Region of Interest Selection for GC×GC-MS Data using a Pseudo Fisher Ratio Moving Window with Connected Components Segmentation

Ryland T. Giebelhaus [1, 2], A. Paulina de la Mata [1, 2], and James J. Harynuk [1, 2 *]

[1] Department of Chemistry, University of Alberta, 11227 Saskatchewan Dr. NW, Edmonton, Canada

[2] The Metabolomics Innovation Centre, 11227 Saskatchewan Dr. NW, Edmonton, Canada

* Correspondence: james.harynuk@ualberta.ca

ORCID

RTG - 0000-0002-7625-3077

APDLM - 0000-0002-5133-7979

JJH - 0000-0002-9627-5206

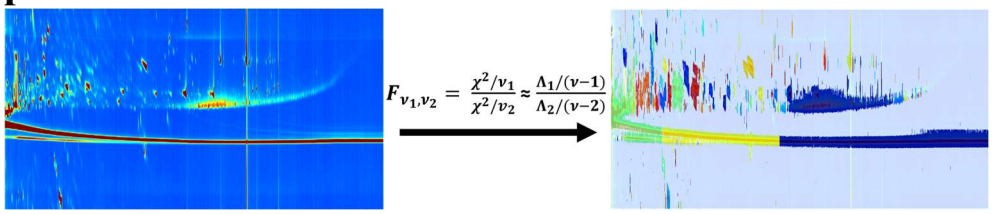
1. Abstract

Comprehensive two-dimensional gas chromatography mass spectrometry (GC×GC-MS) data present several challenges for analysis largely because chemical factors drift along the chromatographic modes across different chromatographic runs, and there is frequently a lack of reliable molecular ion measurements with which to align data across multiple samples. Tensor decomposition techniques such as Parallel Factor Analysis (PARAFAC2/PARAFAC2×N) allow analysts to deconvolve closely eluting signals for quantitative and qualitative purposes. These techniques make relatively few assumptions about chromatographic peak shapes or the relative abundance of noise and allow for highly accurate representations of the underlying chemical phenomena using well-characterized and scrutinized principles of chemometrics. However, expert intervention and supervision is required to select appropriate Regions of Interest (ROI) and numbers of chemical components present in each ROI. We previously reported an automated ROI selection algorithm for GC-MS data in Giebelhaus *et al.* where we observed the ratio of the first and second eigenvalues within a moving window across the entire chromatogram. Here, we present an extension of this work to automatically detect ROIs in GC×GC-MS chromatograms, while making no assumptions about peak shape. First we calculate the probabilities of each acquisition being in a ROI, then apply connected components segmentation to discretize the regions of interest. For sparse chromatograms we found the algorithm detected spurious peaks. To address this, we implemented an iterative ROI selection process where we autoscaled the moving window to the standard deviation of the noise from the previous iteration. Using three user-defined parameters, we generated informative ROIs on a wide range of GC×GC-TOFMS chromatograms.

2. Key words

Comprehensive two-dimensional gas chromatography; Region of interest selection; Chemometrics; Fisher Ratio Analysis

3. Graphical abstract



1 Introduction

Comprehensive two-dimensional gas chromatography mass spectrometry (GC×GC-MS) is a powerful separation technique that utilizes two columns with different stationary phase chemistries, connected by a modulation device [1]. GC×GC-MS offers more peak capacity, chromatographic resolution, and sensitivity compared to GC-MS separations, offering profound insight into the volatile and/or semi-volatile chemistry of a variety of sample types. Recently, GC×GC-MS has been utilized in metabolomics, the study of all metabolites in a biological sample [2,3], to analyze challenging biological samples including plants and bodily fluids such as urine [4–6], and in environmental studies to detect trace contaminants [7,8]. While incredibly powerful, GC×GC-MS data presents several challenges for processing and analysis. These stem largely from high sample complexity, drift of chemical factors along the chromatographic modes between chromatographic runs, and a lack of reliable molecular ion measurements to align peaks across multiple samples.

A number of approaches have been proposed for the analysis of GC×GC-MS data, many of which have been deployed as part of commercial software [9–12]. Chemometric methods such as Multivariate Curve Resolution (MCR), Independent Component Analysis (ICA), or Parallel Factor Analysis (PARAFAC/PARAFAC2) have been utilized for the decomposition of multiple GC-MS and GC×GC-MS samples [12–15]. Parallel Factor Analysis allows for analysts to deconvolve closely eluting signals for quantitative and qualitative purposes, while making few assumptions about chromatographic peak shape or the abundance of the signal, providing accurate representations of the underlying chemical phenomena. Recently, Armstrong *et al.* described a PARAFAC algorithm, dubbed PARAFAC2×N, which allows for the direct decomposition of 4^{th} order tensors, making GC×GC-MS data easily amenable to multiway decomposition [12]. While PARAFAC2×N is a powerful technique for analyzing GC×GC-MS data, it requires pre-defined regions of interest (ROIs) within each of the chromatograms, and an appropriate *k*-component number for each region. These are difficult determinations to automate with GC×GC-MS data.

Region of interest selection is a broad field in computer science and statistics, which focuses on selecting regions in data that contain useful data [15,16]. Automated ROI selection is popular in signal and image processing as it selects ROIs and provides an output based on a series of predefined metrics, eliminating bias and subjectivity that is introduced by manual ROI selection [16,17]. In Giebelhaus *et al*. we described a novel method for automated ROI selection in GC-MS data, where we used a pseudo Fisher ratio moving window to compute the probability of an acquisition containing chemical components [15]. Here, we present an extension of this work to automate ROI selection on GC×GC-MS data, requiring only three user-defined input parameters. Two ROI selection approaches using the pseudo Fisher ratio-moving window are described here; the first is a direct extension of Giebelhaus *et al*. where we compute the probabilities over each modulation then reconstruct the data to represent the GC×GC-MS data. The second approach is iterative, where the ROIs are computed then the standard deviation of the noise regions of the chromatogram is used to auto-scale the moving window in subsequent iterations. We found the iterative approach is necessary for most GC×GC-MS data due to the higher chromatographic resolution and peak capacity, resulting in more noise acquisitions than in GC-MS data, artificially inflating the auto-scaled values of the moving window in noise regions of the chromatographic plane.

2 Experimental

2.1 Data preprocessing

Development and application of the region of interest algorithm was performed in Matlab® R2022a. Each chromatogram was imported into Matlab as a data matrix, $X \in R^{M \times N}$, where M is individual acquisitions and N is ions. These data matrices were then folded into an I × J × K tensor, where I is acquisitions in the

first separation dimension, J is acquisitions in the second dimension, and K is individual ions. Each *i*, *j*, *k* entry describes the relative abundance of a particular molecular fragment at a particular moment in time along both dimensions of separation.

2.2 Region of interest selection algorithm

2.2.1 Fisher ratio moving window

Our region of interest algorithm for GC-MS data described in Giebelhaus *et al.* was adapted to detect regions of interest in comprehensive two-dimensional gas chromatography mass spectrometry [15]. In brief, the pseudo F-ratio moving window for GC-MS data works as follows: data contained within the moving window with size *wndw* is auto-scaled, and then the first two singular values are computed. An approximation of a Fisher ratio is then computed as the ratio of the squares of the first two singular values. This process is then repeated in each iteration of the algorithm as the window is shifted along the raw data by one spectrum in each iteration. A vector of pseudo F-ratios is generated, which is then used to generate a vector of probabilities that each spectrum is within a ROI based on the *F* cumulative distribution function with *wndw*-1 numerator degrees of freedom and *wndw*-2 denominator degrees of freedom. For each value in the vector of probabilities, the χ^2 was calculated using the χ^2 inverse cumulative distribution function, with one degree of freedom, and stored as a vector. This function is computationally intensive, and is called from a MEX file reduce computational time. Vector addition was then performed, combining this vector with the previously summed χ^2 values, with indexing adjusted to account for the moving window. The probability for each mass spectrum was then calculated from the χ^2 values using the χ^2 cumulative distribution function. The number of degrees of freedom for this calculation was the number of instances where the acquisition was considered by the moving window; for instance, the first mass spectrum has one degree of freedom, while a mass spectrum in the middle of a sufficiently large chromatogram (where $n > 2 \times wndw$) will have *wndw* degrees of freedom.

2.2.2 GC×GC-MS region of interest selection

Two different algorithm modes for GC×GC ROI selection are reported here (Figure 1). The first is a direct extension of the GC-MS algorithm reported in Giebelhaus *et al*. [15], where probabilities are computed for each modulation then reconstructed into an I × J matrix of probabilities. The second mode of ROI selection uses an iterative approach to calculate ROIs. In this mode, an initial set of ROIs are computed across the chromatogram with a *wndw* size of *wndw*×2, with the window auto-scaled to the standard deviation of the entire chromatogram rather than auto-scaling to the standard deviation of the spectra in the moving window. Once these initial ROIs are computed, the noise regions are extracted by selecting acquisitions below the user input probability threshold cut-off, and the standard deviation of all noise is computed. The ROI calculation is then performed again, auto-scaling to the standard deviation of the noise from the first ROI calculation. This process is then repeated as many times as desired by the user.

2.2.3 Connected components segmentation

To discretize the ROIs within the chromatographic plane, connected components segmentation was utilized. First, the I × J matrix of probabilities was binarized. The bwconncomp() Matlab function was then used to calculate the individual connected components. Finally, the labelmatrix() function was then applied to create a label for each unique connected component, assigning each component a number starting from 1. The connected component labels are then used to extract each individual ROI from the chromatogram as a tensor. A figure can be generated with the ROIs labeled as individual colors to represent their discretization. Some ROIs, which are not connected share the same color, which is a result of the limited color palette of Matlab.

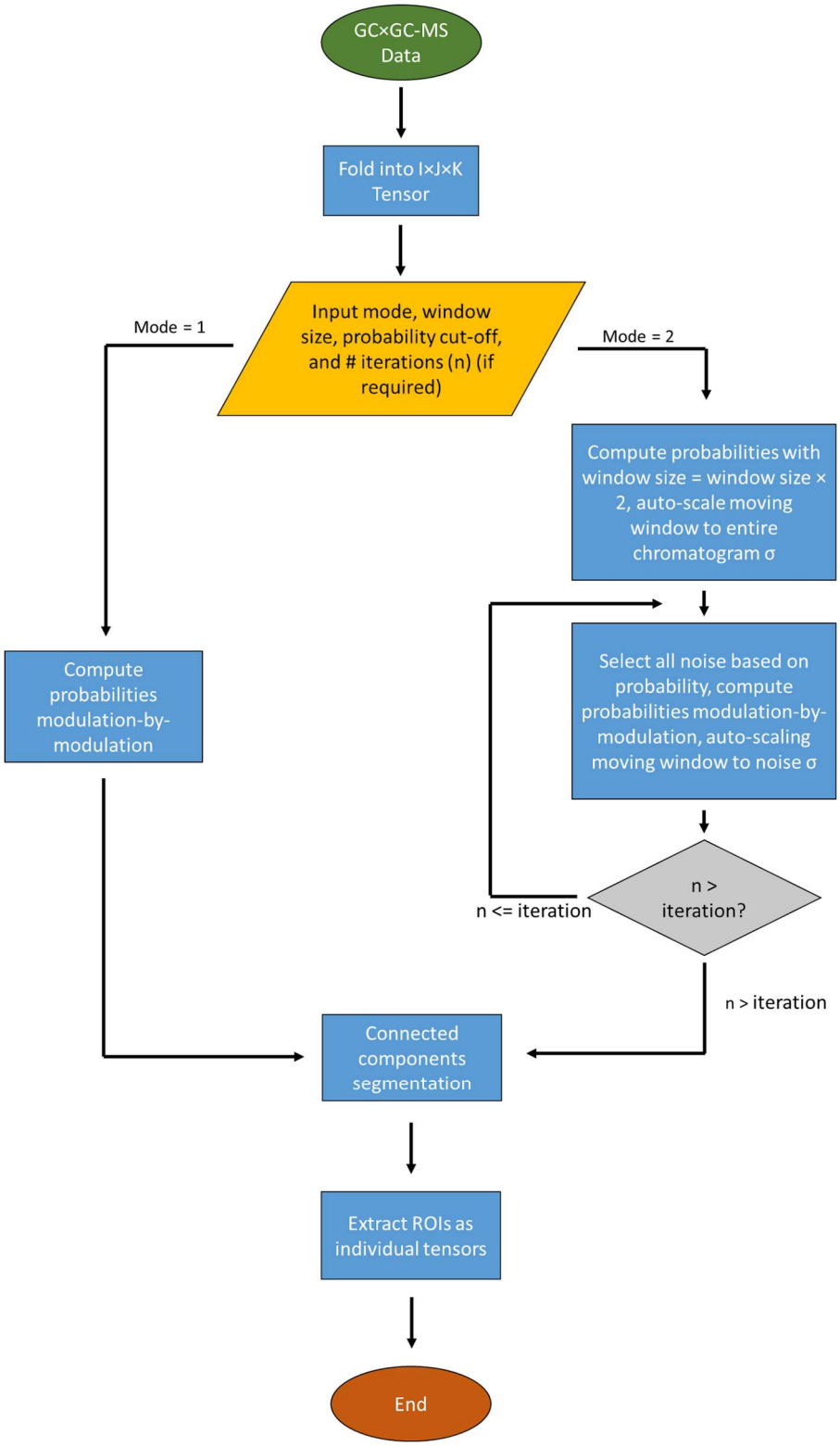


Figure 1. Algorithm flowchart demonstrating the two region of interest selection (ROI) algorithm modes developed for GC×GC-MS data. Mode 1 computes ROIs as an extension of the GC-MS algorithm and mode 2 uses an iterative approach, where the moving window is auto-scaled to the noise regions of the previous ROI calculation.

2.3 Algorithm inputs

The ROI algorithm only requires the raw chromatographic data and five input parameters to perform ROI selection. Users upload the raw chromatographic data as an I × J × K tensor, where I is acquisitions in the first separation dimension, J is acquisitions in the second dimension, and K is individual ions. The moving window (*wndw*) size is an integer value. The probability threshold cut-off (*cutOff*) is the minimum probability (between 0 and 1) for a spectrum to be deemed to be within a ROI. The mode is a binary input (1 or 2) which selects the ROI selection algorithm mode (Figure 1). Users can select to plot the chromatogram after ROI selection by setting plot as 1. If users select mode 2 for ROI selection, the number of iterations must be input (*itters*), which is an integer value greater than 1. If mode 1 is selected, *itters* can be set to any integer. The file gcxgcROIMain.m handles all inputs for this algorithm.

2.4 Algorithm outputs

All algorithm outputs are stored in the structure *dataOut*. The original chromatogram as a tensor with the non-ROI regions removed is stored as the tensor *noiseDropped*. The probabilities across the entire chromatographic plane are stored as an I × J matrix, *arrayPvals*, with probabilities on a scale of 0 to 1. The probabilities above the user defined probability threshold are stored in the I × J matrix *pValCutOff*. The two dimensional total ion chromatogram (TIC) is stored as *ticDataReshaped*. The original chromatogram is stored as the tensor *originalTensor*. An I × J matrix of Boolean values (0 or 1) denoting which acquisitions are in an ROI is stored as *boolArray*. The ROI number for each acquisition is stored in the I × J matrix as *labMatrix*. The number of ROIs detected in the data is stored as an integer as *numROIs*. The user input moving window size, probability cut-off, mode, and number of iterations are stored in the table *metaData*. Additionally, each ROI is excised from the chromatogram and stored as a tensor in the structure *subTensors*. As most ROIs are not rectangular, the excised tensor is padded with zeros to allow for storage as a tensor.

2.5 GC×GC-TOFMS datasets

All data used in the development of this algorithm was collected on a Pegasus 4D TOFMS (LECO, St. Joseph, MI, USA), with an Agilent 7890 (Agilent Technologies, Palo Alto, CA, USA) gas chromatograph and a quad jet liquid nitrogen-cooled thermal modulator. In all cases, the first dimension (^{1}D) column was a 5% phenyl polysilphenylene-siloxane phase (Rtx®−5MS; 60 m × 0.25 mm i.d.; 0.25 µm film thickness) connected by means of a SilTiteTM µ-Union (Trajan Scientific and Medical, Victoria, Australia) to a second dimension (^{2}D) trifluoropropylmethyl polysiloxane-type phase (Rtx-200; 1.6 m × 0.25 mm i.d.; 0.25 µm film thickness). All columns were from Restek Corporation (Restek Corp., Bellefonte, PA, USA). The ^{2}D column was installed in a separate oven located inside the main GC oven.

The derivatized algae data was previously published in Armstrong *et al.* [6]. Blueberry samples were collected using a dynamic headspace (DHS) method previously reported in Dias *et al.* [4]. An alkane standards mix from C_7 to C_{30} in hexane (Sigma-Aldrich, Oakville, Ontario, Canada) was analyzed using a liquid injection method previously reported in Nam *et al.* [5].

3 Results and discussion

3.1 Iterative approach for auto-scaling moving window

Our initial ROI algorithm for GC×GC-MS data, which is a modulation-by-modulation extension of our GC-MS ROI algorithm described in Giebelhaus *et al.*, successfully identified peaks in busy chromatograms (Figure 2). However, when this approach was applied to the simple linear alkanes chromatogram, the algorithm performed poorly, describing the majority of the chromatogram as being within ROIs (Supporting Figure 1).

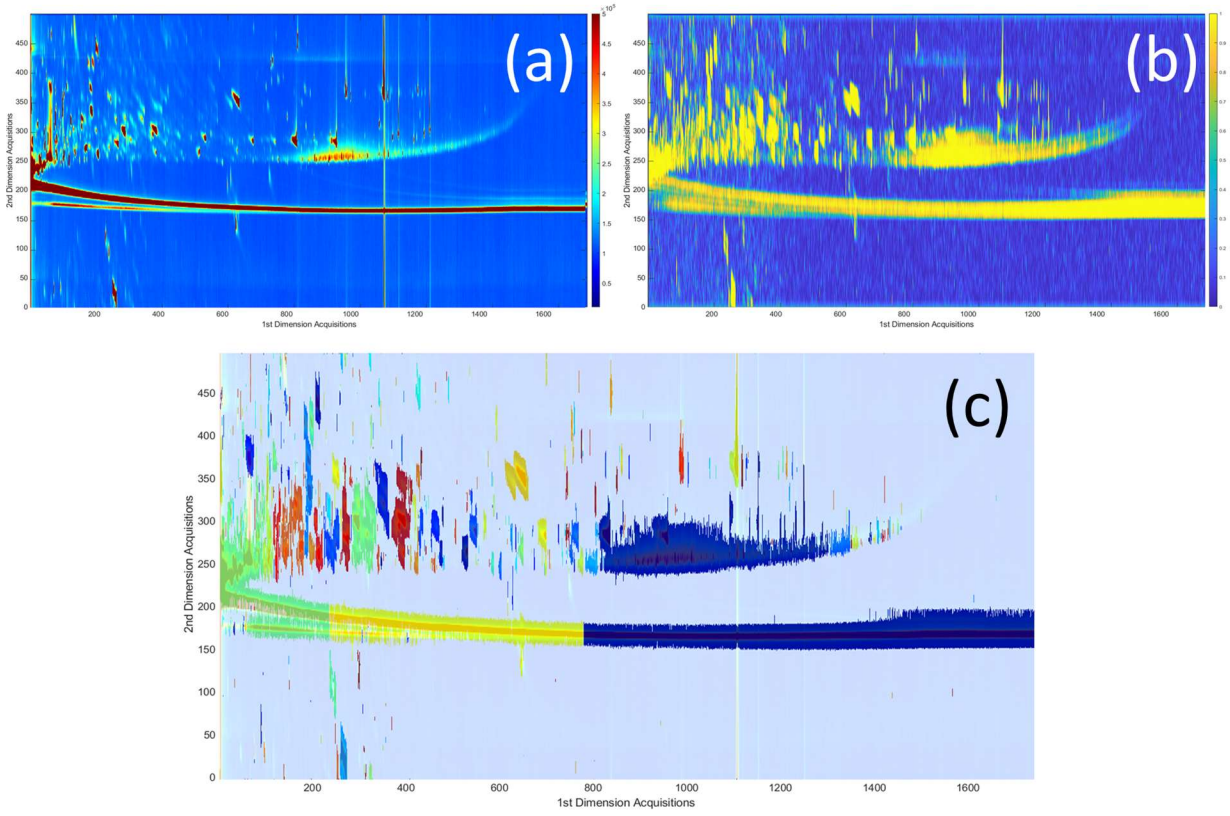


Figure 2. ROI selection on a derivatized algae sample [6] using the modulation-by-modulation extension of the GC-MS ROI detection algorithm. (a) The GC×GC-TOFMS chromatogram of the algae sample and (b) the probability values of the chromatogram, and (c) the ROIs overlaid on the chromatogram.

We hypothesized that this was a result of the moving window's auto-scaled spectra being artificially inflated from low standard deviation (σ). Auto-scaling (Equation 1) is mean centering followed by scaling the data to σ of the data. If a moving window encounters a large section of a modulation absent of any chemical components, σ will be low, which has the potential to artificially inflate the values of the auto-scaled data, especially if only a limited number of x_{ij} are significantly different from $\underline{x_i}$. This phenomenon is well known in chemometrics, where auto-scaling is known to lead to inflation of measurement errors [18].

$$x_{as} = \frac{x_{ij} - \underline{x_i}}{\sigma_i} \tag{1}$$

To eliminate this issue, we first modified our ROI algorithm to auto-scale the moving window using σ of the entire chromatogram. This reduced the false positives in ROI selection, but reduced the sensitivity, due to mass spectra containing chemical components inflating σ. To improve the sensitivity, we then performed the same ROI selection but auto-scaled all moving windows to σ of the non-ROI portions of the chromatogram, equating this to chromatographic noise. We performed this in an iterative process, with each iteration auto-scaling with the σ of the previous ROI calculation. This process is summarized by Figure 1. To decrease false negatives in the initialization step where the moving window is auto-scaled to σ of the entire chromatogram, the moving window size *wndw* is set to *wndw*×2. We found that this approach correctly selected ROIs after about three iterations, with the number of ROIs staying consistent even after 10 iterations (Figure 3). Additionally, to speed up processing of multiple samples, the σ of noise from one

sample after the iterative approach could be used to calculate ROIs in subsequent homogenous samples, circumventing the need to perform multiple iterations for each sample.

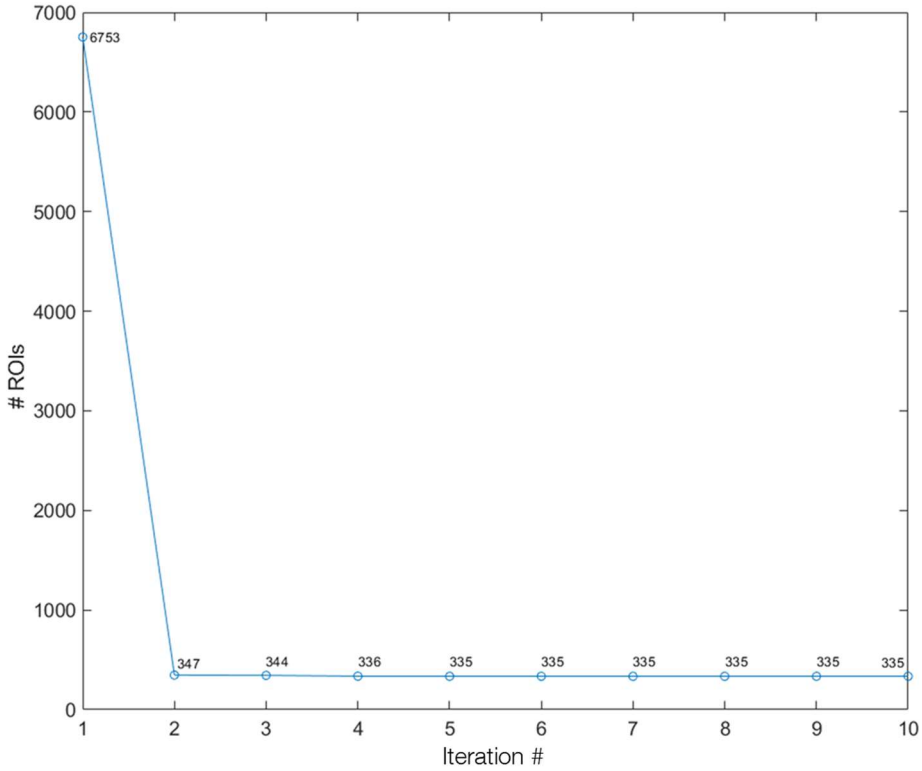


Figure 3. The number of ROIs detected with each iteration of the iterative GC×GC-MS ROI algorithm on a chromatogram of alkanes, with a moving window size of 10 and a probability threshold of 0.8. After five iterations, the number of ROIs remain unchanged.

This iterative approach correctly assigned ROIs in a chromatogram of linear alkanes after 5 iterations, with a moving window size of 10 and a probability cut-off of 0.8 (Figure 4). Spurious ROIs in Figure 4 were identified as real chemical components upon manual inspection, and are likely siloxanes or contaminants in the hexanes solvent.

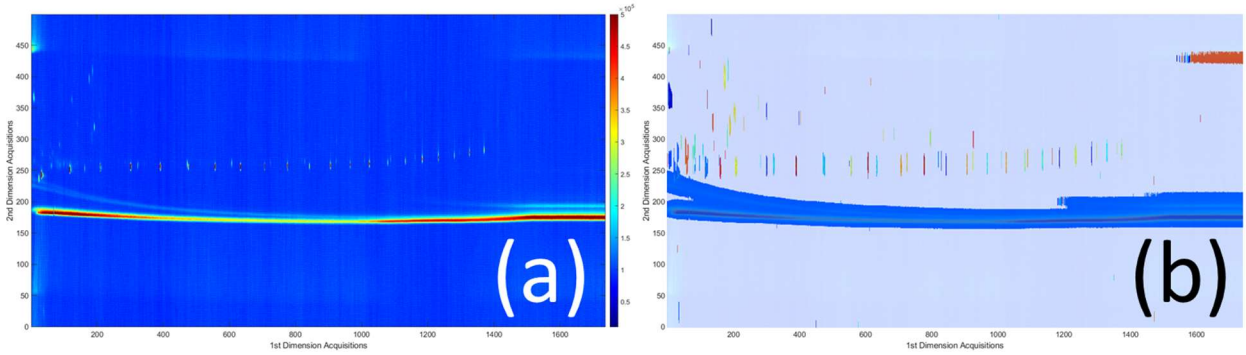


Figure 4. ROI selection on a chromatogram of linear alkanes. (a) Chromatogram of linear alkanes and (b) ROIs overlaid on the chromatogram.

3.2 Optimal settings for region of interest selection

A major advantage of this approach compared to other strategies, including the watershed algorithm [19], is that this algorithm only requires 3 user specified parameters (probability cut-off, moving window size, and iterations), and 2 inputs to select ROI selection mode and the output of a figure. Much like in our GC-MS ROI selection algorithm [15], probability increases rapidly around a ROI, causing the cut-off to have little impact on the size of ROIs (Supporting Figure 2). Decreasing the probability can increase the sensitivity of the ROI algorithm by selecting ROIs with lower probabilities, but this can also increase the size of ROIs and lead to the amalgamation of multiple ROIs into one large ROI, which could make decomposition more challenging. We found a probability cut-off between 0.60 to 0.85 was appropriate for the GC×GC-TOFMS data analyzed in this study.

A moving window size that is approximately the ^{2}D peak width of the average peak over its most intense modulation. We found that the appropriate window size for the GC×GC-TOFMS data explored here was between 10 to 20. Similar to our GC-MS ROI algorithm, larger window sizes help to increase the sensitivity, however this comes with the risk of producing larger ROIs which may not be suitable for multiway decomposition [15]. Additionally, we found that a larger window size is also more sensitive to noise, resulting in the selection of spurious ROIs (Figure 5). A smaller window size may be helpful in resolving peaks eluting closely to the column bleed (Figure 5a). Both the probability cut-off ad the moving window size require minimal user optimization to obtain informative ROIs from GC×GC-TOFMS data.

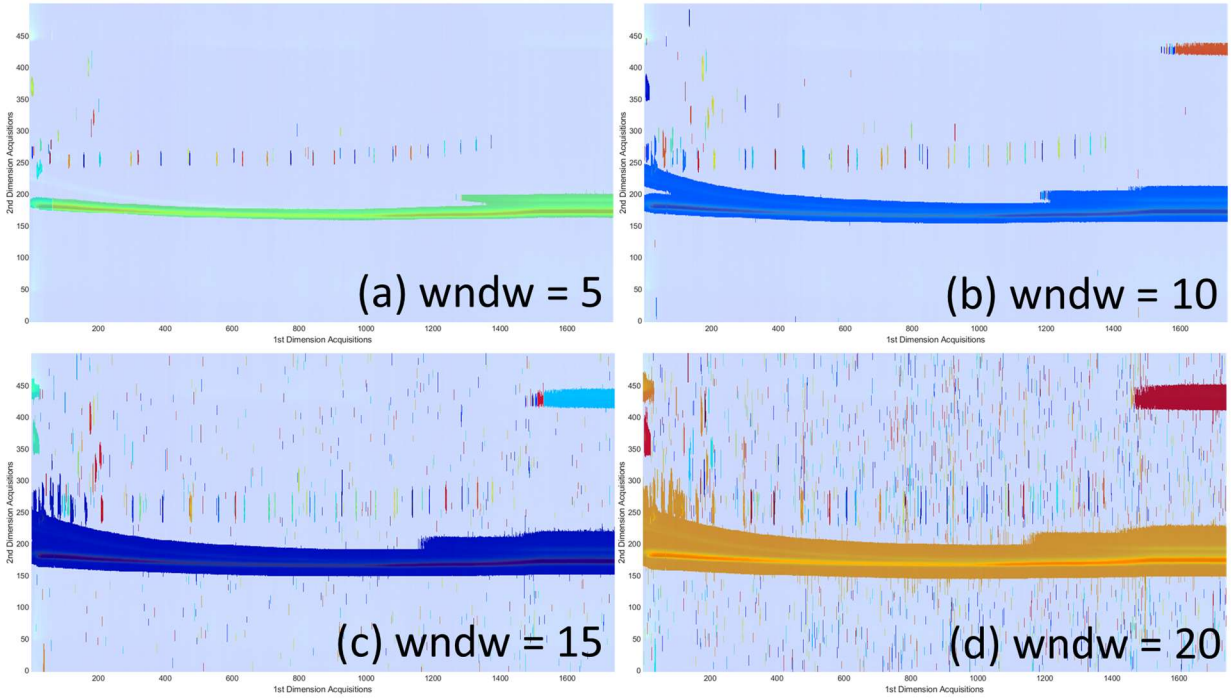


Figure 5. Effects of changing the moving window size on ROI selection.

For GC×GC-TOFMS data we suggest trying the modulation-by-modulation ROI selection algorithm first to determine if the data are amenable to this approach. If this approach performs poorly (Supporting Figure 1), we then suggest trying the iterative approach as previously described, with approximately 5 iterations (Figure 3).

3.3 Performance on complex samples

To assess the performance of this algorithm on GC×GC-TOFMS data representative of real metabolomics studies, two GC×GC-TOFMS datasets analyzing complex mixtures were subjected to ROI selection [4,6]. Our ROI selection algorithm successfully assigned all chromatographic features to an ROI in the complex derivatized algae mixture (Figure 2), with a moving window size of 15 and a probability cut-off of 0.8. The ROI selection on the algae data was performed without the need for the iterative ROI selection approach, likely given the complexity of the chromatogram. We explored the iterative approach on this data, however we found the ROIs became amalgamated (Supporting Figure 3), due to the increased sensitivity of this approach, which would make the data less amenable to multiway decomposition methods.

The blueberry chromatogram [4] is considerably less busy than the algae chromatogram, and therefore performed poorly with the traditional modulation-by-modulation approach, even after trying a wide range of window sizes and probability cut-offs, necessitating the need for the iterative approach (Figure 1). Using 10 iterations, with a moving window size of 15 and a probability cut-off of 0.8, our ROI algorithm successfully selected ROIs in the blueberry chromatogram (Figure 6).

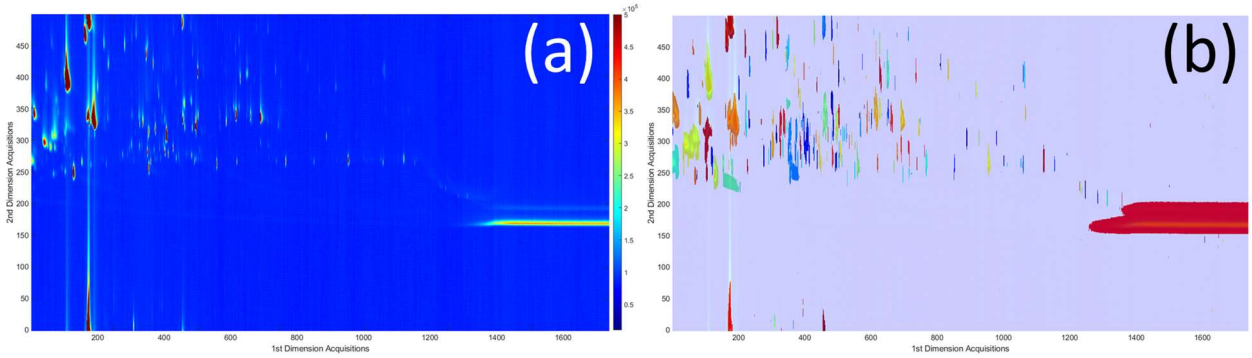


Figure 6. ROI selection on a chromatogram of blueberries [4]. (a) The TIC and (b) ROIs overlaid on the chromatogram.

Complex samples, with closely eluting peaks, often produce irregularly shaped ROIs. Our algorithm extracts the irregular ROIs by extracting the ROI as a cuboid tensor indexed to the ROIs minimum and maximum acquisitions in ^{1}D and ^{2}D. Excess acquisitions in the cuboid are zero-filled. This also addresses the case where an ROI may completely encircle another ROI. A number of strategies have been developed for handling excess zeros when performing decomposition methods, such as MCR and PARAFAC [20,21]. Additionally, we did not encounter peak splitting when computing ROIs, a common issue with the watershed algorithm [19]. This is because our algorithm makes no assumptions about chromatographic peak shape, and uses the mass spectra to select ROIs rather than relying on the TIC [15].

4 Conclusion

This work demonstrates the application of pseudo-Fisher ratio analysis for sensitive, untargeted ROI selection in GC×GC-MS data, by building on our previous GC-MS ROI selection algorithm. Additionally, we addressed the issues of detecting spurious peaks when the GC×GC chromatogram is relatively sparse through our implementation of the iterative ROI selection approach. We generated informative ROIs using only three user-defined input parameters, moving window size, probability cut-off, and number of iterations. Complex samples are efficiently handled by the algorithm, successfully incorporating all chromatographic peaks into ROIs. Our algorithm makes no assumptions about peak shape, making it amenable to a wide range of complex samples and peak shapes. The ROIs are extracted into individual

cuboid tensors, which are agnostic to the ROI shape, allowing for further chemometric analysis. This approach demonstrates an effective component in a workflow for fully automated GC×GC-MS data analysis.

4. Supporting information

All code for this project is available in GitHub at https://github.com/ryland-chem/2d_gc_region_of_interest as release v1.1.

5. Acknowledgements

We would like to thank Dr. Robin J. Abel for providing his expertise in data science, and Dr. Michael D. Sorochan Armstrong for his expertise in chemometrics. We would also like to thank Ryan P. Dias for providing data for the development of this approach. JJH acknowledges the Natural Science and Engineering Research Council of Canada (NSERC) for providing support through grants. RTG acknowledges the Canadian Institutes of Health Research (CIHR) and NSERC for providing financial support through scholarships.

6. CRediT authorship contribution statement

Ryland T. Giebelhaus: Conceptualization, Methodology, Software, Validation, Writing – original draft, Data curation, Visualization, Writing – review & editing. **A. Paulina de la Mata:** Validation, Writing – review & editing. **James J. Harynuk:** Validation, Writing – review & editing, Supervision, Project administration, funding acquisition.

7. Declaration of Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

8. Data availability

Data available upon request.

9. References

[1] T. Górecki, J. Harynuk, O. Panić, The evolution of comprehensive two-dimensional gas chromatography (GC×GC), J. Sep. Sci. 27 (2004) 359–379. https://doi.org/10.1002/jssc.200301650.

[2] R.T. Giebelhaus, L.A.E. Erland, S.J. Murch, HormonomicsDB: a novel workflow for the untargeted analysis of plant growth regulators and hormones, F1000Research. 11 (2022). https://doi.org/10.12688/f1000research.124194.1.

[3] L.A.E. Erland, R.T. Giebelhaus, J.M.R. Victor, S.J. Murch, P.K. Saxena, The Morphoregulatory Role of Thidiazuron: Metabolomics-Guided Hypothesis Generation for Mechanisms of Activity, Biomolecules. 10 (2020) 1253. https://doi.org/10.3390/biom10091253.

[4] R.P. Dias, T.A. Johnson, L.F.V. Ferrão, P.R. Munoz, A.P. de la Mata, J.J. Harynuk, Improved sample storage, preparation and extraction of blueberry aroma volatile organic compounds for gas chromatography, J. Chromatogr. Open. 3 (2023) 100075. https://doi.org/10.1016/j.jcoa.2022.100075.

[5] S.L. Nam, A.P. de la Mata, J.J. Harynuk, Automated Screening and Filtering Scripts for GC×GC-TOFMS Metabolomics Data, Separations. 8 (2021) 84. https://doi.org/10.3390/separations8060084.

[6] M.D. Sorochan Armstrong, O.R. Arredondo Campos, C.C. Bannon, A.P. de la Mata, R.J. Case, J.J. Harynuk, Global metabolome analysis of Dunaliella tertiolecta, Phaeobacter italicus R11 Co-cultures using thermal desorption - Comprehensive two-dimensional gas chromatography - Time-of-flight mass

spectrometry (TD-GC×GC-TOFMS), Phytochemistry. 195 (2022) 113052. https://doi.org/10.1016/j.phytochem.2021.113052.
[7] K. Tarazona Carrillo, N.S. Béziat, G. Cebrián-Torrejón, O. Gros, A.P. de la Mata, J.J. Harynuk, Metabolomic analysis of secondary metabolites from Caribbean crab gills using comprehensive two-dimensional gas chromatography - time-of-flight mass spectrometry—New inputs for a better understanding of symbiotic associations in crustaceans, J. Chromatogr. Open. 2 (2022) 100069. https://doi.org/10.1016/j.jcoa.2022.100069.
[8] T.A. Johnson, M.D.S. Armstrong, A.P. de la Mata, J.J. Harynuk, Exploration of Extraction and Separation Techniques for Routine Trace Analysis of Organic Compounds in Water: Dispersive Liquid-Liquid Microextraction vs Liquid-Liquid Extraction, J. Chromatogr. Open. 2 (2022) 100070. https://doi.org/10.1016/j.jcoa.2022.100070.
[9] K.M. Pierce, J.C. Hoggard, J.L. Hope, P.M. Rainey, A.N. Hoofnagle, R.M. Jack, B.W. Wright, R.E. Synovec, Fisher Ratio Method Applied to Third-Order Separation Data To Identify Significant Chemical Components of Metabolite Extracts, Anal. Chem. 78 (2006) 5068–5075. https://doi.org/10.1021/ac0602625.
[10] S.E. Prebihalo, K.L. Berrier, C.E. Freye, H.D. Bahaghighat, N.R. Moore, D.K. Pinkerton, R.E. Synovec, Multidimensional Gas Chromatography: Advances in Instrumentation, Chemometrics, and Applications, Anal. Chem. 90 (2018) 505–532. https://doi.org/10.1021/acs.analchem.7b04226.
[11] M.J. Wilde, B. Zhao, R.L. Cordell, W. Ibrahim, A. Singapuri, N.J. Greening, C.E. Brightling, S. Siddiqui, P.S. Monks, R.C. Free, Automating and Extending Comprehensive Two-Dimensional Gas Chromatography Data Processing by Interfacing Open-Source and Commercial Software, Anal. Chem. 92 (2020) 13953–13960. https://doi.org/10.1021/acs.analchem.0c02844.
[12] M.D. Sorochan Armstrong, J.L. Hinrich, A.P. de la Mata, J.J. Harynuk, PARAFAC2×N: Coupled decomposition of multi-modal data with drift in N modes, Anal. Chim. Acta. 1249 (2023) 340909. https://doi.org/10.1016/j.aca.2023.340909.
[13] J.M. Amigo, T. Skov, R. Bro, J. Coello, S. Maspoch, Solving GC-MS problems with PARAFAC2, TrAC Trends Anal. Chem. 27 (2008) 714–725. https://doi.org/10.1016/j.trac.2008.05.011.
[14] I.H.M. van Stokkum, K.M. Mullen, V.V. Mihaleva, Global analysis of multiple gas chromatography–mass spectrometry (GC/MS) data sets: A method for resolution of co-eluting components with comparison to MCR-ALS, Chemom. Intell. Lab. Syst. 95 (2009) 150–163. https://doi.org/10.1016/j.chemolab.2008.10.004.
[15] R.T. Giebelhaus, M.D.S. Armstrong, A.P. de la Mata, J.J. Harynuk, Untargeted Region of Interest Selection for GC-MS Data using a Pseudo F-Ratio Moving Window, J. Chromatogr. A. 1682 (2022) 463499. https://doi.org/10.1016/j.chroma.2022.463499.
[16] R.A. Poldrack, Region of interest analysis for fMRI, Soc. Cogn. Affect. Neurosci. 2 (2007) 67–70. https://doi.org/10.1093/scan/nsm006.
[17] Q. Zhang, H. Xiao, Extracting Regions of Interest in Biomedical Images, in: 2008 Int. Semin. Future Biomed. Inf. Eng., 2008: pp. 3–6. https://doi.org/10.1109/FBIE.2008.8.
[18] B.G.M. Vandeginste, D.L. Massart, L.M.C. Buydens, S. De Jong, P.J. Lewi, J. Smeyers-Verbeke, Chapter 35 - Relations between measurement tables, in: B.G.M. Vandeginste, D.L. Massart, L.M.C. Buydens, S. De Jong, P.J. Lewi, J. Smeyers-Verbeke (Eds.), Data Handl. Sci. Technol., Elsevier, 1998: pp. 307–347. https://doi.org/10.1016/S0922-3487(98)80045-2.
[19] G. Vivó-Truyols, H.-G. Janssen, Probability of failure of the watershed algorithm for peak detection in comprehensive two-dimensional chromatography, J. Chromatogr. A. 1217 (2010) 1375–1385. https://doi.org/10.1016/j.chroma.2009.12.063.
[20] J.-G. Jang, J. Lee, J. Park, U. Kang, Accurate PARAFAC2 Decomposition for Temporal Irregular Tensors with Missing Values, in: 2022 IEEE Int. Conf. Big Data Big Data, 2022: pp. 982–991. https://doi.org/10.1109/BigData55660.2022.10020667.
[21] N. Omidikia, M. Ghaffari, R. Rajkó, Sparse non-negative multivariate curve resolution: L0, L1, or L2 norms?, Chemom. Intell. Lab. Syst. 199 (2020) 103969. https://doi.org/10.1016/j.chemolab.2020.103969.