Dear Dr Schopfer,

I am pleased to resubmit a revised version of the Manuscript PONE-D-17-07220 “GCalignR: An R package for aligning Gas-Chromatography data”. I appreciated the constructive criticisms of the Associate Editor and the reviewers. I have addressed each of their concerns as outlined below.

The most substantial concerns were related to the novelty and performance of algorithm implemented in the R package GCalignR in comparison to other suggested peak alignments tools that are available for GC-MS data. Furthermore, the comments revealed that our manuscript was partly no precise enough to convey that our algorithm and the package are primarily intended to process peak lists derived from one-dimensional chromatography (GC-FID) and not two-dimensional GC-MS data.

REVIEWER 1 COMMENTS:

Reviewer #1: Ottensmann et al. introduce GCalignR, an R package to align GC data. They justify the necessity of their package since numerous tools exist for GC-MS data alignment, but there is no alternative for GC coupled to other (not-defined) detectors. However, there are indeed packages and tools for GC-MS peak alignment [1-4], and concretely [1,4] an R package called ‘ptw’.

Adapting those solutions, from GC-MS to GC should be straightforward, since all these align each m/z channel independently, and thus, a TIC (or any other single signal) generated by GC can be aligned as such (as if it was a single m/z channel). Second, methods and results are difficult to understand. From the text, it is not clear how GCalignR aligns peaks (see other comments). Also, the metrics used for the results section (Fig 6, the number of substances) are not indicative of the method performance. My suggestions about what this study needs to reach the novelty and quality required for publication are as follows:

1. *We do not disagree that there are numerous alignment algorithms for multidimensional chromatography data e.g. GC-MS or LC-MS. However, these do not support GC-FID and other one-dimensional chromatograms, because they rely on mass spectra or putatively other spectra that are available for each peak. In the case of a GC-FID dataset no spectral information is available that is why the retention time is the only measure that allows aligning. We realised that our wording was imprecise. Therefore, we edited the manuscript to point out that our algorithm and our package are thought to fulfil the alignment of peak lists derived from one-dimensional chromatography techniques, illustrated on GC-FID.*

1. Authors use a peak list, but they do not specify how they obtain it. Are all the GC vendors’ software allow for automated peak picking? If so authors should state that, so users are aware that the peak list can easily be obtained from their vendors’ software.

1. *We included a more concise description of the peak list. However, we are confident that all GC-FID software allows to output a peak list, as the aim of this technique is to detect and quantify chemicals within a sample, whereby the annotation is based on the retention time until detection. We checked three commonly used software applications Xcalibur, ChemStation and XXX that are distributed by Thermo Fisher Scientific (United States), Agilent Technologies (United States) and Shimadzu (Japan) respectively. All provide automatized tools to detect and export peaks in form of a list.*

2. However, a peak picking algorithm could be easily implemented, for instance, with the ‘MassSpecWavelet’ R package (Bioconductor). This would enable GCalignR to perform the complete GC data processing pipeline without third party vendor software.

1. *We appreciate this remark to consider enhancing the pipeline. However, peak detection by* *continuous wavelet transform-based pattern matching as implemented in MassSpecWavelet (Du, Kibbe, Lin 2006) relies on spectral features that are captured by m/z ratios in GC-MS experiments. This method is not applicable using one-dimensional chromatography techniques (e.g. GC-FID), as each ionized molecule produces only an electric current that is indicative of a peak but not substance specific as variation in quantities is not unexpected. Hence, there is nothing comparable to a spectrum in GC-FID that could be used for an alignment except the retention time. Moreover, there are many software solutions available that can be used to call peaks, some of which are specific to used hardware and are available from vendords (eg. Agilent Technologies, Shimadzu). Moreover, any suitable peak calling software written in the R language can be easily implemented before working on the so obtained peak list with GCalignR.*

2. A robust metric to evaluate the method performance is to analyze the same experiment with GC-MS and GC. This is how authors have tested GCalginR, however, I do not believe that the metrics employed are the most illustrative. The same peak list could be aligned with other GC alignment tools, such as [1].

1. *Wehrens et al. (2015) implemented parametric time warping for aligning spectra from LC-MS or LC-DAD. These data provide a spectrum (m/z ratios) for every compound i.e peak. Contrary, the peak list used in GCalignR cannot be aligned using the package ‘ptw’ as there is no spectral information available for GC-FID.*

Then, the performance could be assessed by introducing a random noise, and then evaluate the quantitative performance (R-squared) between GCalignR and [1], given a reference (ground truth) quantitative peak area of a selective m/z for each compound. In other words: provide the R-squared for each molecule, between GCalignR and the reference (selective m/z) peak area across samples; and the same between ‘ptw’ and the selective m/z.

I believe that the implementation of these suggestions have a low technical complexity, and that they will greatly improve the applicability of GCalignR in GC studies.

1. *As stated in (4), GCalignR does not support spectral features and is not intended to align them. Therefore, we do not see a possibility to conduct a benchmark experiment against other software in R based on a peak list as used in GCalignR. We rephrased ambiguous text in our manuscript to avoid confusion here. GCalignR is intended for GC-FID (cf. (1)) and peak list with one dimension in general. It read clearer now.*

Minor:

- Line 42: anonymous should read unknown

*Done*

- Although analytical methods are described in another paper, a brief courtesy methods paragraph should be included. Authors could name which type of detectors coupled to GC is GCalginR compatible with.

*A list of compatible datatypes is now included*

- Section (ii) Peak alignment, this section is confusing and more effort should be devotted to clarify the GCalignR algorithm. For instance, how is the first matrix (line 109) calculated?

*Done*

- Line 248, 257-258. Units are mandatory.

*Done*

[1] Ron Wehrens Tom G. Bloemberg Paul H.C. Eilers. Fast parametric time warping of peak lists. Bioinformatics (2015) 31 (18): 3063-3065.

[2] Perera, V., De Torres Zabala M., Florance, H., Smirno , N., Grant, M., Yang, Z. R. Aligning extracted LC-MS peak lists via density maximization. Metabolomics. 8 (2012) 175–185.

[3] Wei X, Shi X, Merrick M, Willis P, Alonso D, Zhang X. A method of aligning peak lists generated by gas chromatography high-resolution mass spectrometry. Analyst. 138 (2013) 5453-60.

[4] Bloemberg, T. G. et al. (2010) "Improved Parametric Time Warping for Proteomics", Chemometrics and Intelligent Laboratory Systems, 104 (1), 65-74

REVIEWER 2 COMMENTS:

The authors present a technique implemented in R for retention time alignment of GC data. They demonstrate the data on real data to show noise insensitivity.  
  
Major:  
Page 2, paragraph starting with line 48:  
Algorithms and software are not synonymous. You have opted to describe both in one manuscript, which is fine, but should realize that the description and evaluation of each are very different. What you’ve done is created a piece of software for alignment (and anyone could substitute your algorithm for their own), and created an algorithm for alignment.

1. *We tweaked the manuscript by separating the description of the algorithm from the package itself and emphasise that other algorithms (if available) could be used as a substitute*   
     
   A split treatment would facilitate the reproduction of the algorithm, including for external validation, without being tied to the implementation, and would allow the evaluation of the implementation independent of the alignment algorithm it is used with.  
     
   I would like to see you tweak the text to separately discuss them. If you are introducing a novel algorithm, its description should be in the methods, not the introduction. It should also be described in sufficient detail that it can be reproduced from the paper. The current description is vague enough to not only be unimplementable, but also too vague to see how your algorithm differs from the over 100 alignment algorithms already published.
2. *The algorithm for aligning individual peaks is formalised in formulas 1 and 2 of the manuscript. We moved the description to the appropriate section as suggested.*  
     
   It is very difficult to write English language in a way that carries the mathematical precision of code. I would recommend replacing English descriptions of code with pseudocode or a flowchart. These descriptions should definitely not be in Figure legends (Fig 2), or the introduction.
3. *We inserted a flowchart as recommended.*  
     
   Your method has a dozen user-set parameters. In your evaluation against a ground truth benchmark, you explored 100 combinations of parameters. What strategy should users use to find optimal settings for 12 parameters on real (non-ground truth) data, where supervised scoring is not available? This is a very significant issue.
4. *We added a paragraph suggesting a strategy, simultaneously emphasised to critically think of the own data, how do the chromatograms look like? Which range of retention times are there? All these a specific for a dataset and are expected to influence alignment results.*  
     
   On a related but separate note, I would like to see a figure showing the results of those 100 combinations to give the user a sense of what random parameter settings (which may be the best they can do with real data) will yield. In a sense, this would be the most valuable result in the paper, as it shows what a user can expect in terms of real world performance.
5. *This experimental outcome is depicted in figure 7 that shows a three-dimensional plot of all 100 parameter combinations independently for three sets of data.*   
     
   This paper is well written, but the content is written from the perspective that there are no other alignment methods published. In reality, there are dozens. This paper cannot be published without a comparison to existing methods. While it is too much to ask for a comparison to the 100+ papers that have been published, the following is a bare-minimum to show that the algorithm is superior to what is already available:  
   -A quantitative comparison of results for 2-3 top (either quantitatively best or most popular) alignment methods.  
   -An algorithmic comparison to these other methods (how is what you are doing different from what they did).  
     
   An extensive review on alignment was published in 2015 by Smith et. al: “LC-MS alignment in theory and practice: a comprehensive algorithmic review.”
6. *The mentioned publication reviews alignment methods of spectral features and is therefore not comparable to the gap that GCalignR aims to fill, which is the alignment of peak list from one-dimensional chromatography techniques (GC-FID etc.). However, we added a paragraph that summarises the main techniques that are used for GC-MS and highlight why they are not applicable to peak lists of GC-FID data.*

The authors will find a comparison of 50 alignment algorithmic approaches, and an analysis of continuing limitations of all published algorithms. Notable among the latter: Model assumptions that fail to capture real behavior, long run times due to algorithm complexity and user-defined parameter optimization, pairwise comparisons and reference samples, current methods have not been empirically compared. The authors will note that their method as described in this manuscript is subject to all of those shortcomings.

1. *Interesting note. We elaborated our statement of caution for careful data inspection by the user and included an approximation for a-priori prediction of run times.*   
     
   If your algorithm is as non-novel as your current description suggests, the paper will have to be re-written with the contribution being an R program for facilitating an alignment algorithm, rather than an R program with a novel alignment algorithm. The focus and content would need to be revised to highlight how users can implement the algorithm of their choice, how they can use the program’s visualization and other inspection tools to tune algorithm parameters, etc.
2. *As outlined in responses (1-4) we do argue that our algorithm is novel and was not implemented elsewhere. For the fine-tuning of the parameters and the usage of the package we refer to an enlarged version of the package vignette, distributed with the package as this is standard for writing R packages. The software is completely open source code that can be accessed easily on our GitHub repository. Therefore, users are free to change and extend the R package, which includes the implementation of other functions or algorithms. It is easy to incorporate these into the project to make it available for the community.*   
     
   Minor:  
   1. What data formats are accepted by the application? Are they standard? Does any other pipeline use these data formats? Is there a mapping between standard data mappings and what your program accepts/outputs?
3. See response (2). Not surprisingly, output formats do differ among software. This is why we use a common format by which users simply concatenate the output from different samples either by pasting the data in a program like Microsoft Excel or by using a simple custom script in R. Output for further analysis are data frames, which one would term to be a sophisticated format within R that is expected by many methods including the vegan package as shown.

2. Page 2, line 64. There are literally dozens of open source alignment programs available. This is not an advantage over what is currently available.

1. *We tweaked the text to be more precise that our arguments exclusively aim on GC-FID and other one-dimensional techniques, for which the published algorithms are not applicable (cf. responses 1-4).*