



Natural Resources
Canada

Ressources naturelles
Canada

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

SMILE_s and be SMART_s

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

Natural Resources Canada

CanmetENERGY

CanmetENERGY Devon

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) ²

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

Vareennes

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



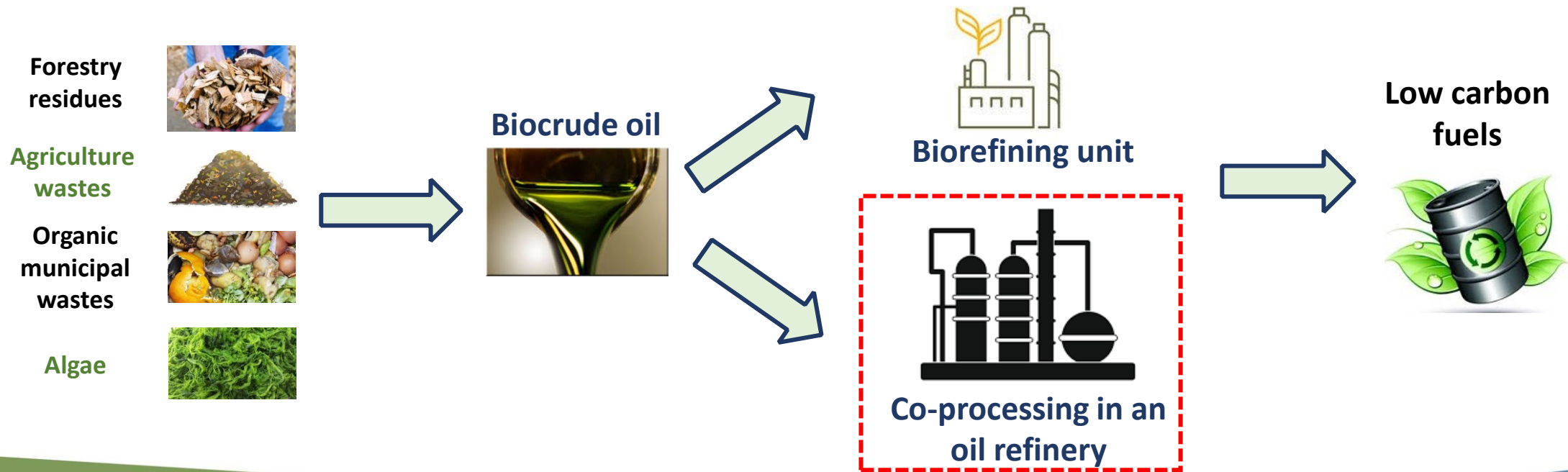
Natural Resources
Canada

Ressources naturelles
Canada

Canada

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

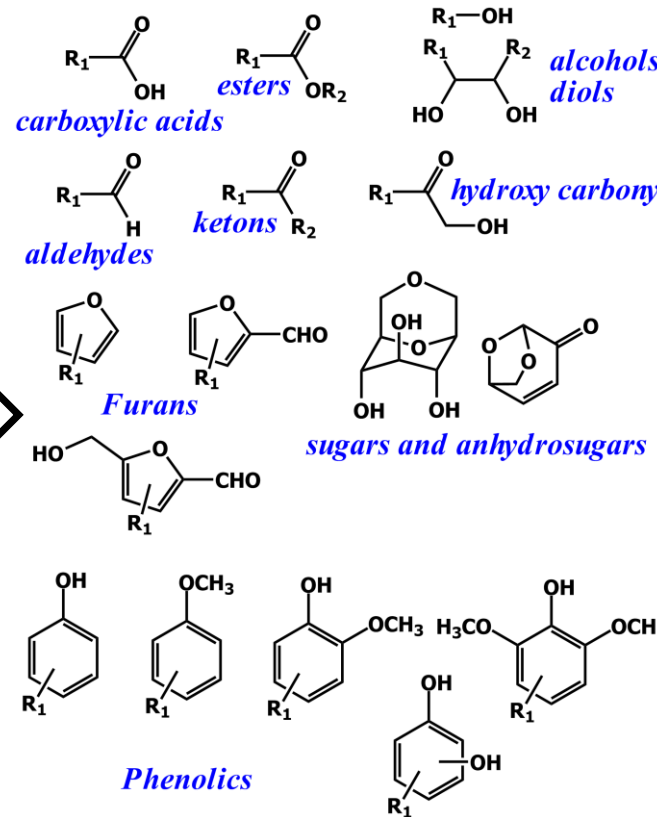
Cellulose
(38-50%)

Hemicellulose
(23-32%)

Lignin
(15-25%)

Other
(5-13%)

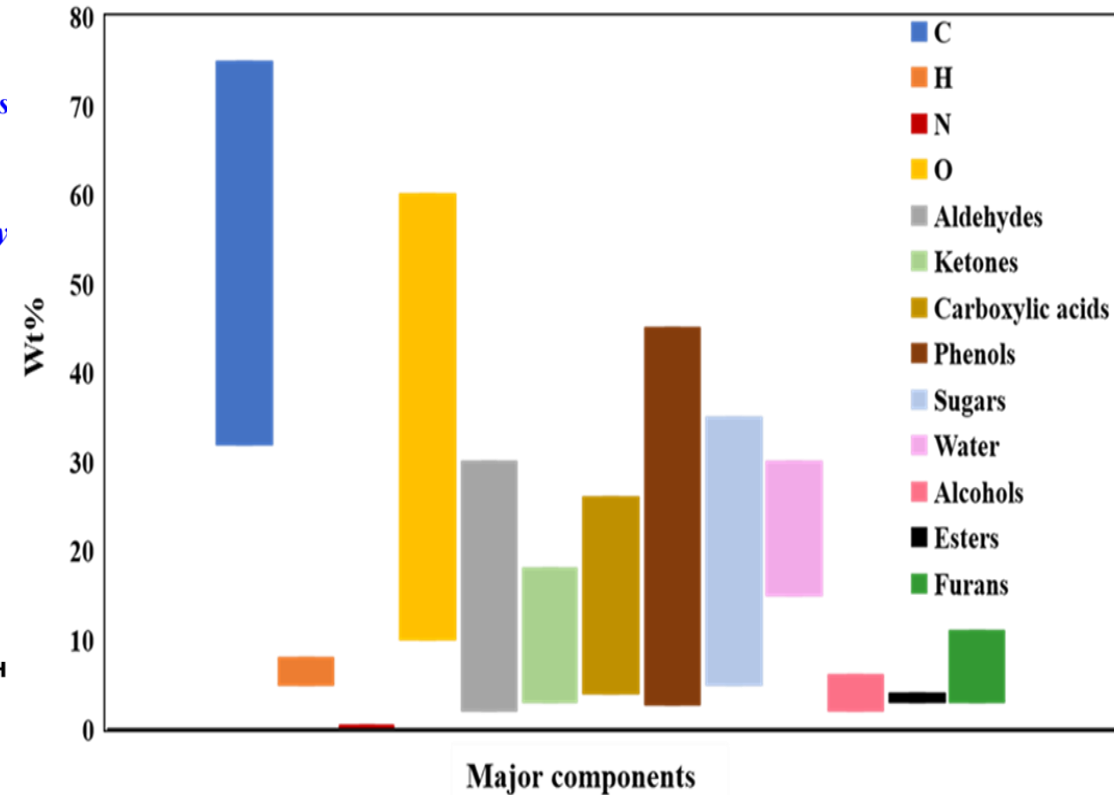
Conversion



+ many others

Biocrude

(quite a complex mixture)



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

Biomass



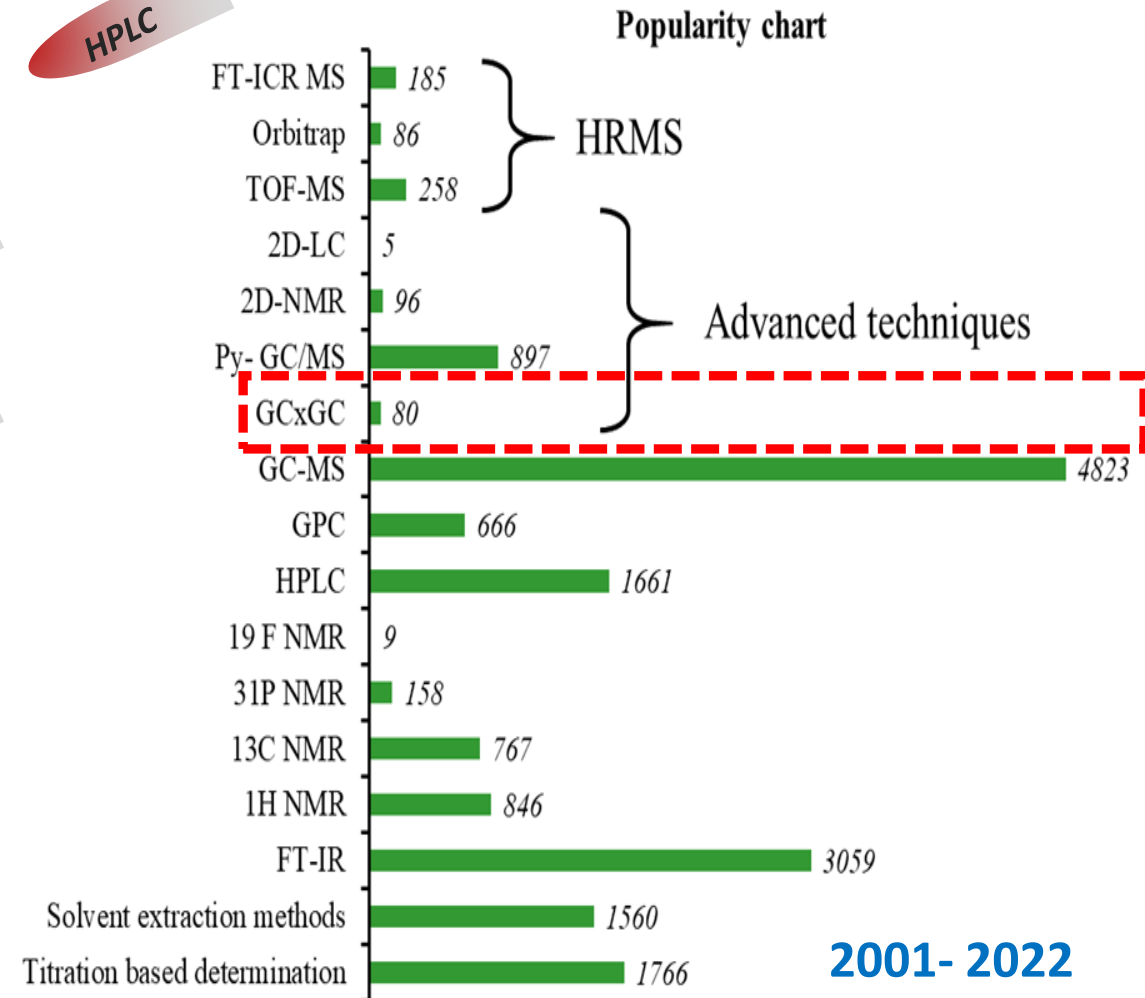
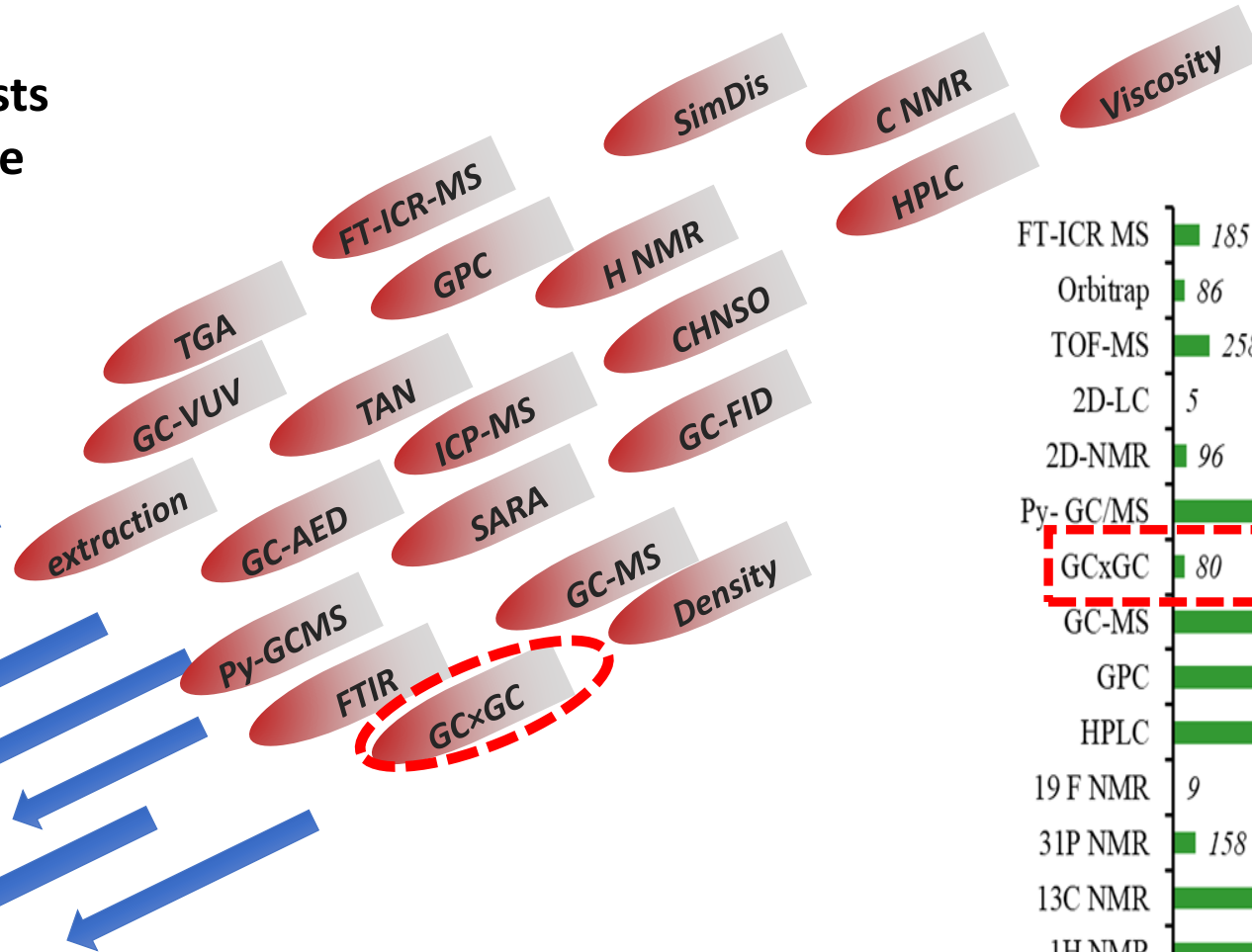
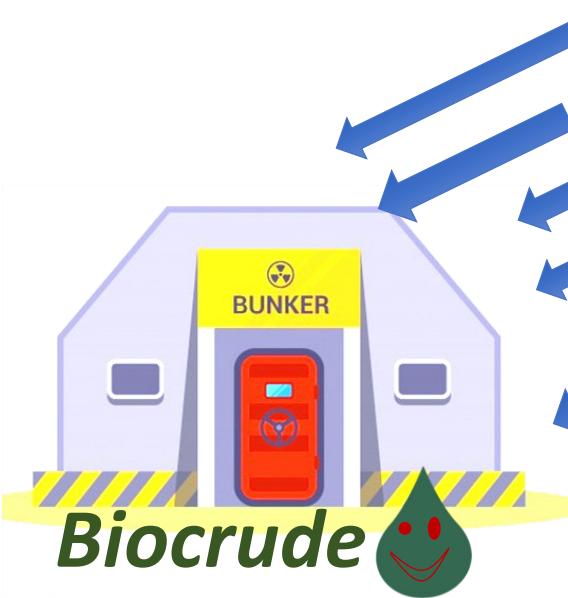
Natural Resources
Canada

Ressources naturelles
Canada

Canada

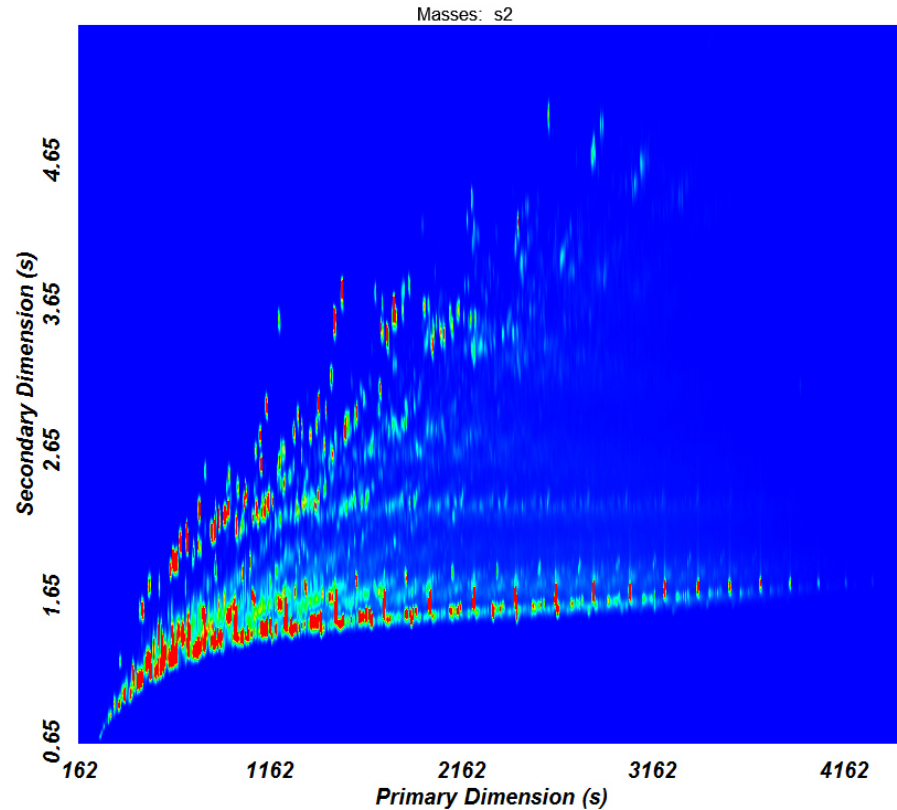
Analytical methods or tests

Analytical methods or tests that work to characterize *petroleum* oil may not work for biocrude



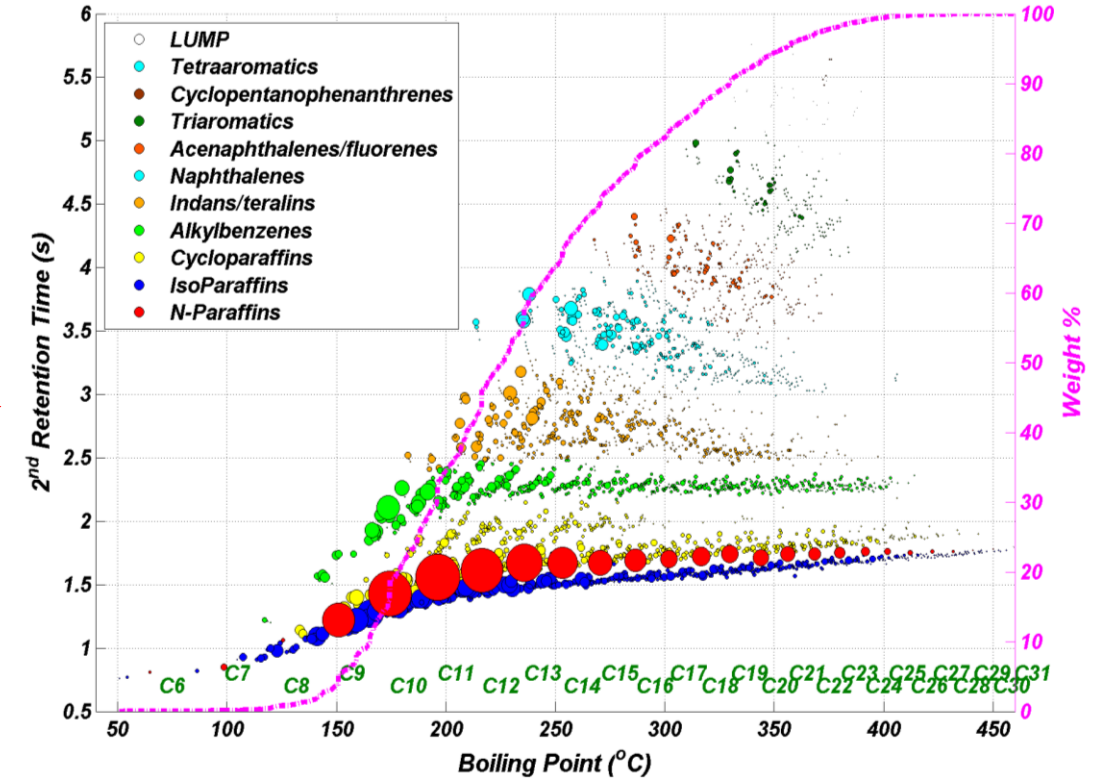
Speciation: 2D gas chromatography (GC×GC)

Structured chromatogram (petroleum-based oil)



GC × GC-FID contour map

maximizing
information

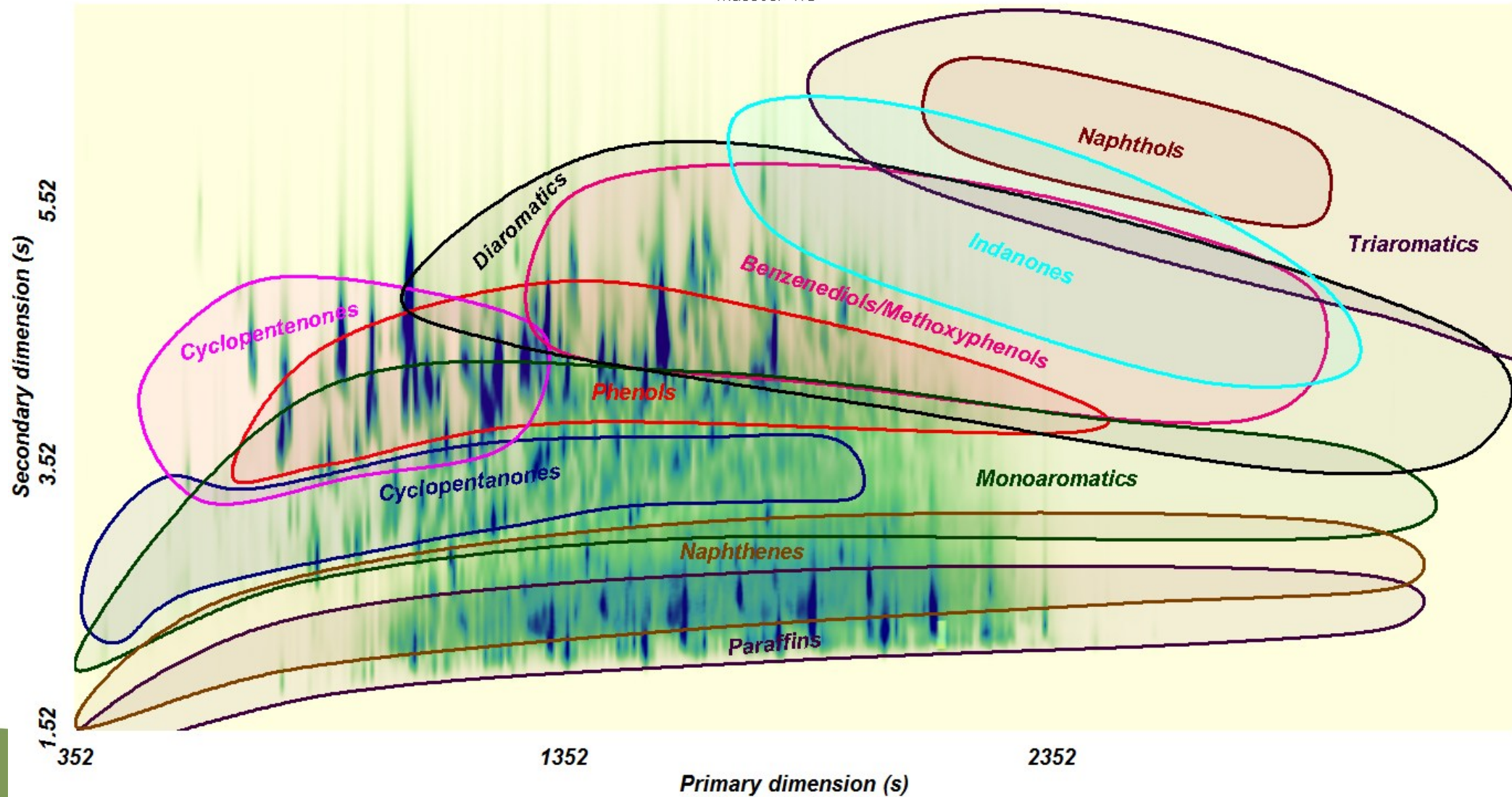


Boiling Point vs Polarity + HC classes by colours + GC × GC SimDis on the plot

Classification types - mapping

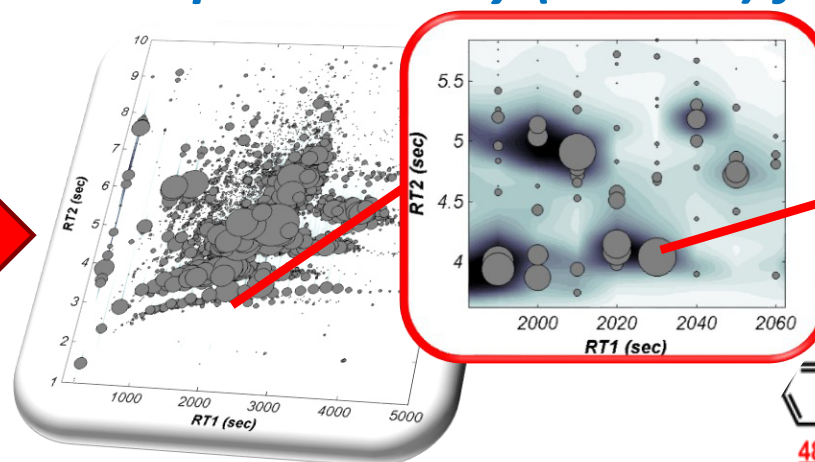
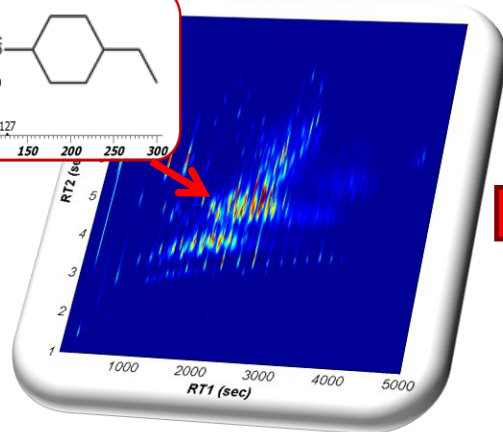
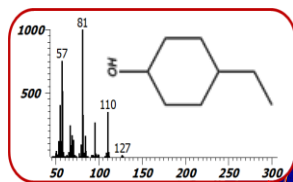
GC×GC-FID - Upgraded (lightly) biocrude sample

Masses: TIC

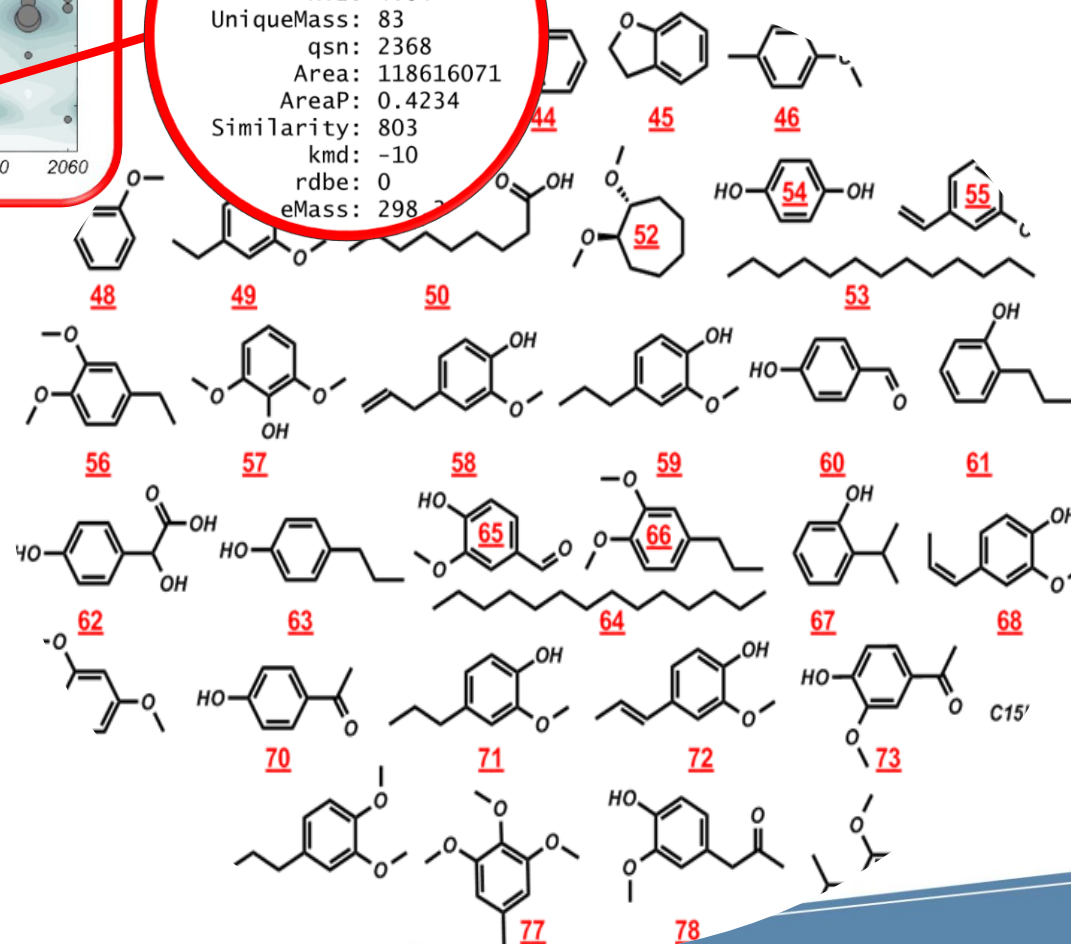


GC×GC: Biocrude

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



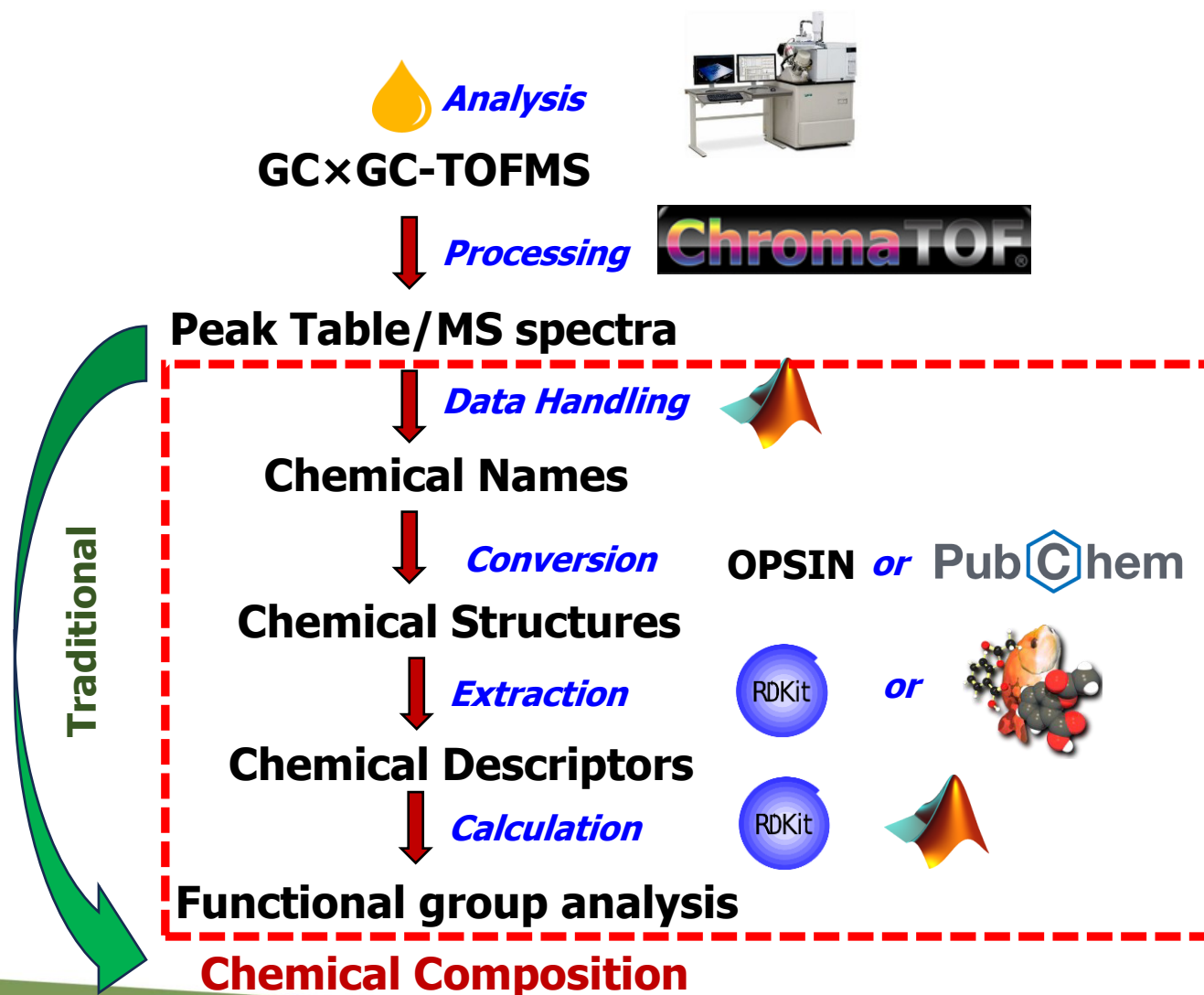
PeakNo: 1386
 Name: 1-EICOS
 Weight: 298
 MolForm: C20H42O
 RT1: 2030
 RT2: 4.04
 UniqueMass: 83
 qsn: 2368
 Area: 118616071
 AreaP: 0.4234
 Similarity: 803
 kmd: -10
 rdbe: 0
 eMass: 298



- 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.



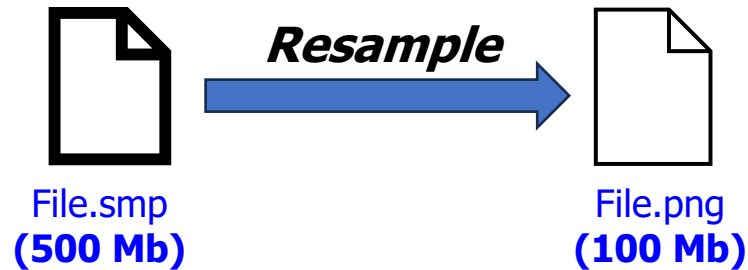
Proposed Workflow for ~~Advanced~~ ^{Alternative} Data Processing





ChromaTOF software (4.50)

Peak Table (setup)



Baseline & Peak Find & Area

Select the task or tasks you wish to perform from the list below:

- ☒ Baseline - computes baseline
- ☒ Peak Find - finds peaks above the baseline
- ☒ Library Search - identifies all peaks found
- ☒ Calculate Area / Height - computes the area and height
- ☐ Retention Index Method
- ☐ Classifications
- ☐ Apply Calibration(s) - computes the absolute concentration of peaks
- ☐ Apply Reference(s) - computes the relative concentration of peaks with respect to a reference
- ☐ Semi Quantification - computes concentration
- ☐ Tune Check
- ☐ Tailing Factor Check - checks to see if the peaks are symmetrical
- ☐ Calibration Check
- ☐ Blank Check - checks to make sure none of the peaks are from the blank
- ☒ Report - prints selected reports for each sample
- ☐ Export peak information in ASCII CSV format
- ☐ Export data in Andi MS format (.cdf)
- ☐ Export data file
- ☐ Cache script results

Enter baseline tracking info below:

#	Start	End	Mode
1*	Start of Run	End of Run	Span

Enter the baseline offset below (0.5-3.0): 0.5

Enter the number of data points that should be averaged: Auto

2nd Dimension

Match Required to combine: 650

☒ Override the allowed second dimension RT shift for combine

Early: 0.010 Late: 0.000

Enter the expected peak width in seconds below: (as measured from the chromatogram)

☐ Peak widths broaden throughout the chromatographic run

Peak Width: 0.2 Retention Time: 30

Subpeak Settings

Minimum S/N: 12

Enter the maximum number of unknown peaks to find: 100000

☐ Keep False Peaks

Enter segmented processing info below:

#	Start	End	Peak Find	S/N	Masses	Notes
1*	Start of Run	End of Run	On	100.0		2

Common masses in derivatized products:

Library Search

Library Identity Search Mode:

- ☒ Normal ☐ Quick

Library Search Mode:

- ☒ Forward ☐ Reverse

Enter the number of library hits to return: 10

Enter the masses to library search below:

☐ Enable Additional Library Search Criteria

Minimum molecular weight allow: 0

Maximum molecular weight allow: 500

Mass Threshold (Relative abundance): 10

Minimum similarity match before: 0

Add the libraries to use for search:

mainlib
replib

Number of elements' atoms in the chemical formula:

Element	<=	>	than
F	=	0	
I	=	0	
P	=	0	
Br	=	0	
Cl	=	0	

Exporting

Information not exported:

- Area %
- BaselineModified
- CAS
- Calculated Ion Ratio 1
- Calculated Ion Ratio 2
- Calculated Ion Ratio 3

Exported Information:

- Peak Number
- Name
- 1st Dimension Time (s)
- 2nd Dimension Time (s)
- UniqueMass
- Deconvolution Purity
- Quant S/N
- Area

Select directory and columns to export below:

Directory: C:\Users\giele\Work\Individuals\Project1_20

Export Library-Hits as part of the information for each peak.

☐ Remove repetition from sample name whether of library hits to export.

Header: (Leave Blank if no header information)

Functions...

Information not exported:

- Area %
- BaselineModified
- CAS
- Calculated Ion Ratio 1
- Calculated Ion Ratio 2
- Calculated Ion Ratio 3

Exported Information:

- Name
- CAS
- Formula
- Similarity
- Reverse
- Probability
- Weight
- Exact Mass
- Mass Defect

Promote Demote

Field Separator:

- ☐ Comma
- ☒ Tab

Filter:

- ☒ Calculate area percentage from filtered peak true spectrum as part of the information for each peak.
- ☒ Quantified
- ☒ Analytes ☒ Surrogates ☒ Not Found
- ☐ Match
- ☒ Out of Tolerance
- ☒ Contaminants / Unknowns
- ☐ Group:
- ☐ Expand Semi Quantifications
- ☐ Total Row

Threshold (Relative abundance of base ion (0 - 998))

not of intensity values

- ☐ Absolute
- ☒ Relative



Natural Resources
Canada

Ressources naturelles
Canada

Canada

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
	PeakNumber	Name	stDimensio...	ndDimensi...	UniqueMass	Deconvolut...	QuantSN	Area	Hit1Name	Hit1CAS	Hit1Formula Hi
	Number	Text	Number	Number	Number	Number	Number	Number	Text	Text	res5 Categorical Number
1	Peak Numb...	Name	1st Dimensi...	2nd Dimensi...	UniqueMass	Deconvolut...	Quant S/N	Area	Hit 1 Name	Hit 1 CAS	Hit 1 Formu...
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-Cvclophen...	190	9.500	47	0	7661.8	524450583	4-Cvclophen...	694-47-3	C5H8O2 558
9	1	Spectra									
10	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 40:2 36:1 2 188:2 358:2 156:2 324:2 51:2 436:2 87:2 362:2 493:									
11	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.Hit1 =
    name: {N x 1 cell}
    cas: {N x 1 cell}
    formula: {N x 1 cell}
```

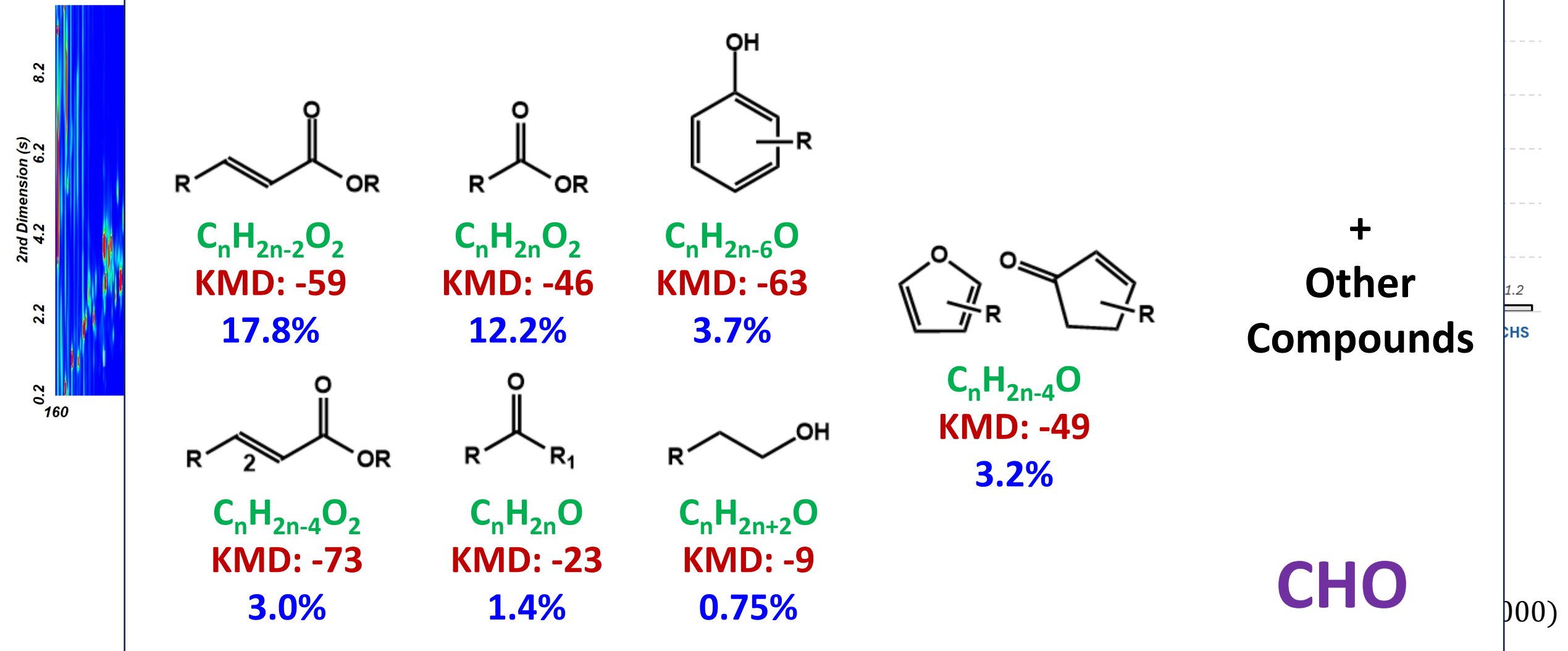
```
similarity: [N x 1 double]
reverse: [N x 1 double]
probability: [N x 1 double]
weight: [N x 1 double]
exactmass: [N x 1 double]
massdefect: [N x 1 double]
formulapeel: {N x 1 cell}
mass: [N x 1 double]
kmd: [N x 1 double]
dbe: [N x 1 double]
iabund: [N x 9 double]
chnossi: [N x 6 double]
z: [N x 1 double]
formulageneral: {N x 1 cell}
```

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

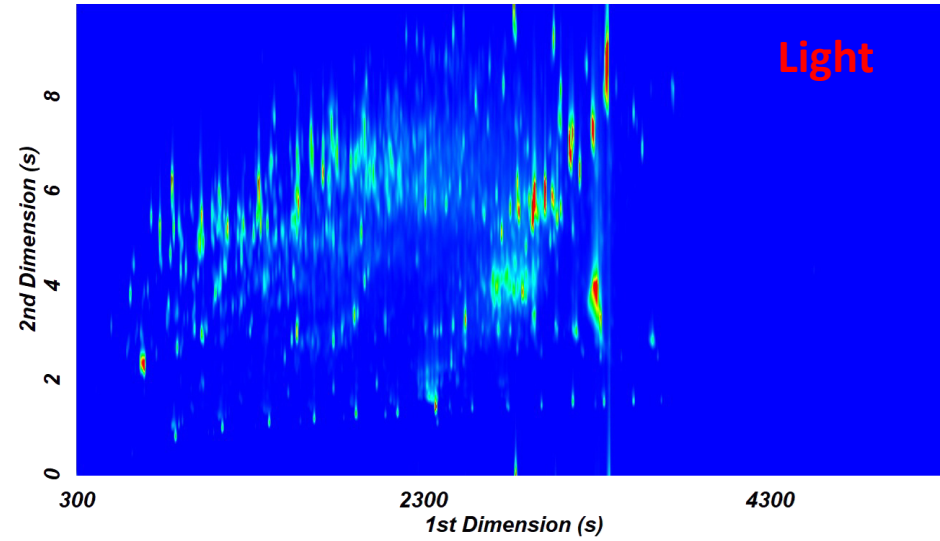
```
fTOFMS.MS =
    peaknumber: [N x 1 double]
    spectra: {N x 1 cell}
```



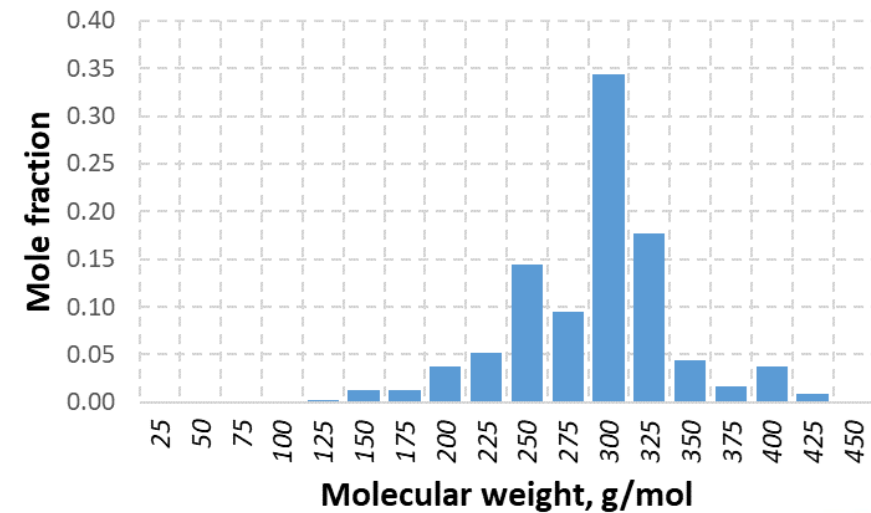
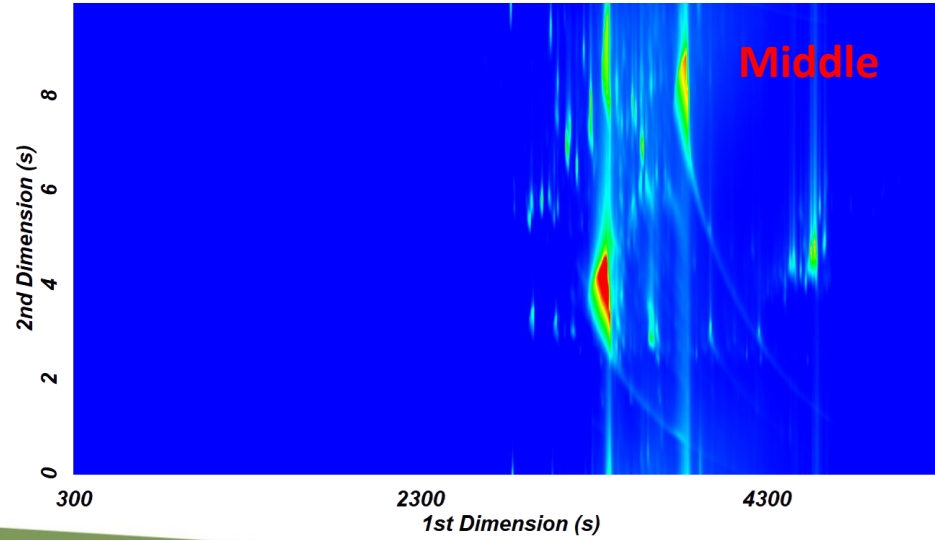
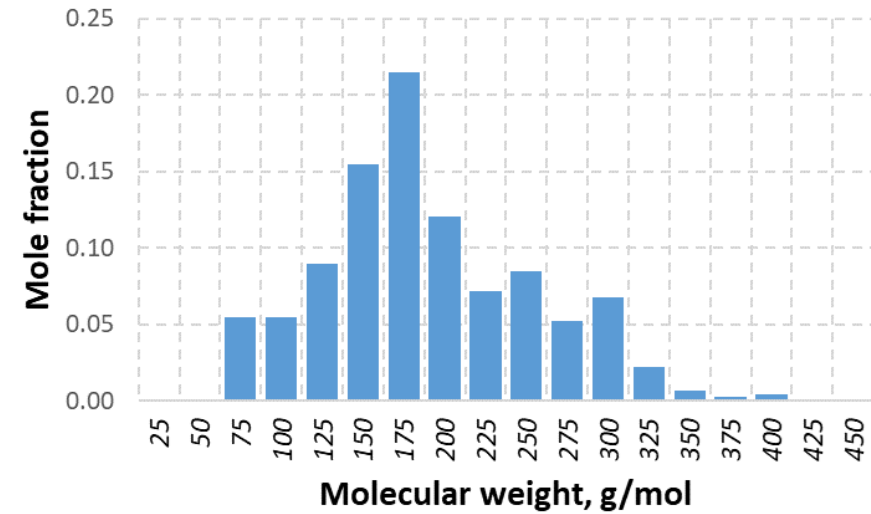
Example: Algae biocrude



Example: HTL biocrude fractions



Molecular Weight Distribution

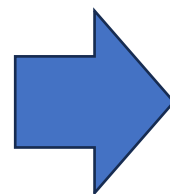


Extracting Every Bit of Information

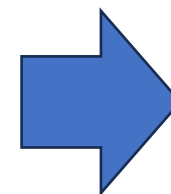


Name	CAS
p-Cresol	106-44-5
Tetradecane	629-59-4
Toluene	108-88-3
Naphthalene	91-20-3
Phenol, 2,4,6-trimethyl-	527-60-6
Phenanthrene	85-01-8
Phenol, 2,6-dimethyl-	576-26-1
Phenol, 4-ethyl-2-methoxy-	2785-89-9
Benzene, pentyl-	538-68-1
Toluene	108-88-3
Phenol, 2,3,6-trimethyl-	2416-94-6
1,2-Benzenediol,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

How to make
use of that
information?



A database of
chemical
structures can
be created



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

SMILES 😊

*Simplified Molecular
Input Line Entry System*

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

...

Chemical Names

CAS #

(Chemical Abstracts Service
Registry #)



Natural Resources
Canada

Ressources naturelles
Canada

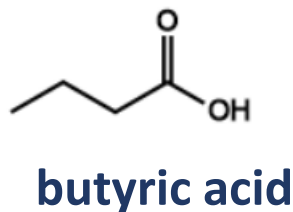
Canada

Name 2 Structures

SMILES (Simplified Molecular Input Line Entry System)



CCCCC



CCCC(O)=O



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)



Natural Resources
Canada

Ressources naturelles
Canada

Canada

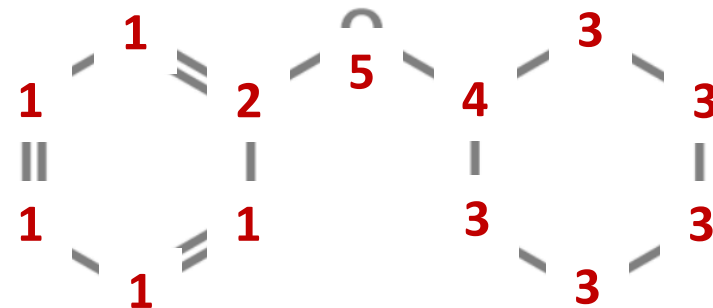
Property Calculations

Group Contribution Method (GCM)

- GCM estimates molecular properties by summing atomic contributions
- Quick, computationally efficient for large datasets
- Accuracy depends on the quality of the data

Recipe:

1. Decompose molecule into fragments
2. Assign pre-defined group numbers to each fragment
3. Add group contributions



1: [cX3H1;R]

2: [cX3H0;R]

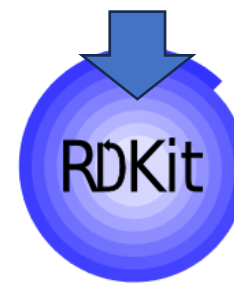
3: [CX4H2;R]

4: [CX4H1;R]

5: [OX2H0;!R;\$([OX2H0]-[cX3H0]);\$([OX2H0]-[C])]

SMARTS 
SMILES Arbitrary Target Specification

Substructure Searching



Open-Source Cheminformatics
and Machine Learning

534.2 K
982 cm³/g

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

$$D = 176.25/171.51 = 1.028 \text{ cm}^3/\text{g}$$

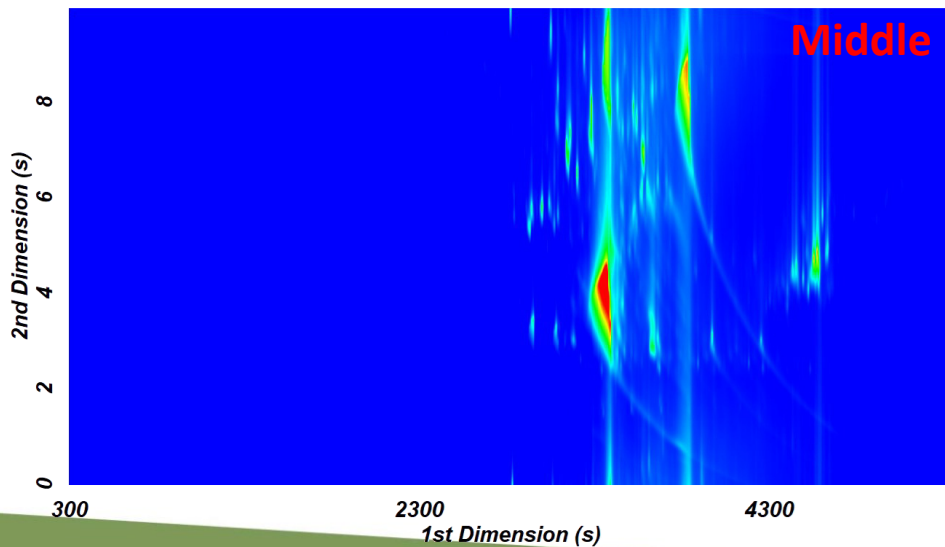
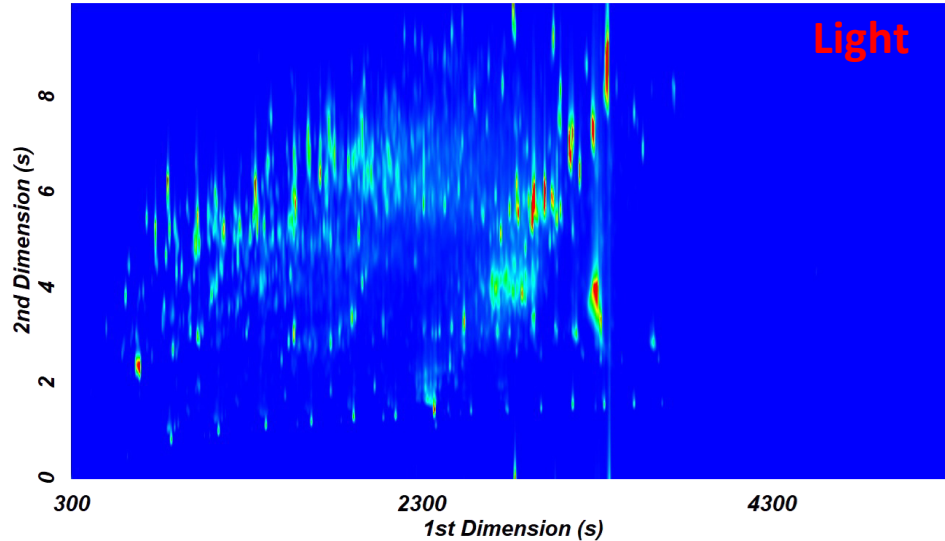


Natural Resources
Canada

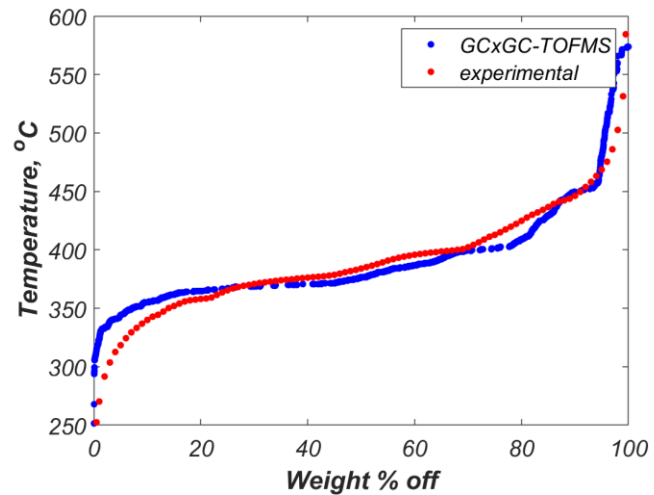
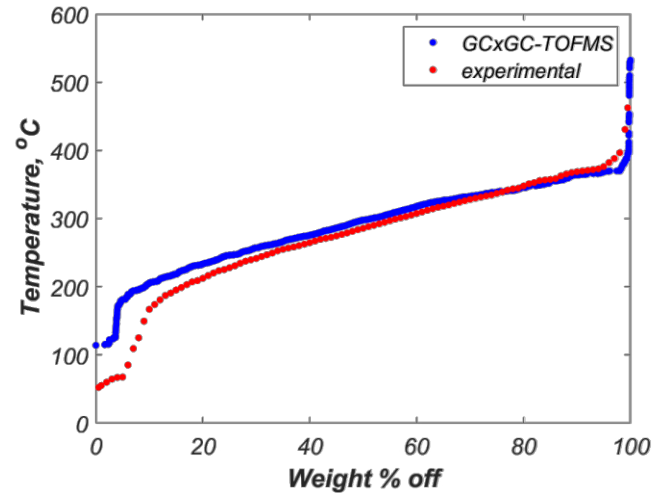
Ressources naturelles
Canada

Canada

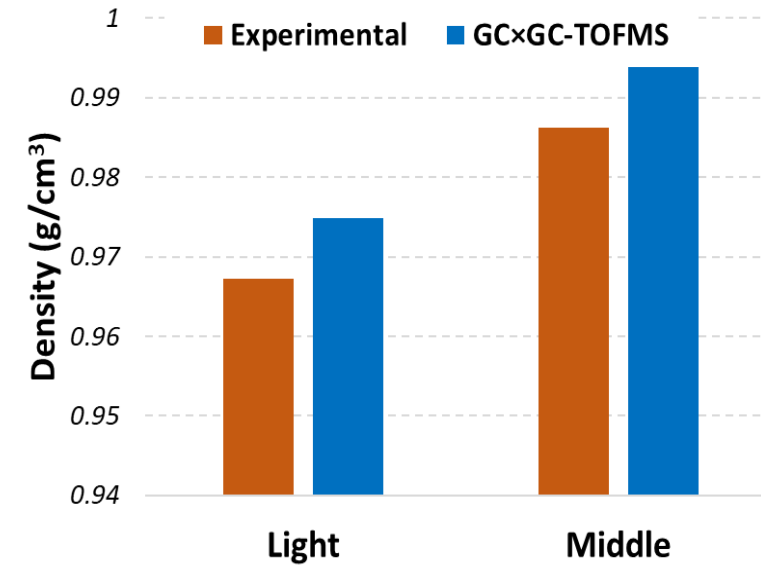
Example: HTL biocrude fractions



Simulated Distillation

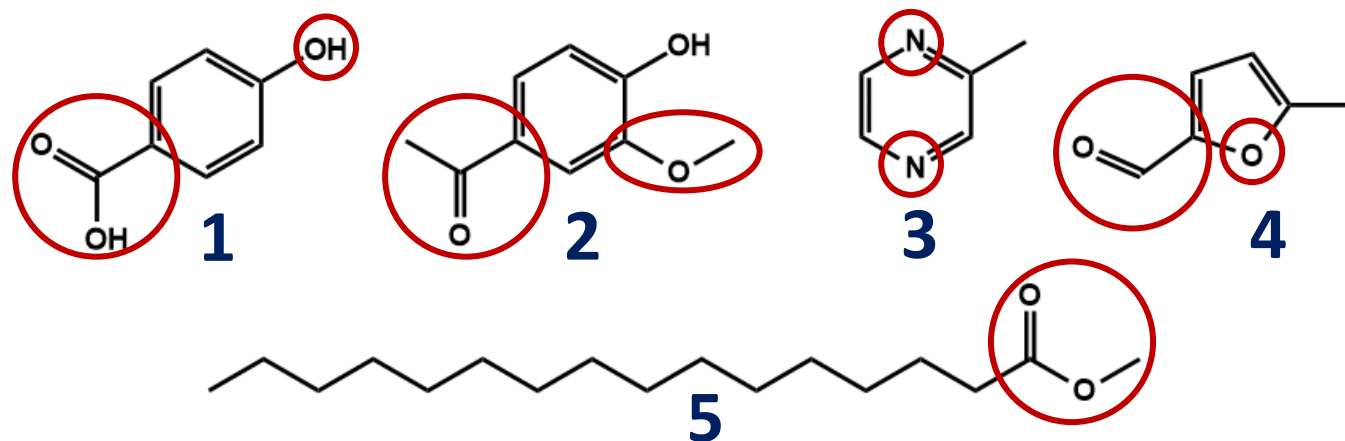


Density (15.6 °C, g/cm³)



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

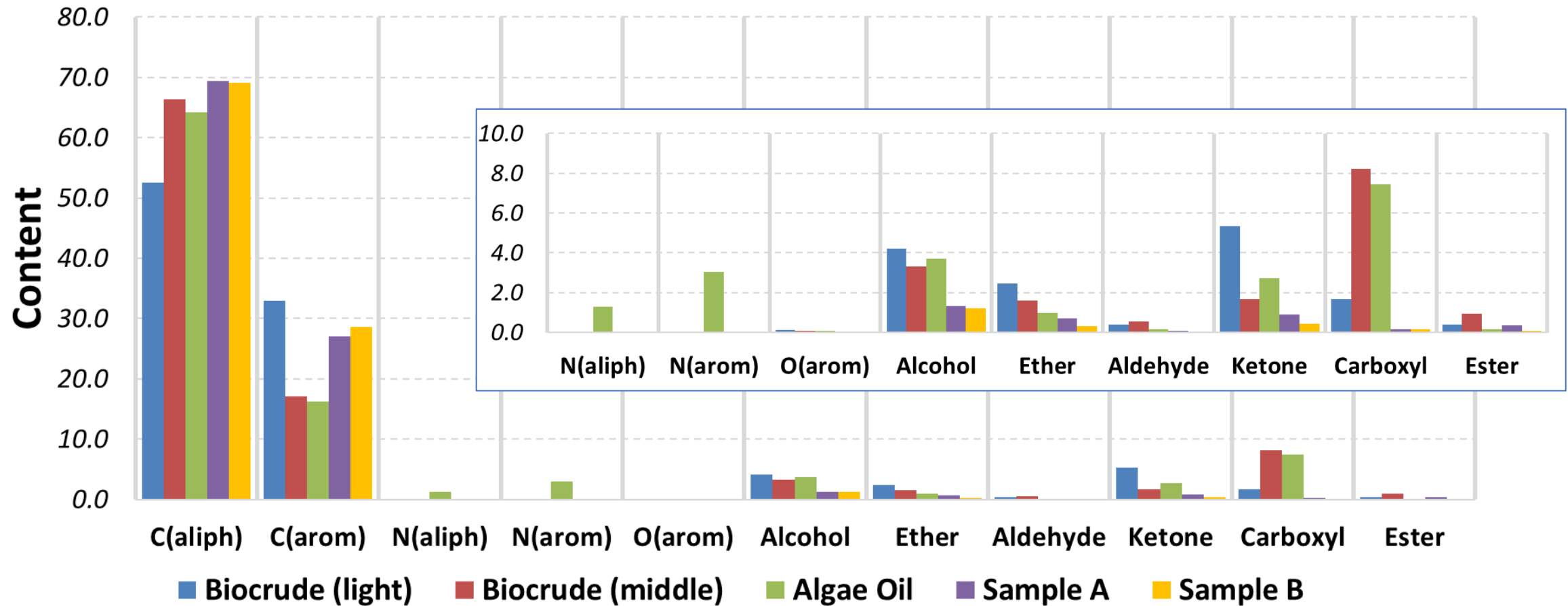
C_i - Concentration of a compound

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



Examples



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

© His Majesty the King in Right of Canada, as represented by the Minister of Natural Resources, 2025



Natural Resources
Canada

Ressources naturelles
Canada

Canada