



HAL
open science

Coding and tracking contaminants of food contact materials by wavelet analysis

Julien Kermorvant, Phuong-Mai Nguyen, Cedric Lyathaud, Sandra Domenek, Olivier Vitrac

► **To cite this version:**

Julien Kermorvant, Phuong-Mai Nguyen, Cedric Lyathaud, Sandra Domenek, Olivier Vitrac. Coding and tracking contaminants of food contact materials by wavelet analysis. FoodSim 2022, Apr 2022, Ghent, Belgium. hal-03759593

HAL Id: hal-03759593

<https://hal.science/hal-03759593>

Submitted on 30 Aug 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

CODING AND TRACKING CONTAMINANTS OF FOOD CONTACT MATERIALS BY WAVELET ANALYSIS

Julien Kermorvant^{1,2}, Phuong-Mai Nguyen¹, Cédric Lythaud¹, Sandra Domenek², Olivier Vitrac²

¹Laboratoire National de métrologie et d'Essai
29 Avenue Roger Hennequin
78190 Trappes, France

²UMR 0782 SayFood, INRAE, AgroParisTech
Université Paris-Saclay
91300 Massy, France

E-mail: julien.kermorvant@lne.fr

KEYWORDS

Food packaging, Food safety, Fingerprints, Clustering, Circular, Recycling

ABSTRACT

The management and assessment of chemicals in food contact materials (FCM) are mandatory according to the European framework regulation (EC) 1935/2004. Recent crises (bisphenol A, printing inks, mineral oils) confirm emerging risks, which have been insufficiently anticipated. The generalization of recycled materials with plastic or cellulosic origin for food contact raises concerns about their sources and decontamination levels. Leachable content of FCMs is conventionally recovered by extraction and subsequent analysis by GC-MS, LC-MS, and ¹H-NMR. Their interpretation is particularly complex for non-intentionally added substances (NIAS), such as by- and cross-contamination products. This study proposes a new processing technique to tackle the vast information acquired during non-targeted analyses, including the forest of peaks. The method has been tested on recycled plastic extracts analyzed by GC-MS. Continuous wavelet transform (CWT) tags ionized fragments and encodes each ion signal in monotonic sub-signals at different scales. The entire chromatogram is translated into text sequences, which can be aligned and compared across samples and databases. Contaminants are easily detected as insertions or substitutions in sequences. A back-transformation is proposed to reconstruct the likely signal associated with the changes and offer compatibility with conventional tools in chemometrics.

INTRODUCTION

Food packaging is essential to protect, preserve and distribute processed foods and beverages from producers to consumers. However, plastics, papers, and boards have been involved in various crises (bisphenols from can coatings, printing inks and mineral oils from paper and board, phthalates from plastics), eroding consumers' confidence in the safety of food contact materials (FCM). New environmental laws around Europe, including the single-use plastic directive (EU) 2019/904 and the French anti-waste law N°2020-105, make recycled materials mandatory for food contact. These new trends tend to oppose the principles of FCM inertia and the requirement not to endanger human health, including the identification of genotoxic substances and endocrine disruptors. Several safeguards exist, but they are either too general – obligation to assess and manage the risks of the

framework regulation (EC) 1935/2004 – or too specific – validation of the mechanical recycling processes of plastics according to the regulation (EC) 282/2008. On a day-to-day basis, recycled material and non-intentionally added substances (NIAS) originating from multiple sources (post-consumer contaminations, bad practices, material aging) are difficult to trace throughout the supply chain and for all packaging components. This constraint reduces the possibility of deploying preventive and safe, and responsible packaging design approaches (Nguyen, 2013; Zhu, 2019).

The contaminants are usually identified in solvent extracts of FCM in GC-MS. With recycled materials, the chief difficulty is the prevalent presence of complex peaks forest (Grob, 2014). Tracking specific chemicals and sourcing becomes particularly challenging in the circular economy. This study explores new procedures to align chromatograms composed of peaks and peaks forests, while being back-compatible with previous deconvolution methods (Gillet, 2011; Kassouf, 2013; Nguyen, 2015). The alignment is carried out on the continuous wavelet transform (CWT) with Ricker wavelet of each ion signal. CWT is commonly implemented in software packages (eg. XCMS) for peak detection (Du, 2006). New capabilities are demonstrated by encoding monotonic regions as letter sequences interpretable with genomic-proteomic alignment tools. Differences between samples can be then expressed as Levenshtein distances.

The manuscript is organized as follows. The second section introduces the principle of wavelet analysis of chromatograms. The aim is not only to eliminate biases (baselines, pollution due to column bleeding) but also to tag the monotonic segments of each ionic signal. The mathematical properties of the chosen time-frequency analysis are briefly outlined, as well as the possibility to build an inverse approach to reconstruct a filtered signal usable along with the conventional GS-MS identification databases. The cases studied are presented in the following section. They cover synthetic signals, reconstructed blends with varying fractions of oligomers, and recycled polyethylene terephthalate (rPET) samples from various recycling streams. The results demonstrate the different possibilities offered by the new coding method for the isolation of NIAS or the construction of fingerprints allowing a filiation of the streams. Implementation and computational cost considerations are also discussed. The main conclusions and perspectives for constructing an observatory of recycled material are finally addressed in the last section.

THEORY

Motivation

Blind deconvolution methods offer an approach to isolating signals and identifying unknown compounds. By contrast with pursuit methods which compares the remaining signal with a dictionary of sub-signals, they require no or a minimal amount of information a priori. Among them, independent component analysis is the most popular. It factorizes signals by minimizing the mutual information (mMI) between sub-signals or by maximizing the non-Gaussianity of sub-signal values (Common, 1994; Hyvarinen, 2000). However, both methods suffer severe limitations to directly tackle the peak forest problem. They cannot identify the actual number of compounds or sub-signals, order signals according to their importance or scales. Weak signals and infrequent signals are likely to be considered as noise. Enough realizations of the mixtures with sufficiently different weights must be available to expect the mixing matrix to be invertible. In the context of FCM, NIAS and cross-contaminants appear randomly, and only oligomers and additives are likely to be met across different samples. Besides, column bleeding at high temperatures pollutes signals with degradation products of the column.

This study explores a different route but with mathematical similarities with mMI and projection pursuit techniques to identify minor compounds and factorize peak forests. Our initial idea was motivated by the analogy between genetic mutations and the contamination of post-consumer FCM. Mutations are identified by the alteration of the original sequence, by insertions or deletions in the original code. In addition, it is usually thought that only mutations, which preserve functions and structures globally, propagate through generations. Similarly, NIAS in FCM are thought not to be capable of altering the intrinsic signature of the host polymer or cellulosic material. As a result, the non-mutual information between streams, recycled and virgin materials codes for the desired chemical information. This information can be addressed directly as mass spectrograms (*i.e.*, ion signals code for actual chemicals) or processed subsequently with existing ICA approaches. In the framework of information theory, the randomness of the contamination and degradation of polymers ensures that the conditional entropy reflects an endogenous (aging) or exogenous mass transfer (cross-contamination). Only the mutual information measure (the nature of the material) is symmetric, and any distance based on conditional entropy requires a quasi-symmetric construction (to satisfy the triangle inequality) or the identification of neat/virgin references *a priori*. For intended use by a national reference laboratory, no degree of belief about the cause of the contamination was introduced in this study. The source of the contamination or the ancestor of the material before recycling is assumed unknown.

Regularizing properties of continuous wavelet transforms

Continuous wavelet analysis generalizes harmonic analysis while preserving important features. It provides an overcomplete representation of signals with filtering and gain flattening capabilities. In this work, the best results were obtained with the Mexican hat or Ricker's function as mother wavelet $\psi(t)$. The continuous wavelet transformation (CWT) reads (Bernstein, 2019):

$$C(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} s(t) \psi\left(\frac{t-b}{a}\right) dt \quad (1)$$

$$\psi(t) = \frac{2}{\sqrt{3}\pi^{\frac{1}{4}}} (1-t^2) e^{-\frac{t^2}{2}} \quad (2)$$

$\psi(t)$ is the normalized Laplacian of Gaussian function and is plotted in Figure 1.

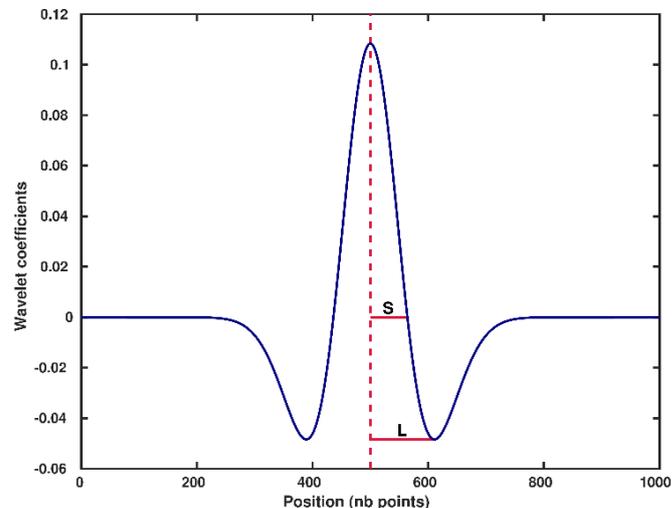


Figure 1 Ricker wavelet with scale $S = 64$

It includes the three lobes: one with positive surface area and two with negative ones. Due to the negative values, $\psi(t)$ is not a kernel and has no probabilistic interpretation. Positive lobes detect peaks with Gaussian shapes and matching the scale S . The two negative lobes enable to cancel any linear trend or stationary Poisson noise. The main CWT properties are listed in Table 1.

Table 1 Ricker CWT properties

Property	Consequence
Continuous wavelet transform is a bilinear application	Continuous wavelet transform of a sum of signals is equal to the sum of each transformed signal
Integral of the Ricker wavelet is equal to zero	Continuous wavelet transform of a linear signal is the zero signal at any scale
Integral of the squared Ricker wavelet is equal to one	Variance of an infinite signal s is equal to the variance of its continuous wavelet transform at any scale

The filtering capabilities are exemplified in Figure 2 on synthetic signals with baselines, positive and negative ramps. The signals are unaffected by the presence of baselines. CWT tags all peaks with negative lobes, coding for start and stop sequences. Choosing proper scales enables to filter out noise while keeping acceptable separation of peaks. Shouldering can be extracted but usually with a poorer quantifiability.

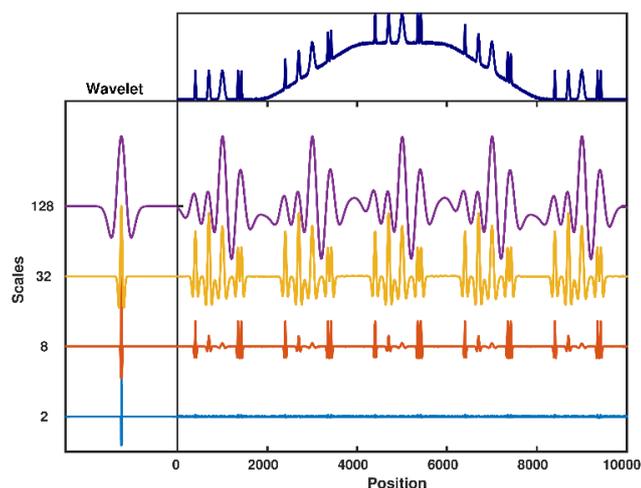


Figure 2 CWT of a synthetic signal (Top) at scales $S = 2, 8, 32, 128$ (From bottom to top) with corresponding dilated mother wavelet shown in the first column

MATERIALS AND METHODS

Materials

The methodology was applied to synthetic signals and polymer extracts. One-dimensional synthetic ion signals were obtained from the linear superposition of baseline, noise and Gaussian peaks with prescribed widths and positions.

Two types of polymer extracts were studied to represent the two typical applications of the methodology: the identification of additives in polyolefins and the comparison of recycled PET. Tailored extracts of low-density polyethylene (LDPE) with increasing complexity were prepared by mixing twenty common additives of the plastics positive list with different levels of oligomers of polyethylene. Three oligomeric contamination levels were prescribed: 0/100, 50/50 and 100/0 PE extract/dichloromethane (v:v). The studied set of PET extracts in chloroform/methanol 25/75 (v:v) mixture included two commercial food-grade materials, one unknown material, samples (flakes) from ten different streams of post-consumer packaging. One-half of studied streams had a food origin, whereas the other had a mixed-origin with less than 5% of non-food-origin.

GC-MS parameters

All samples were analyzed in GC-MS using standards protocols of LNE. Samples were analyzed in gas-chromatography coupled with a single quadrupole mass spectrometer. A $30 \times 0.25 \times 0.25$ DB-5MS non-polar column was used to separate compounds. Injector temperature was 250°C , and oven temperature gradient started at 50°C for 4 min, followed by an increase of $15^\circ\text{C}/\text{min}$ for 18 min up to 320°C and held for 12 min for a total run of 34 min. Data were acquired in scan mode from 40 Da to 650 Da with a mass resolution of 1 Da and an electronic ionization of 70 eV.

Wavelet transform and encoding

All ion signals were processed independently and transformed via CWT at the desired scales. Transformed signals were encoded as monotonic segments. The six possible letters are shown in Table 2. While peak lines are not overlapping and the Shannon rule is fulfilled, the obtained coding is almost invariant with scale. All combinations of

peaks can be encoded as increasing/decreasing, positive/negative start, crossing x-axis, or not segments. The information is strongly compressed, and any pattern can be searched in real-time using regular expressions. Accepting insertions and deletions in the sequence makes the search method resistant to contaminations and noise. This approach is based on an earlier methodology proposed by Du (2006) with the inclusion of signal encoding and regular expressions for local maxima detection.

Table 2 Encoding system

Letter	A	B	C	Z	Y	X
Crossing X	1	0	0	1	0	0
Positive start	0	1	0	1	0	1
Increasing	1	1	1	0	0	0

Chromatogram reconstruction

Because CWT offers a redundant representation of signals, there is no unique manner to reconstruct individual peaks in the physical space for chemical identification and quantitative interpretation. The two-step strategy was applied independently to each peak. The first step consisted of detecting the optimal scale maximizing the cross-correlation with the wavelet. The second step relied on prior calibration of dimensionless Gaussian peaks with their transforms. Scaling was recovered by the height and width of each peak. The procedure reconstructs chromatograms fully without a baseline, with strong and weak peaks assuming symmetry. Asymmetric patterns were recovered. The outputs are compatible with existing software and databases.

Fingerprints and pairwise alignment of sequences

The 611 text sequences coding for the 611 ion lines represent the fingerprint of each material. Sequences originating from the same materials can be aligned the same way as biological sequences (protein, DNA, RNA), natural language, and financial data are. Sequence similarities represent the stable contribution of the polymer or its formulation. Alignments are obtained by adding gaps to the two query sequences, as shown in Figure 3. The dissimilarities between strings were measured by the normalized pair Levenshtein edit distance, as justified by (Berger, 2021). The algorithm detailed in (Barton, 1993) was used along with a custom identity scoring matrix.

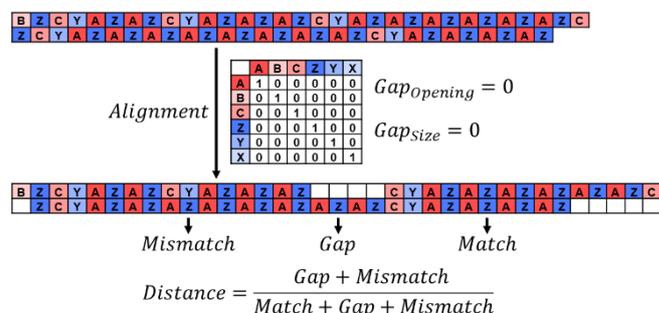


Figure 3 Pairwise alignment and distance criterion, raw (Top) and aligned sequences (Bottom)

The distance counts insertion, deletion, and substitution as gaps or mismatches indifferently. Since the distance is symmetric, there is no causal/parenting linkage between samples. For the comparison of multiple samples, the chosen distance satisfies the axioms of a metric, at least when the

number of matches is homogeneous across samples. If not, absolute distances were preferred.

RESULTS AND DISCUSSION

Results are presented on three datasets: synthetic signals, reconstructed extracts from polyethylene and extracts from various recycled PET streams.

Illustration of the encoding monotony stability

CWT of chromatograms at multiple scales transforms 2D datasets into 3D datasets. Continuous scale representation introduces redundancy, which is not desirable for most applications. Increasing scales improves the signal-to-noise ratio, but too large scales cause peaks to merge. A robust approach requires the conclusions independent of the scale choice, at least within an acceptable range.

The encoding stability is illustrated in Figure 4 for six Gaussian peaks with different widths and proximities. The corresponding monotonic segment coding is shown for length scales ranging from 1 to 128 times the sampling period. Most of the transformed signal is made of noise at low scales, and changing the CWT scale slightly modifies the coded monotonic sequence. For scales commensurable to peak width, coding is stable at least while the scale is not greater than the distance with two peaks. Interactions between peaks shift the location but preserve the distance. Local peak merging induces only local changes that affect the entire sequence progressively. The stability of the coding over a broad range of scales appears as preserved color orders. The edit distance between reconstructions at different scales offers a straightforward methodology to estimate upper and lower critical scales on real-life samples.

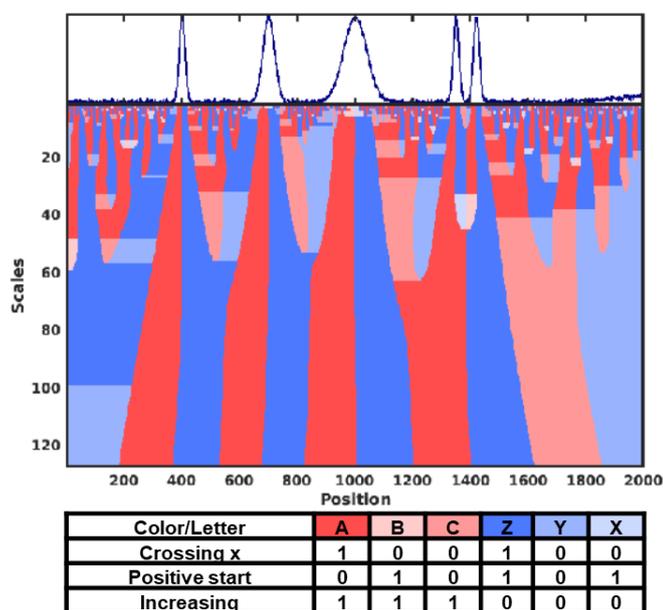


Figure 4 Evolution of encoded sequences with wavelet scale from 1 to 128 (Bottom) for a synthetic signal (Top)

Detection and reconstruction of peaks forests in polyethylene extracts

Polyolefins constitute a significant source of leachable substances, including and additives and polyolefins. Additives have independent signatures, but oligomers and their degradation products coelute with additives. High-

temperature ramps used to elute bulky additives generate several background ions due to the aging of the column coating. Baselines and background pollution are usually removed by considering only ion signals, which are likely to be correlated together. These corrections require strong signals and sufficient statistics (Likic, 2009) and are likely to confuse the forest of peaks with column bleeding. A recent study showed that ion backgrounds reduce the transmission rate to mass detectors and affect their sensitivity (Schulze, 2020).

This work does not introduce any other hypotheses that peaks at a given scale are symmetric. The method's capability is shown in Figure 5 on samples with substantial column bleeding and three levels of peak forests produced by oligomers. Background reduction was very efficient while keeping minor ion signals and peak forest. CWT filtered out stationary mass signals ion-wise without prior chemical information or correlation. Reconstructed chromatograms included 300 to 450 ions, corresponding to approximately 18 to more than 100 compounds. Multiresolution analysis was required to achieve sharp reconstructions. In practice, 5 to 11 scales are required in routine. The complete analysis is performed in 2 to 3min per chromatogram.

Mass spectra and elution-times databases can be directly used to identify compounds or suggest chemical structures. Aliphatic ions have a typical signature in polyolefins due to the absence of heavy atoms besides carbons. It was confirmed that the 20 considered additives were still identifiable despite the forest of peaks.

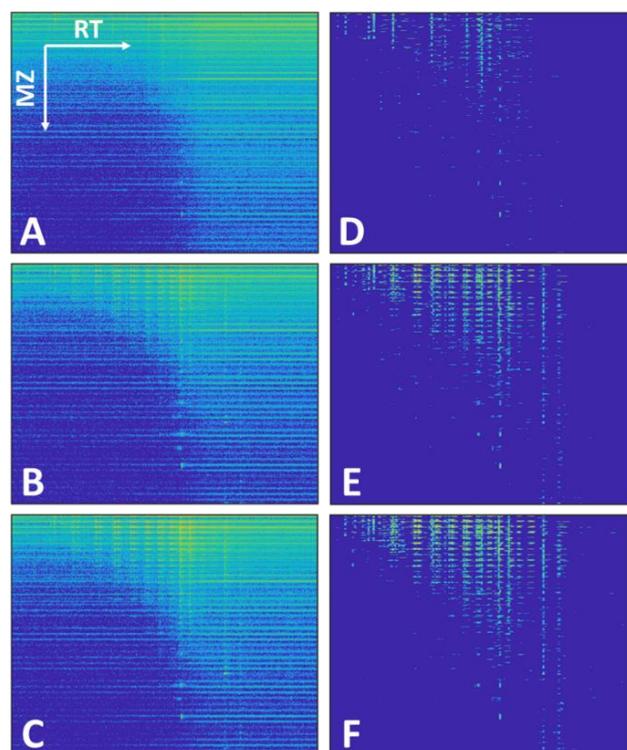


Figure 5 (a-c) raw and (d-f) reconstructed chromatograms with MZ the mass over charge ratio and RT the retention time: (a, d) no oligomers, (b, e) 50 v% and (c, f) 100 v% oligomers of polyethylene

Clustering recycled PET streams

Fingerprints of PET present many substrings (subsequences) which facilitate their alignment. At the scale $S = 32$, the fingerprint consists of 611 strings of length ranging from 200 to 300. The distances between the 13 samples are associated with a minimum of 20% to a maximum of 55% of gaps and mismatches in the aligned sequences.

The pair correlation matrix along each coordinate was used to generate a hierarchical clustering of samples from a $k - d$ tree approach. A simplified representation was finally achieved using representation was made using principal coordinates analysis (PCoA) to reduce the number of dimensions from 611 to 3. The results are shown in Figure 6. The distances between samples are only related to their differences. Three to four different groups were identified with a clustering matching the origin of the different streams. The visualizations help identify that recycled materials (not decontaminated) were not equivalent to food-grade ones. The differences were associated with different post-consumer contamination profiles. According to the origin of recycled materials, they were significantly different: food or non-food. After projection, the first dimension, labeled x , separated recycled and non-recycled materials, whereas the second one labeled z could discriminate the sourcing. It is worth noticing that the analysis is not based on concentrations but information and counts the same way the differences associated with a minor or a significant compound.

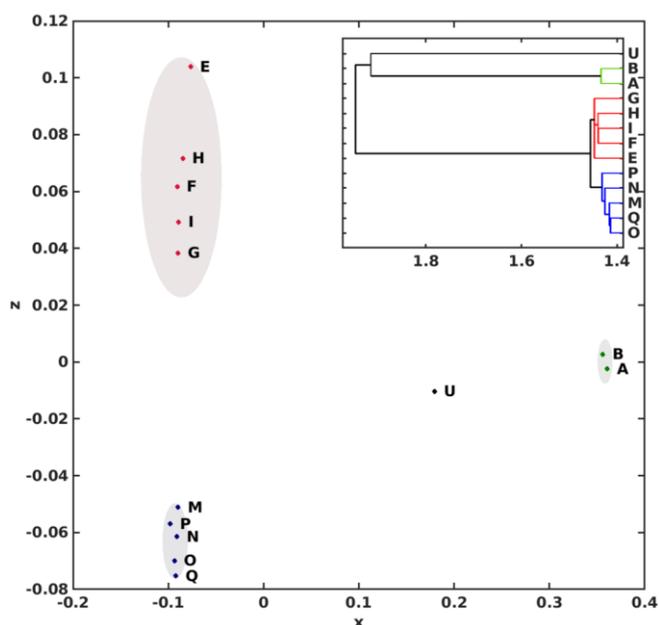


Figure 6 Clustering of rPET Samples. The samples are tagged: (A & B) for food-grade rPET, (U) for unknown rPET, (E to I) for mixed-origin and (M to Q) food-origin flakes

CONCLUSION

Until recently, the safety of food contact materials was managed mainly by certificates of conformity issued at all stages of production and processing of food contact materials. Only an inspection of the finished product was required to verify compliance with local laws. The development of the circular economy changes the problem entirely by requiring high throughput controls along the whole supply chain: from waste to recycled material streams, etc. The search for contaminants must necessarily be untargeted and extended to

all substances in our exposome. This work proposes to combine GC-MS widely used in laboratories and industry with high volume spectra processing tools: thousands of data, accumulated and stored for several years. Because the most toxic molecules are not necessarily the most concentrated, the approach used relies on information theory to isolate the signals of substances and compare them between reference samples. This study presents the feasibility and robustness of the approach in various contexts and in the presence of important biases: column bleeding, peak forests, low resolution. The coding of monotonic segments into strings after wavelet transforms presents sufficient invariance with the scale used to allow automated processing. The one scale fingerprint is sufficient to allow sample comparison, but five scales are recommended to allow the reconstruction of the optimized reconstructed signal. This construction ensures compatibility with existing substance identification tools. This method is at the heart of the construction of the first recycled material observatory.

REFERENCES

- Barton, G. J. 1993. "An Efficient Algorithm to Locate All Locally Optimal Alignments between Two Sequences Allowing For Gaps". *Computer Applications in the Biosciences* 9 (6): 729-734.
- Berger, B., Waterman, M. S., Yu, Y. W. 2021. "Levenshtein Distance, Sequence Comparison and Biological Database Search". *Ieee Transactions on Information Theory* 67 (6): 3287-3294.
- Bernstein, B. and Fernandez-Granda, C. 2019. "Deconvolution of Point Sources: A Sampling Theorem and Robustness Guarantees". *Communications on Pure and Applied Mathematics* 72 (6): 1152-1230.
- Comon, P. 1994. "Independent Component Analysis, a New Concept". *Signal Processing* 36 (3): 287-314.
- Du, P., Kibbe, W. A., Lin, S. M. 2006. "Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching". *Bioinformatics* 22 (17): 2059-2065.
- Gillet, G., Vitrac, O., Desobry, S. 2011. "A Fast Method to Assess the Composition of a Polyolefin: An Application to Compliance Testing of Food Contact Materials". *Journal of Applied Polymer Science* 119 (3): 1492-1515.
- Grob, K. 2014. "A Personal Review on 40 Years at the Kantonales Labor Zurich: Success-Failure - Conclusions". *Chimia* 68 (10): 682-691.
- Hyvarinen, A. and Oja, E. 2000. "Independent component analysis: algorithms and applications". *Neural Networks* 13 (4-5): 411-430.
- Kassouf, A., Maalouly, J., Chebib, H., Rutledge, D. N., Ducruet, V. 2013. "Chemometric tools to highlight non-intentionally added substances (NIAS) in polyethylene terephthalate (PET)". *Talanta* 115: 928-937.
- Likic, V. A. 2009. "Extraction of pure components from overlapped signals in gas chromatography-mass spectrometry (GC-MS)". *Biodata Mining* 2: 11.
- Nguyen, P. M., Goujon, A., Sauvegrain, P., Vitrac, O. 2013. "A Computer-Aided Methodology to Design Safe Food Packaging and Related Systems". *Aiche Journal* 59 (4): 1183-1212.
- Schulze, B., Bader, T., Seitz, W., Winzenbacher, R. 2020. "Column bleed in the analysis of highly polar substances: an overlooked aspect in HRMS". *Analytical and Bioanalytical Chemistry* 412 (20): 4837-4847.
- Zhu, Y., Guillemat, B., Vitrac, O. 2019. "Rational Design of Packaging: Toward Safer and Ecodesigned Food Packaging Systems". *Frontiers in Chemistry* 7: 20