Manuscritpt

Fitting segmented mixed models with random changepoints in R

Dear editor

I would like to thank you and the three referees for your insightful comments and suggestions which greatly improved the paper.

I have followed all recommendations (**Required** and **Recommended**) by you and referees but two: 1) I don't have added any "medical comment' on the results in the first example: I don't have the skills to do it; 2) no diagnostic analysis on the fitted models (it would require a not-so short treatment and I think it is out the aims of the paper). With respect to the first version, I have modified the text to follow the referees' suggestions, I have modified (slightly)the title, and I have added treatment of a couple of new functions (mainly breakpoint()). Also, the results in the 1st example are slightly improved (in terms of Likelihood) as I have modified the fitter functions to avoid early stopping of the algorithm.

In addition to this letter, I am uploading the new manuscript (pdf), the source file including the command lines used in the paper, and the new version of the segmented package. I plan to upload it to CRAN, along with detail on this manuscript (link, doi?...), as soon as I receive your (positive) reply about the acceptance of the paper.

Below my point-by-point reply to your comments

Thank you very much for your work,

Kind regards,

Vito Muggeo

Required (the decision rests on these):

Reviewer 1 Example 2 states its aim is to validate the functions against the true simulation parameters, then reports "results not shown here". Show the summary(oseg1) comparison to the true values, or drop the validation framing and present the example as illustration only.

I have eliminated the "validation framing", in that I use the verb "to assess". Moreover, I show the results by means of the new function fixef and getVarCorr (rather than summary)

Reviewer 3 Example 1 is not reproducible as written because bootstrap restarting is stochastic and no seed is set, while Example 2 does set seeds. Set a seed (or the relevant seg.control option) so the printed estimates reproduce, or state in the text that bootstrap-restarting results can vary across runs.

I have added in the text the comment about the seed for boot restarting: "segmented.lme() uses a seed coming from the response values" (see page 4)

Recommended:

[methods-soundness] The text states the breakpoints make the model non-regular, yet the summary and slope outputs report Wald t-values and p-values for the changepoint (G0) and slope-difference (U) parameters, and the narrative reads significance directly from them ("significantly different", "a larger changepoint"). Near-boundary and changepoint parameters are exactly where Wald inference is least reliable under non-regularity. Please add one or two sentences stating which quantities have trustworthy standard errors and advising caution for the changepoint-related terms, so readers do not transfer the displayed p-values uncritically.

Yes, you right. Done.

[software-quality] I am assessing from the manuscript and the described public interface. The package source, its help pages, and any tests/ directory were not supplied, so I cannot confirm that the documented behaviour of segmented.lme, plot.segmented.lme, sim.seg, and slope is covered by automated tests. State in the paper whether the demonstrated functions are exercised by the package test suite, and point readers to where those tests live.

At the submissions (the first and this second one), I also uploaded the source file with code and the new version of segmented which will be available on CRAN (if and) when the paper will get accepted.

[software-quality] The examples set seeds (set.seed(123), seed=15, seed argument of sim.seg) and use shipped data with no hard-coded paths, so the runs are portable. The paper does not record the R version or a captured session state. Add one line giving the R version and sessionInfo() (or the exact segmented/nlme versions) used for the printed output, so the numeric results can be reproduced.

I have added the output of sessionInfo()

[software-quality] The code appears across the text as interleaved snippets rather than one runnable file. Provide the two examples as a single script or a package vignette that runs end to end, so a new user can reproduce Figures 1 to 3 and the printed estimates without reassembling fragments.

At the submissions (the first and this second one), I also uploaded the source file with code and the new version of segmented which will be available on CRAN (if and) when the paper will get accepted. Analyses and plots can be reproduced by running the file I also uploaded.

[statistical-rigor] The analysis of glucoseData describes the Obese curve as "significantly different" from Control on the basis of several contrasts read off one summary table (intercept, left slope, slope difference, changepoint) with no correction across them. Please restate these conclusions in terms of the estimated differences and their intervals, which the example already produces, and drop the threshold wording so the example does not model uncorrected multiple testing for readers who copy it.

I have eliminated the "significantly different" wording

[statistical-rigor] The multinormal random-effects distribution in equation (4) and the zero-mean residual assumption are stated but never examined in either example. One sentence pointing to a residual or random-effects diagnostic, alongside the standardized-residual summary already printed, would show readers how to check the fit rather than assert it.

Yes, you are right. Diagnostics on residuals are crucial to check the model reliability and underlying assumptions. However, such discussion is beyond the aim of this paper. I hope you understand.

[statistical-rigor] The simulated example validates recovery on a single draw (seed 15) at one noise level. Since the Conclusions state the method "work[s] reasonably well when the signal-to-noise ratio is modest to high", a short note on the range of conditions behind that statement, using runs you have already done, would let readers judge where the recovery claim holds.

I have added a small simulation study at the end of section 3 to show the performance on several simulated datasets.

[visualisation] Figure 1 packs 33 subjects into a 3 by 11 grid. At that size the per-panel axis annotation and the "time"/"phos" labels are hard to read, and only a handful of panels carry tick values. Consider a 5 by 7 or 6 by 6 layout, or a single faceted plot with shared axes and one axis label per edge, so the reader can compare profiles without squinting.

Yes, you are right. I use a 5x7 grid.

[visualisation] The red/blue coding in Figures 1 and 2 reads for the common red-green colourblindness because the two hues differ in blue content, so the palette is broadly safe. Figure 2 reinforces colour with line type (lty=3 for Obese), which is good practice. Carrying that same line-type distinction into Figure 1, and stating the red/blue meaning once in the main text as well as the caption, would make the group contrast legible in greyscale print too.

Yes, I have followed your suggestion

[visualisation] Axis labels give variable names ("time", "phos") without units. One phrase in each caption, naming the time unit and that phos is plasma inorganic phosphate, would let the figures stand on their own.

Yes, I have followed your suggestion (in the first plot).

[writing-quality] Section 1 ends the sentence "the R package segmented (version 2.3-0 or higher)" without a full stop, and the paragraph in Section 2 ending "in the segmented, in particular" and "demanding … more demanding" trails off without a closing period in two places. A pass to close these sentences would help the reader.

Yes, done.

[writing-quality] The summary(os1) output shows no t-value or p-value for the U and G0 rows while neighbouring rows report both. One sentence saying why these entries are blank (for example that the changepoint and reference slope-difference are reported without the associated test) would prevent readers from reading it as a defect in the output.

Yes, I have added comments about that on page 6, just after the first call to summary.

[ethics-and-integrity] The empirical example uses glucoseData, a de-identified dataset of 33 patients (Control and Obese) shipped inside the segmented package, analysed as a secondary illustration. No individuals are identifiable and no new data were collected, so the absence of an ethics-approval or consent statement does not meet the threshold for a violation. Add one sentence naming the original source of glucoseData and its public availability, so the provenance of the human measurements is on record.

Yes, done.

[ethics-and-integrity] The paper is a tutorial for the author's own segmented package, a relationship visible from the author line and the citations. A short competing-interests line stating that the author maintains the software would make the interest explicit. Add a competing-interests statement.

Not sure to understand what you mean, I am sorry. However, people can find information on the package, including the maintainer, on CRAN.

Reviewer 1 Provide a single self-contained script that runs both examples end to end, with the output of sessionInfo() and the segmented/nlme versions used.

I have added the output of sessionInfo()

Reviewer 1 Comment on the near-zero random-effect standard deviations in Example 1 (time and U, around 0.002) and link them to the overparameterization guidance already given in the Conclusions.

Yes, I have followed your suggestion by including some comments in the text.

Reviewer 1 In summary(os1) the U and G0 rows carry no p-value (and G0 no t-value). State what these entries mean, or confirm they are intentionally omitted.

Yes, you are right. I have added some comments about the "missing" p-values just after the first call to summary.

Reviewer 2 The summary of os1 omits t and p columns for G0 and the p column for U. Add one sentence explaining the omission and the non-regular nature of inference on the changepoint.

Yes, you are right. I have added some comments about the "missing" p-values just after the first call to summary.

Reviewer 2 Add a small repeated-simulation check (bias and interval coverage over replicates) to support the validation claim, since one dataset recovers one draw.

Yes, done. I have added a small simulation study

Reviewer 2 Give a short account of the approximate restricted log-likelihood (the "approx" in the print output) and what the approximation affects.

I have added a short comment about the approximate Likelihood. However I provide no detail about that by mentioning the original methodological paper. I hope you are fine with my choice

Reviewer 2 Expand the guidance for choosing among the four covariance specifications into a short worked rule, since convergence difficulty is the practical obstacle you identify.

Yes, you are right. I have added comment about that.

Reviewer 3 Report R version, operating system, and the exact versions of segmented and nlme used, and pin an exact segmented version instead of "2.3-0 or higher" for the numeric outputs.

optional:

I have added the output of sessionInfo() at the end of manuscript.

[methods-soundness] The statement "the more complex the model, the better the fit and corresponding predictions" is loose, since richer random-effect structures risk overfitting and the convergence problems the same paragraph describes. A brief note tying model choice to an information criterion, which the AIC/BIC output already provides, would make the guidance concrete. This is beyond the illustrative scope of the current examples.

Yes, you are right. I have added a short comment in the text about model selection via BIC/AIC.

Reviewer 1 (conditional) One or two sentences on the clinical reading of the Control-versus-Obese changepoint difference would help an applied reader connect the fit to the glucose tolerance test.

Yes, it could be but I am a statistician and I am not able to comment that from a medical viewpoint. I am sorry, I hope you understand that

We will correct these in production, and you confirm them in the proof:

[consistency-check] In Example 2 the text names the additional segmented profiles as "the 3rd, the 4th and the 10th", but the code call plot(oseg1, id=c(4,28,3,5,10), ...) and Figure 3 show subjects 3, 5

and 10 (subject 4 is one of the two linear profiles). The “4th” should read “5th” to match the code and figure.

[writing-quality] In Section 2, “In particular, the the ‘plot’ method can be used” repeats “the”. This is a duplicated word, not an extraction artefact.

[writing-quality] In Section 3.1 the summary discussion reads “the Obese group has an higher intercept”; this should read “a higher”.

Yes, your right. Everything has been fixed.