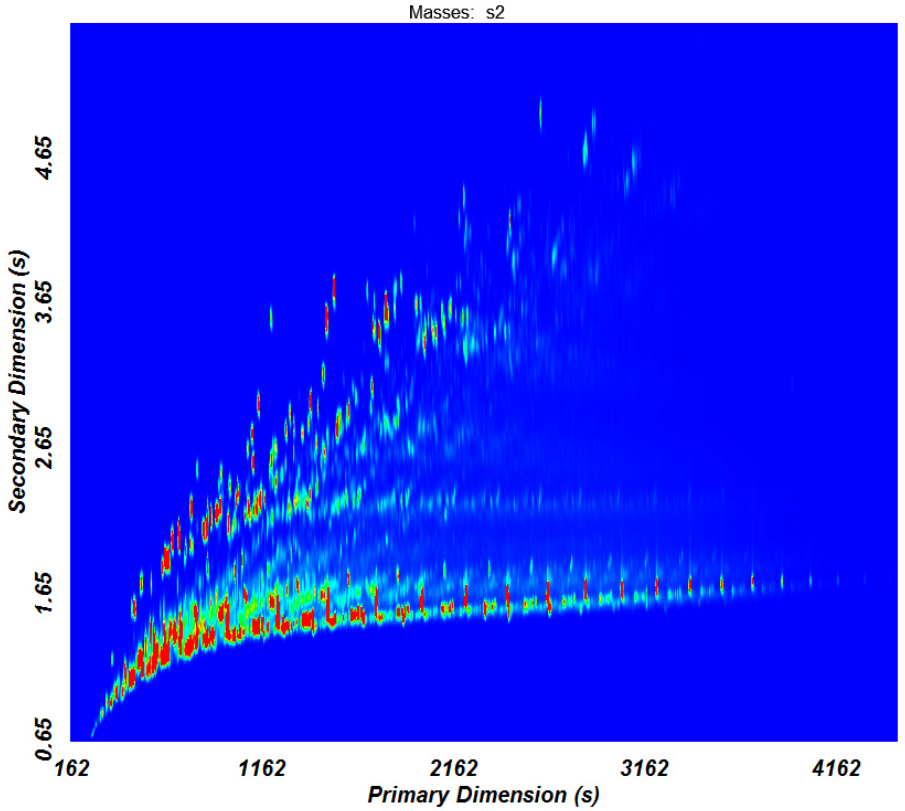




2001- 2022

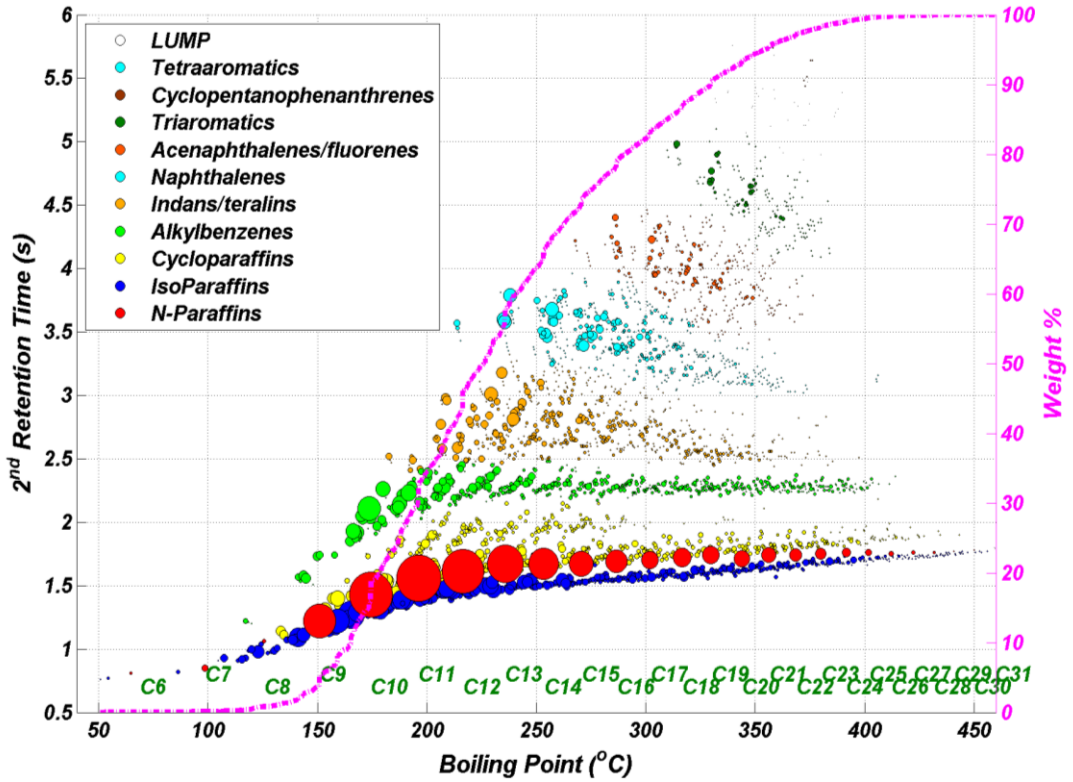
Speciation: 2D gas chromatography (GC×GC)

Structured chromatogram (petroleum-based oil)



GC × GC-FID contour map

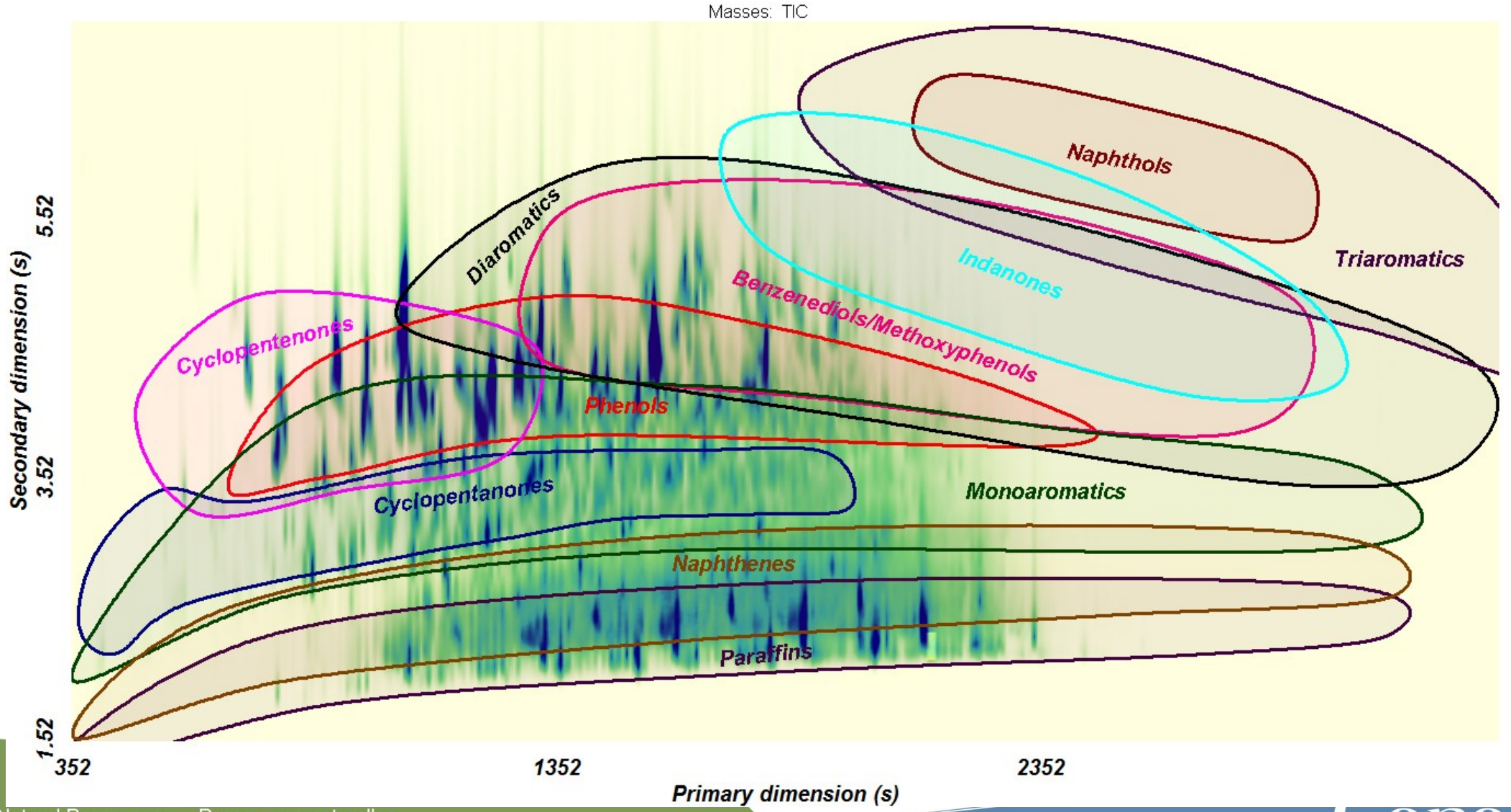
maximizing information



Boiling Point vs Polarity + HC classes by colours + GC × GC SimDis on the plot

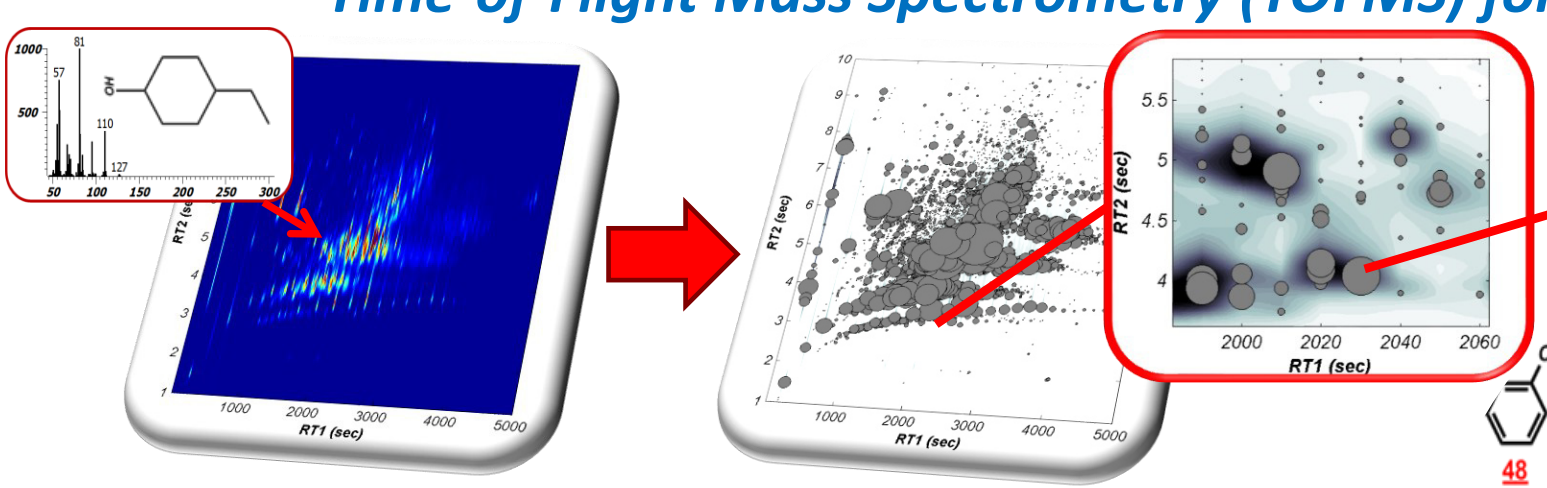
Classification types - mapping

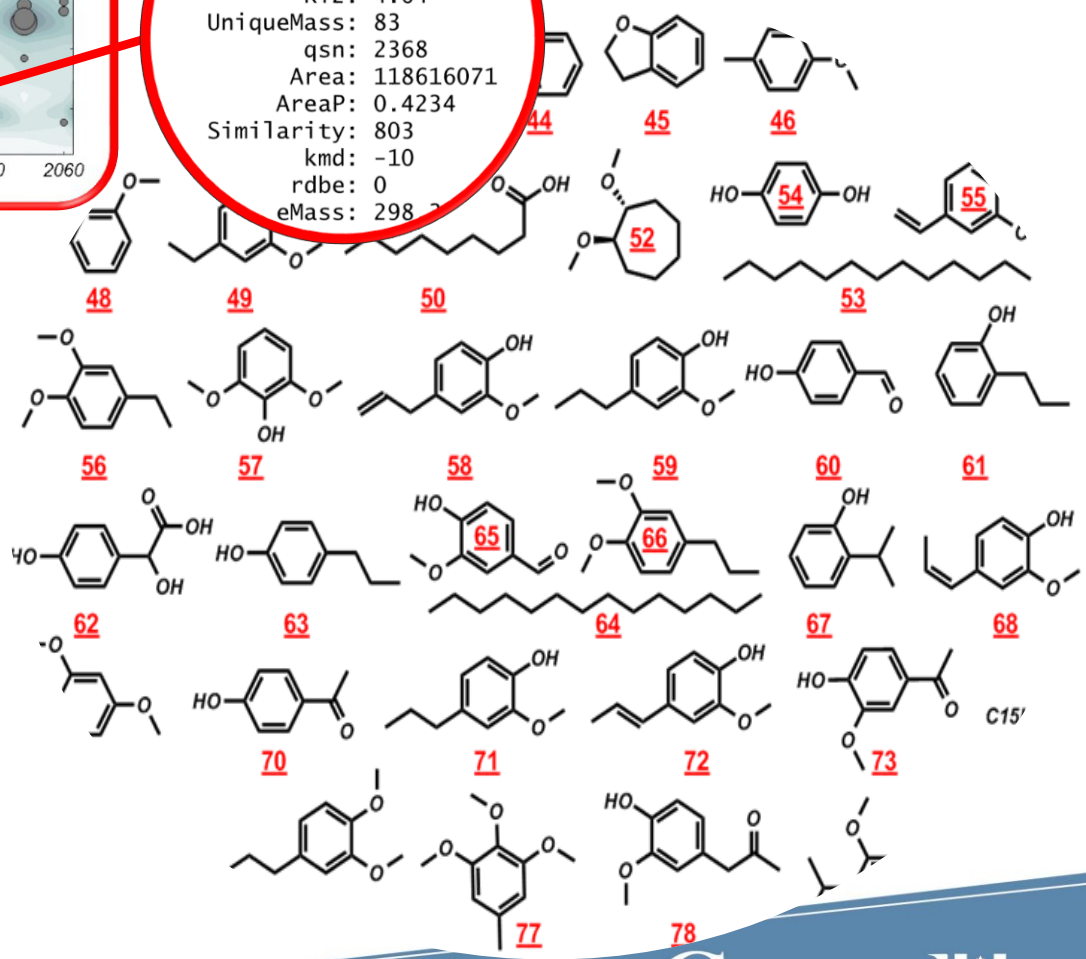
GC×GC-FID - Upgraded (lightly) biocrude sample



GC×GC: Biocrude

Time-of-Flight Mass Spectrometry (TOFMS) for species identification





- GC×GC-FID can speciate and quantify but only for samples with low heteroatom content
- GC×GC-TOFMS identifies hundreds of compounds, forming the basis for accurate sample representation.
- *Large datasets increase the risk of human error in manual analysis.*
- Automatic data handling may improve accuracy and reduces errors in biocrude analysis.

Proposed Workflow for ~~Advanced~~ *Alternative* Data Processing

Analysis

GC×GC-TOFMS

Processing — ChromaTOF

Peak Table/MS spectra

Data Handling

Chemical Names

Conversion — OPSIN or PubChem

Chemical Structures

Extraction — RDKit or

Chemical Descriptors

Calculation — RDKit

Functional group analysis

Traditional

Chemical Composition

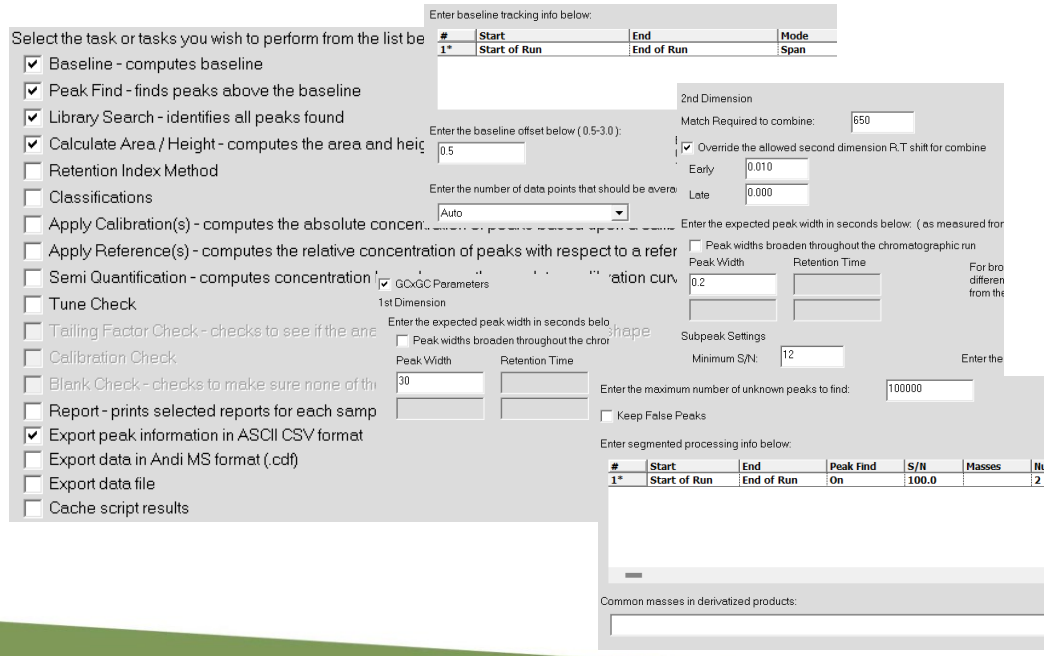
Peak Table (setup)

ChromaTOF software (4.50)

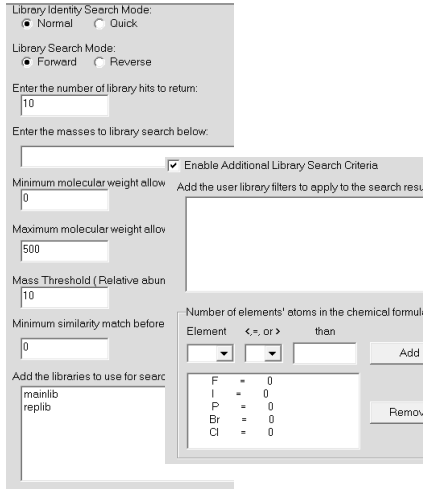
File.smp
(500 Mb)

File.png
(100 Mb)

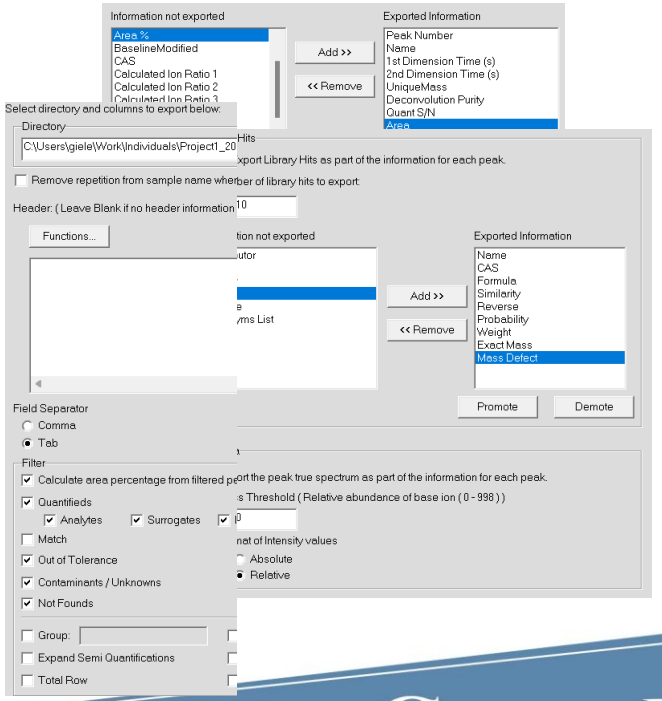
Baseline & Peak Find & Area



Library Search



Exporting



Peak Table

Exporting Data and Handling Files

Peak Table information and Mass Spectra saved in *.txt file

	A	B	C	D	E	F	G	H	I	J	K	
										res5		
	PeakNumber	Name	stDimensio...	ndDimensi...	UniqueMass	Deconvolut...	QuantSN	Area	Hit1Name	Hit1CAS	Hit1Formula	Hi
	Number	Text	Number	Number	Number	Number	Number	Number	Text	Text	Categorical	Number
1	Peak Numb...	Name	1st Dimensi...	2nd Dimens...	UniqueMass	Deconvolut...	Quant S/N	Area	Hit 1 Name	Hit 1 CAS	Hit 1 Formu...	Hit 1 Simila...
2	1	Ethanol	170	9.150	45	0	21.466	347136	Ethanol	64-17-5	C2H6O	896
3	2	dl-Stachydr...	180	4.950	36	0	5442.1	1093234689	dl-Stachydr...	32039-73-9	C7H13NO2	482
4	3	dl-Stachydr...	180	9.250	49	0	1193.4	18368951	dl-Stachydr...	32039-73-9	C7H13NO2	475
5	4	2-Butanone	190	3.350	88	0	1325.2	47765150	2-Butanone	78-93-3	C4H8O	828
6	5	2-Butanone	190	3.400	56	0	1748.7	72680746	2-Butanone	78-93-3	C4H8O	901
7	6	Imidazole, ...	190	4.400	84	0	798.96	78200818	Imidazole, ...	0-00-0	C5H6N2O2	456
8	7	4-Cyclopen...	190	9.500	47	0	7661.8	524450583	4-Cyclopen...	694-47-3	C5H8O2	558

	Text
1	Spectra
2	45:999 46:344 43:343 44:193 42:128 41:56 47:30 38:12 36:11 35:8 37:6 49:6 39:5 58:4 52:3 50:3 400:2 361:2 188:2 358:2 156:2 324:2 51:2 436:2 87:2 362:2 493:
3	49:999 84:523 35:491 47:286 86:265 51:262 37:154 48:134 36:59 50:50 88:39 41:38 42:26 85:20 83:20 38:19 52:12 87:10 70:9 54:7 55:7 57:7 58:7 53:7 59:7 56:7
4	49:999 84:562 86:361 51:309 47:256 35:244 48:116 37:82 88:58 50:47 41:32 36:31 42:21 83:20 85:19 38:10 87:6 70:6 82:6 52:4 72:3 43:3 60:1 59:1 58:1 54:1 89:
5	43:999 72:212 42:113 49:38 51:29 44:27 55:24 84:22 53:20 47:15 37:15 86:14 38:12 48:12 56:10 52:10 73:9 54:9 35:7 36:3 88:3 78:3 85:3 69:2 68:2 70:2 46:1 87:
6	43:998 72:140 42:79 57:62 39:48 41:43 44:28 55:18 38:16 50:14 56:13 37:11 53:11 45:8 51:8 49:7 40:6 73:6 71:6 54:4 52:3 58:2 36:2 46:1 70:1 48:1 35:0 84:0 47:
7	49:999 84:622 86:379 51:333 35:242 47:239 48:135 37:90 88:57 36:30 83:17 46:13 85:10 87:6 82:2 60:2 92:2 90:1 378:1 382:1 94:1 343:1 418:1 157:1 249:1 372.
8	72:998 46:463 48:437 47:433 51:422 53:416 49:415 50:412 45:409 54:407 52:406 55:384 44:349 57:278 37:161 36:152 35:151 73:116 38:112 39:97 74:60 76:55
9	43:999 72:219 57:98 39:22 45:20 44:15 60:14 38:12 73:11 37:10 71:8 56:4 58:3 40:2 88:1 62:1 85:1 84:1 61:1 87:0 59:0 69:0 64:0 227:0 228:0 357:0 230:0 356:0 2
10	49:998 84:538 86:361 47:258 50:208 35:202 37:135 36:123 51:97 48:93 88:86 46:67 77:60 79:52 109:37 58:34 81:32 80:31 64:28 100:27 144:27 107:27 332:24 4

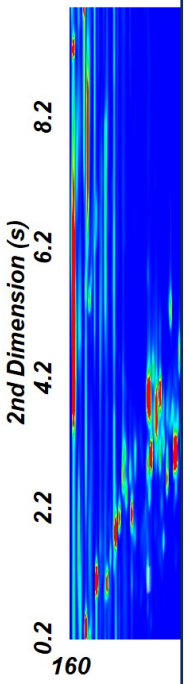
```
fTOFMS =
        Data: [1x1 struct
        Hit1: [1x1 struct
        …
       Hit10: [1x1 struc
          MS: [1x1 struc
```

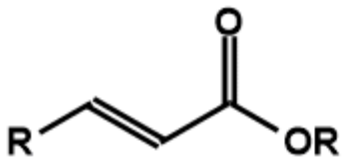
```
fTOFMS.Data =
    peaknumber: [N×1 double]
          name: {N×1 cell}
           rt1: [N×1 double]
           rt2: [N×1 double]
    uniquemass: [N×1 double]
     deconvpur: [N×1 double]
       quantsn: [N×1 double]
          area: [N×1 double]
     areaperc: [N×1 double]
```

```
fTOFMS.Hit1 =
                  name: {N
                   cas: {N
               formula: {N
            similarity: [N×1 d
               reverse: [N×1 d
           probability: [N×1 d
                weight: [N×1 double]
             exactmass: [N×1 double]
           massdefect: [N×1 double]
           formulapeel: {N×1 cell}
                  mass: [N×1 double]
                   kmd: [N×1 double]
                   dbe: [N×1 double]
                iabund: [N×9 double]
              chnossi: [N×6 double]
                     z: [N×1 double]
        formulageneral: {N×1 cell}
```

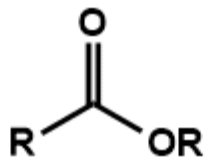
```
fTOFMS.MS =
    peaknumber: [N×1 double]
       spectra: {N×1 cell}
```

Example: Algae biocrude

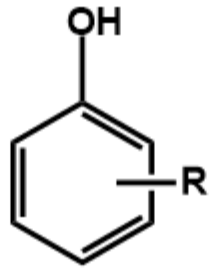




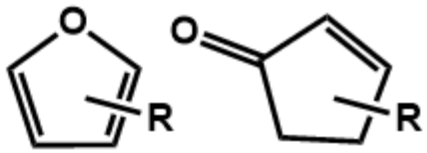
$C_nH_{2n-2}O_2$
KMD: -59
17.8%



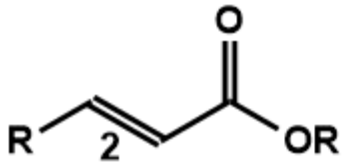
$C_nH_{2n}O_2$
KMD: -46
12.2%



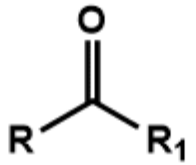
$C_nH_{2n-6}O$
KMD: -63
3.7%



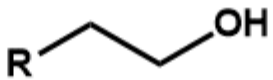
$C_nH_{2n-4}O$
KMD: -49
3.2%



$C_nH_{2n-4}O_2$
KMD: -73
3.0%



$C_nH_{2n}O$
KMD: -23
1.4%

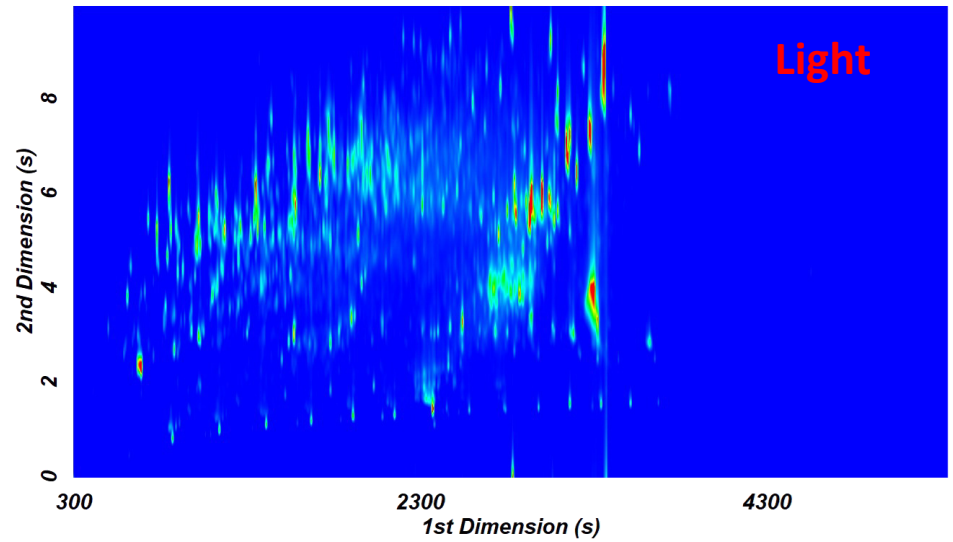


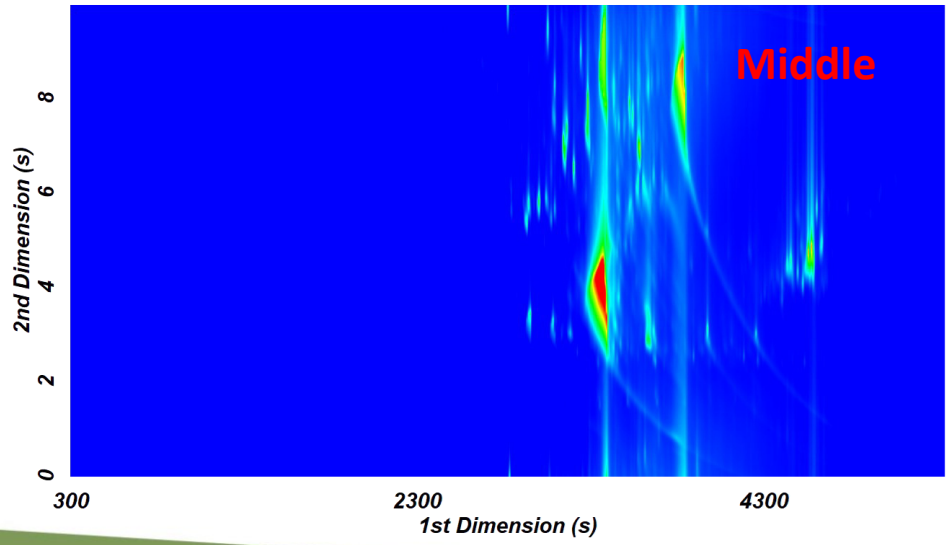
$C_nH_{2n+2}O$
KMD: -9
0.75%

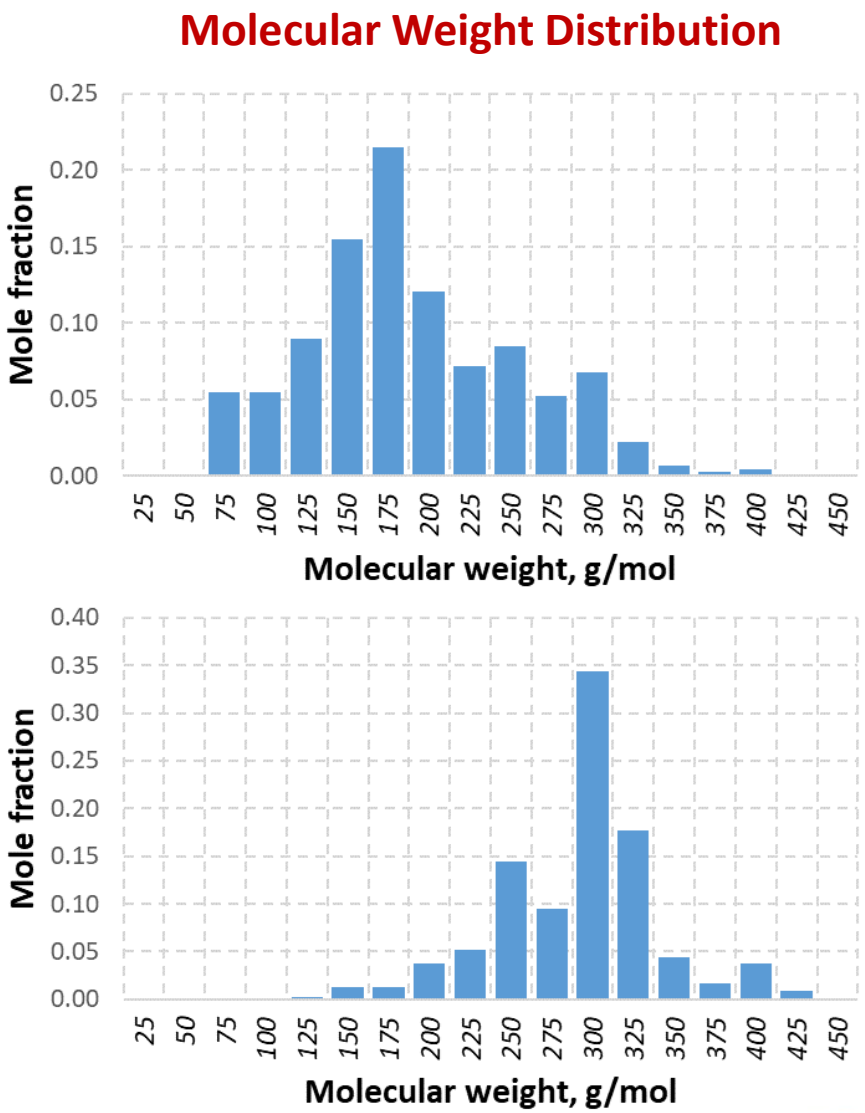
\+
Other
Compounds

CHO

Example: HTL biocrude fractions







HTL – hydrothermal liquefaction

Extracting Every Bit of Information

ChromaTOF

Name	CAS
p-Cresol	106-44-5
Tetradecane	629-59-4
Toluene	108-88-3
Naphthalene	91-20-3
Phenol, 2,4,6-trimethyl-	527-60-6
Phenanthrene	85-01-8
Phenol, 2,6-dimethyl-	576-26-1
Phenol, 4-ethyl-2-methoxy-	2785-89-9
Benzene, pentyl-	538-68-1
Toluene	108-88-3
Phenol, 2,3,6-trimethyl-	2416-94-6
1,2-Benzenediol, 3-methyl-	488-17-5
Heptadecane	629-78-7
Phenol, 4-ethyl-2-methoxy-	2785-89-9
o-Cymene	527-84-4
Benzene, 1,3-dimethyl-	108-38-3
Phenol, 2,3,6-trimethyl-	2416-94-6
Phenol	108-95-2

Chemical Names
CAS #
(Chemical Abstracts Service Registry #)

How to make use of that information?

A database of chemical structures can be created

SMILES ☺

Simplified Molecular Input Line Entry System

Structures enable us to enhance our data by computing compound- or sample-specific characteristics

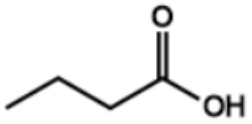
Functional Group Analysis,
Boiling point,
Density,
Retention idx,
Heat of formation,
Melting point,
…

Name 2 Structures

SMILES (Simplified Molecular Input Line Entry System)

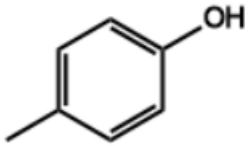
n-pentane

CCCCC



butyric acid

CCCC(O)=O



p-cresol

Cc1ccc(O)cc1

HOW?

OPSIN

Open Parser for Systematic IUPAC nomenclature

J. Chem. Inf. Model. 2011, 51, 3, 739–753

- Converts IUPAC names to structures.
- Free, open-source, and API-friendly.
- Accurate, following IUPAC rules.
- Fast

PubChem

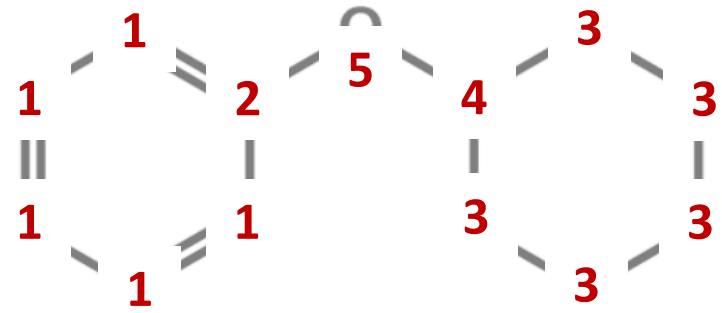
PubChem Power User Gateway (PUG)

- Use CAS No
- Automates access to PubChem data.
- Integrates via web APIs.
- Slow
- Supports property calculations (LogP, PSA, HBD/A, Solubility, Polarity, NMR, MS, etc)

Property Calculations

Group Contribution Method (GCM)

- **GCM estimates molecular properties by summing ator contributions**
- Quick, compu for large datas
- Accuracy depe



1: [cX3H1;R]
2: [cX3H0;R]
3: [CX4H2;R]
4: [CX4H1;R]
5: [OX2H0;!R;$([OX2H0]-[cX3H0]);$([OX2H0]-[C])]

SMARTS

SMILES Arbitrary Target Specification

Substructure Searching

RDKit

Open-Source Cheminformatics and Machine Learning

Recipe:

1. *Decompose m groups.*
2. *Assign predef group.*
3. *Add group co*

534.2 K

982 cm³/g

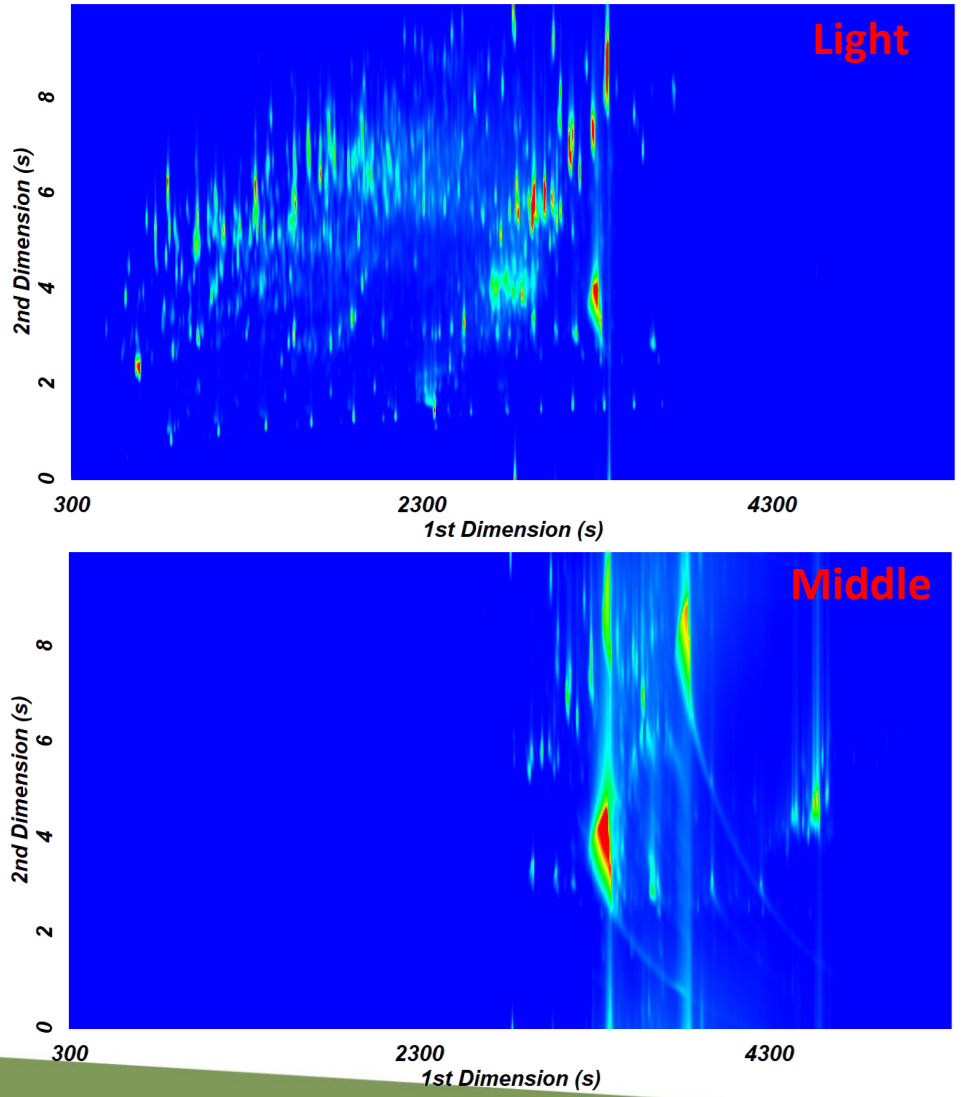
ntribution

	Density
7	56.1
94	-6.15
6	69.65
19	-1.16
36	3.07
96	50
49	171.51
	176.25

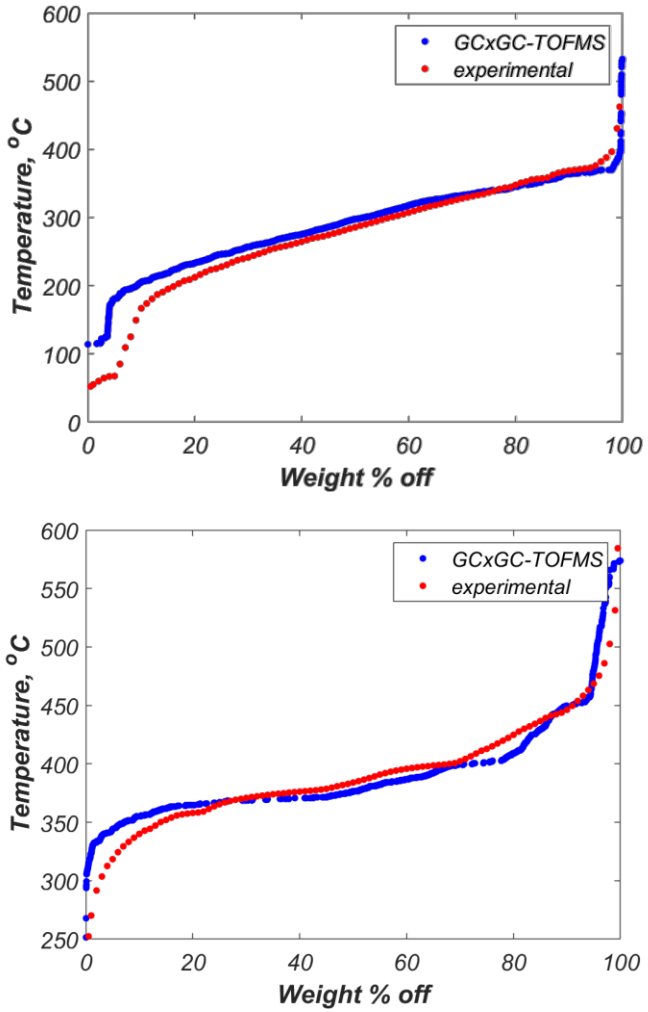
D = 176.25/171.51 = **1.028 cm³/g**

Pereira de Oliveira et al. Energy Fuels 2013, 27, 3622–3641; Joback et al. Chem. Eng. Commun. 1987, 57, 233–243.

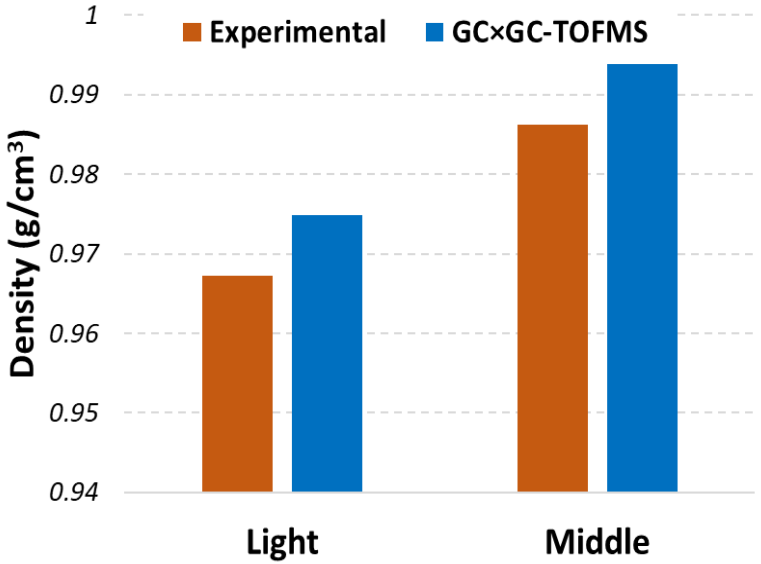
Example: HTL biocrude fractions



Simulated Distillation



Density (15.6 °C, g/cm³)



Automatic Functional Group Analysis

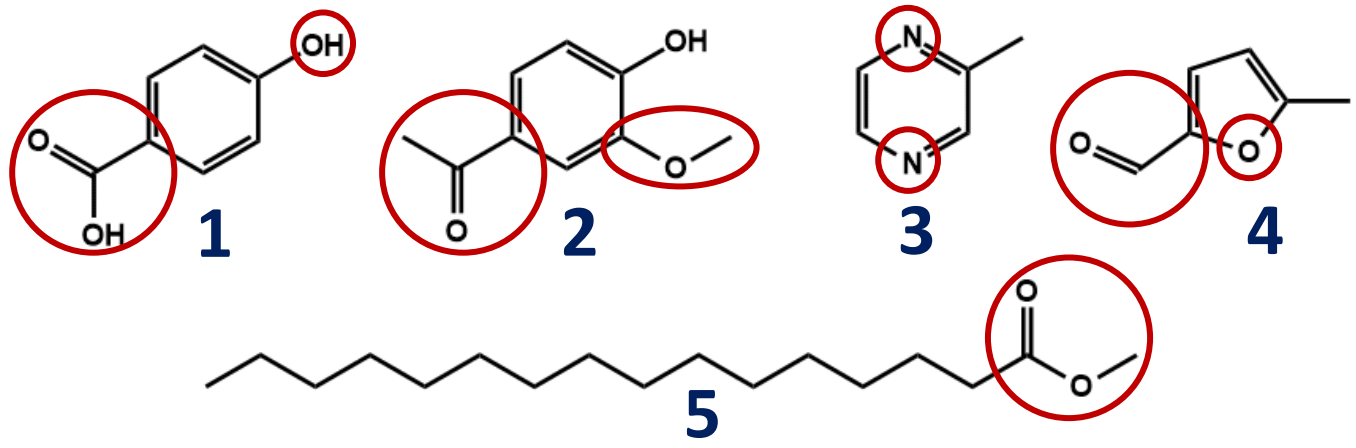
- Functional group analysis: categorizing compounds into chemical families.
- Considers all selected functional groups in a compound/sample.
- Uses a fragmentation algorithm with SMARTS codes.

- *The **automated** method calculates the mass fraction of each functional group present in a compound* ($mf_{fg,i}$) and uses these fractions to weight the contribution of each group to the sample's overall reporting.

$$C_{fg} = \sum_{i=1}^{n} C_i \times mf_{fg,i}$$

C_{fg} - Concentration of functional group in the sample

C_i - Concentration of a compound

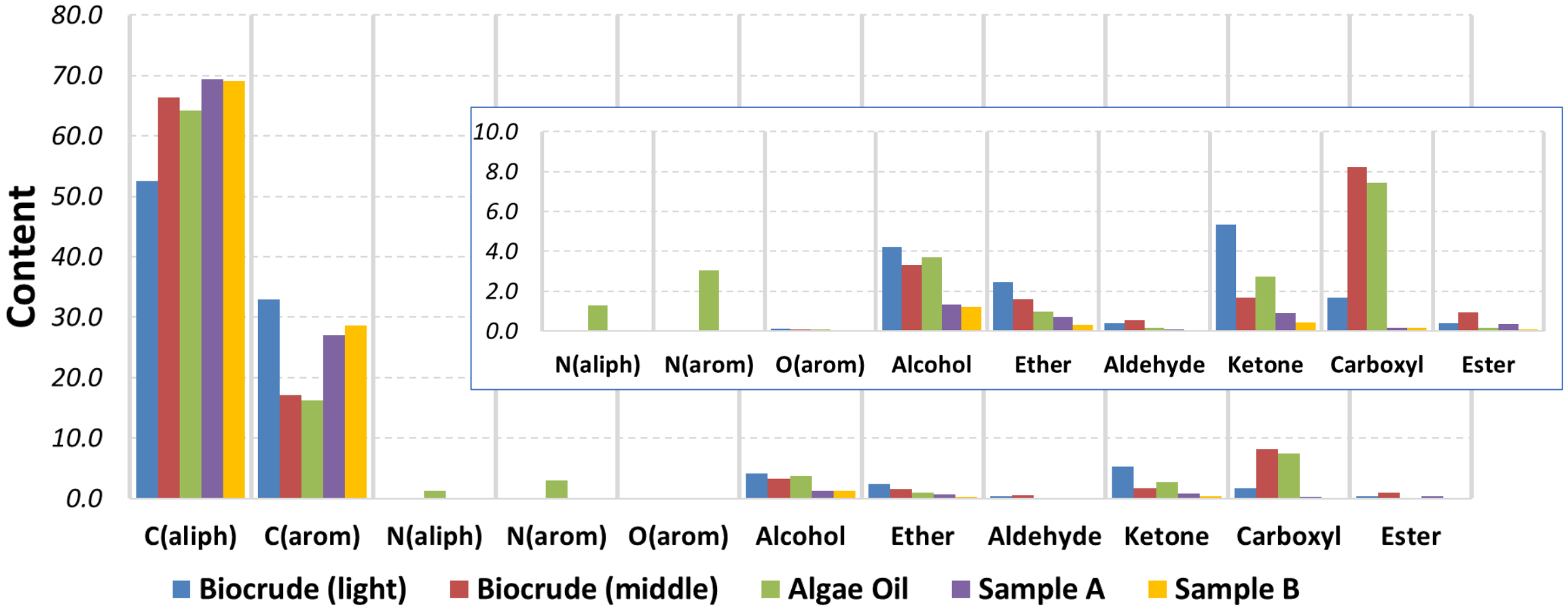


Examples of functional group mass fractions

Molecule	C(aliph)	C(arom)	N(arom)	O(arom)	Alcohol	Ether	Aldehyde	Ketone	Carboxyl	Ester
mol 1	0.0	55.1	0.0	0.0	12.3	0.0	0.0	0.0	32.6	0.0
mol 2	0.0	30.7	0.0	0.0	10.2	25.9	0.0	33.1	0.0	0.0
mol 3	16.0	54.3	29.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0
mol 4	13.7	45.5	0.0	14.5	0.0	0.0	26.4	0.0	0.0	0.0
mol 5	78.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	21.8

Pecchi & Goldfarb RSC Sustainability, 2024, 2, 1444–1455

Examples


Content
80.0
70.0
60.0
50.0
40.0
30.0
20.0
10.0
0.0
C(aliph)
C(arom)
N(aliph)
N(arom)
O(arom)
Alcohol
Ether
Aldehyde
Ketone
Carboxyl
Ester
10.0
8.0
6.0
4.0
2.0
0.0
N(aliph)
N(arom)
O(arom)
Alcohol
Ether
Aldehyde
Ketone
Carboxyl
Ester
Biocrude (light)
Biocrude (middle)
Algae Oil
Sample A
Sample B

Conclusions

Cmpd Names → Structures → Properties

- The proposed method provides a comprehensive and automated approach to GC×GC-TOFMS data analysis by translating NIST data search results into molecular structures/properties
- Custom MATLAB scripts enable full automation, including (1) Functional group mass fraction analysis, (2) Bulk property calculation, etc.
- Offers deeper insights into biocrude composition for process optimization, improving yield and efficiency → and contributing to the idea of **Molecular Reconstruction** of (bio)oil.

Future Directions (Mindful of GIGO: Garbage In, Garbage Out):

- Refine and expand the workflow for diverse biocrude samples.
- Improve quantitative accuracy using proper calibration.
- Implement chemometric models (e.g., Kurt Varmuza's approach) to validate compound identities and reduce false positives.

Acknowledgments

- The Office of Energy Research and Development (OERD) of NRCan
- Government of Canada's interdepartmental Program of Energy Research and Development (PERD)
- CanmetENERGY Devon Pilot Plants and Analytical Lab
- Industrial collaborators and partners

Dr. Rafal Gieleciak
Research Scientist, Team Lead
Downstream and Renewables
CanmetENERGY Devon
Natural Resources Canada
Rafal.Gieleciak@nrcan-rncan.gc.ca
Telephone: 780-233-5853

Canada