PEER REVIEW RECORD

Fitting segmented mixed models with random changepoints in R

Vito M.R. Muggeo

Vol. 1 (2026), Art. 12 · https://doi.org/10.68365/jrepstat.2026.12

The decision letters, the reports the authors read, their answers and the editor's signed decisions, as written. The reports come from the journal's review engine. The authors could have kept this record private until they confirmed the proof.

DECISION PATH

2026-10-04	Minor revision
2026-10-30	5 answers
2026-10-31	Minor revision
2026-11-03	14 answers
2026-11-03	Minor revision
2026-11-06	4 answers
2026-11-18	Accepted

Round 1: Minor revision

SIGNED DECISION

Confirmed and signed by Matthias Templ on 2026-10-04 · ORCID 0000-0002-8638-5276

DECISION LETTER

Thank you for your submission to JRepStat. We are inviting a minor revision. The reviewers' points follow, grouped by weight.

This paper shows, with one real and one simulated dataset, how to fit and visualise segmented mixed models with random changepoints using segmented.lme in the R package segmented. It matters to applied analysts of longitudinal and clustered data who need subject-specific breakpoints estimated in a likelihood framework. We would be glad to see the revised version.

How to read this letter. The decision rests on the items marked required. If a required item seems wrong to you, or cannot be done within this paper, say so in your response letter and give the reason. A reasoned decline is a legitimate answer, and the editor weighs it. Reviewers can be mistaken. Items marked recommended or optional are advice from the reviewers. Adopt them, decline them with a sentence, or leave them for later work. None of that counts against the paper. The reviewer reports below carry the same marks. The editor who signs this decision may weigh an item differently in the accompanying message, and that message prevails.

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.
- [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.

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.

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

- [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.
- [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.

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".

REVIEWS

Report Reviewer 1, recommends minor

Written by the journal's review engine in the role of the application reviewer, and read by the editor before the decision. How the engine works is on the peer review page.

This is a clear, practically useful software illustration from the package author, showing how segmented.lme fits piecewise-linear mixed models with subject-specific breakpoints. The underlying method is already published (Muggeo et al., 2014), so the contribution here is the code, the argument reference, and two worked examples. For an applied reader who has longitudinal data with a plausible turning point per subject, the paper gives a usable path from an lme fit to a segmented fit, to group comparisons, to plots. The structure of the functions section is a good quick reference, and the glucose example reads well.

My main concern is in Example 2, which the text sets up explicitly as a validation exercise, comparing estimates to the true simulation parameters, and then reports the key comparison as "results not shown here". For a software paper the recovery of known parameters is close to the point of the example, and an applied reader cannot judge whether the functions behave as claimed without seeing it. The boundary-detection part (identifying subjects 4 and 28 as linear, matching dd$perc.psi) is shown and is convincing. The parameter-recovery part is asserted. Please either print the summary(oseg1) comparison against the true values, or present the example as an illustration of the workflow and remove the validation language.

Two things in Example 1 would help a practitioner more than they currently do. The diagonal fit reports random-effect standard deviations for time and U of roughly 0.002, which are effectively zero and suggest those random components are not supported by these data. Your Conclusions already warn about exactly this situation, so one sentence connecting the warning to the visible output would teach the reader how to read it. Separately, the summary(os1) table shows no p-value for U and no t-value or p-value for G0. If that omission is deliberate, because testing a zero slope-difference or a zero-valued changepoint is not a standard test here, say so in a line, because an applied reader will otherwise read the blank as missing output.

On reproducibility, the code is embedded in the text and uses set.seed and seed arguments, which is good practice. The submission record lists reproducibility artifacts that I was not given, so I cannot confirm whether a single runnable script exists. If one does not, a self-contained script plus the output of sessionInfo() and the segmented and nlme version numbers would let a reader reproduce both examples without reassembling snippets from the prose. The package version floor (2.3-0 or higher) is already stated, which is helpful.

The interpretive language around the group comparison ("significantly different", "it turns out the curve for the Obese patients is significantly different") leans on significance rather than on the estimated differences and their uncertainty, which you already report. Stating the intercept, slope and changepoint differences with their intervals, as you do for the slopes, would make the applied reading cleaner. This is a presentation point and not a barrier.

Numbered suggestions to the authors:

1. REQUIRED In Example 2, show the summary(oseg1) estimates against the true simulation parameters that the example says it validates, or remove the "validate" framing and present the example as a workflow illustration.
2. RECOMMENDED Supply one self-contained script that runs both examples, together with sessionInfo() output and the segmented and nlme versions used.
3. RECOMMENDED Add a sentence in Example 1 noting the near-zero random-effect standard deviations for time and U, and connect it to the overparameterization guidance in the Conclusions.
4. RECOMMENDED Explain the blank entries for U and G0 in the summary(os1) table, or confirm they are omitted by design.
5. OPTIONAL (conditional) Add one or two sentences on what the Control-versus-Obese changepoint difference means for the glucose tolerance test, to anchor the fit for applied readers.

Report Reviewer 2, recommends minor

Written by the journal's review engine in the role of the methodology reviewer, and read by the editor before the decision. How the engine works is on the peer review page.

This is a clear software illustration of a method that is already established in Muggeo et al. (2014). As a methodology read, my job is to check that the illustration represents the estimator and its caveats correctly, and that the claims the examples make are supported by what the examples show. On the first count the paper is in good shape. The model in equation (1), the decomposition into fixed and random parts in (2), the two changepoint parametrisations in (3) with the logistic form keeping psi inside (a1, a2), and the covariance structure in (4) are all stated correctly, and the four random-effect specifications are ordered sensibly by complexity. I checked the internal arithmetic of the os1 example and it holds together: the Obese left slope (-1.2797 + 0.5461 = −0.7336) and right slope (-0.7336 + 1.6482 - 0.5351 = 0.3795) match the slope() output to the digits shown, and the intercept and changepoint sums match the text. The numbers are coherent.

My main concern is the validation claim in Example 2. The paper states the aim is "to 'validate' the functions by comparing the results with the true parameter values," and then reports "Estimates of the fixed effects and of the variance components can be compared to the true values … (results not shown here)." The one comparison that is the stated purpose of the example is the one withheld. For a software paper this is the evidence a methodology reader looks for first. The fix is cheap because you already have the object: print the fixed effects and variance components against the fixEff and S arguments used in sim.seg. If instead you prefer to keep the example compact, qualify the word "validate" to a single recovery illustration, because one simulated dataset recovers one draw of the random effects and does not on its own establish that the estimator is unbiased or that its intervals cover. A short repeated-simulation summary over replicates, reporting bias and the coverage of the Wald intervals, would carry the stronger word and is worth adding.

A second, lighter methodological point concerns inference on the changepoint. The summary of os1 shows no t or p value for G0 and no p value for U, while G.groupObese and the slope terms carry full columns. A reader will notice the gap. This is the right place to say in one sentence why Wald inference on the breakpoint location is treated differently, given the non-regularity you already flag at the start of Section 3, and what the confint.segmented.lme method reports instead. The print output also labels the log-restricted-likelihood as "approx", which a methodology reader will want explained: a sentence on what is approximated in the likelihood used for REML, and what that affects in the reported standard errors, would close the loop.

The practical guidance in the Conclusions is sound and matches my experience with these models, that convergence is the binding constraint and a diagonal start is wise. It would help to turn that advice into a short rule tied to the four specifications in Section 2, so a reader who hits non-convergence knows which covariance to drop first. None of these points unsettles the method or the examples. They are about showing the evidence the paper says it has and stating the inferential caveats that a changepoint model carries.

Numbered suggestions to the authors:

1. REQUIRED Display the Example 2 comparison of the estimated fixed effects and variance components against the sim.seg truth (fixEff and S), or change "validate" to a single illustrative recovery so the claim matches what is shown.
2. [editor_fix] Remove the duplicated word in "the the 'plot' method" (Section 2).
3. RECOMMENDED Add one sentence in Section 3.1 explaining why the summary omits t and p for G0 and p for U, linking it to the non-regular inference on the changepoint and to what confint.segmented.lme reports.
4. RECOMMENDED Add a small repeated-simulation summary over replicates reporting bias and Wald-interval coverage for the fixed effects and changepoint, to support the validation claim.
5. RECOMMENDED Add a sentence describing the approximate restricted log-likelihood (the "approx" in the print output) and what the approximation affects in the reported standard errors.
6. RECOMMENDED Turn the Conclusions advice into a short rule tied to the four covariance specifications, naming which variance or covariance to drop first on non-convergence.

Report Reviewer 3, recommends minor

Written by the journal's review engine in the role of the software reviewer, and read by the editor before the decision. How the engine works is on the peer review page.

This is a clear, compact software illustration of segmented.lme and its companion methods. Read as code, the manuscript hangs together well. The internal arithmetic across the printed outputs is consistent: in Example 1 the changepoint estimate (1.724) matches the G0 coefficient, and the right slope (0.3507) equals the left slope plus the U coefficient. In Example 1 with the group interaction, the Obese left slope (-0.7337) and both right slopes recovered from slope(os1) reconstruct exactly from the summary table. In Example 2 the boundary-detection check is a genuine small validation, because which(!attr(oseg1$psi.i, "is.break")) and which(dd$perc.psi %in% c(0,1)) both return subjects 4 and 28. That agreement is the strongest part of the paper from a software standpoint.

The reproducibility gap sits in Example 1. The print output reports "Bootstrap restarting on 6 samples; 1 different solution(s)", which tells the reader the fit involves a random component, yet no seed is set before

the segmented() call. Example 2, by contrast, sets set.seed(123) and passes seed=15 to sim.seg, and the text even says "We set the seeds just to reproduce the results". The two examples should be held to the same standard. A reader who runs the Example 1 block today may see different fixed effects and variance components from those in the paper. Setting a seed (or the appropriate seg.control option that governs the restart draws), or a one-line statement that the restarting results can vary, closes this.

The second gap is the one claim I cannot see supported. The paper states the simulated example serves "to validate the functions by comparing the results with the true parameter values", then reports summary(oseg1) as "results not shown here". The only quantities shown against truth are the left slope (-0.1995 against −0.2) and the boundary detection. For a paper that names validation as an explicit aim, the summary output and a short estimate-versus-truth comparison belong in the text. If that is more than intended, the simpler route is to describe the simulated example as an illustration plus the boundary check, and drop "validate".

On environment: the manuscript pins "version 2.3-0 or higher" and names nlme, but gives no R version, operating system, or exact package versions. For a software article whose value is the exact printed numbers, a session-info block removes ambiguity about what produced them. I note that the artifacts declared at submission were not provided to me, so I have assessed the inline code and its internal consistency and have not run anything. A reader reproducing Example 2 also depends on sim.seg naming its output columns Xdif, Zpsi1, and Zpsi2, which the fit formula assumes. A single head(d) or a note on the column names would make that explicit.

Numbered suggestions to the authors:

1. REQUIRED Make Example 1 reproducible: set a seed (or the seg.control option controlling the bootstrap restarts) before the segmented() call so the printed estimates reproduce, or add one sentence that bootstrap-restarting results can vary across runs.
2. REQUIRED Support the validation aim by including the summary(oseg1) output with an explicit comparison of the estimated fixed effects and variance components to the true simulation values, or narrow "validate" to the boundary-detection check you already show.
3. RECOMMENDED Add a session-info block giving the R version, operating system, and exact versions of segmented and nlme, and pin an exact segmented version for the numeric outputs.
4. RECOMMENDED In Example 2, note or show that sim.seg returns the columns Xdif, Zpsi1, and Zpsi2 that the fit formula references, for instance with head(d), so the block runs without the reader guessing the column names.
5. [editor_fix] Remove the duplicated word in Section 2, "the the 'plot' method".

REVISION, RECEIVED 2026-10-30

Authors confirm each required point as done or give their reason for declining it. Further answers are voluntary, because the review compares every revision with the previous version and the next round's reports say what remains open.

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.

> REQUIRED 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.

Done.

> REQUIRED 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.

Done.

> RECOMMENDED 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.

Declined. it would require a not-so short treatment and I think it is out of aims of the paper

RECOMMENDED 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.

Declined. I do not understand what you mean. Sorry.

OPTIONAL (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.

Declined. I don't have added any "medical comment' on the results in the first example: I don't have the skills to do it

The authors' full response letter, as uploaded, follows as Appendix A.

Round 2: Minor revision

SIGNED DECISION

Confirmed and signed by Matthias Templ on 2026-10-31 · ORCID 0000-0002-8638-5276

The reviewers of the first round read the revision. Both required points of that round are met, one reviewer now recommends acceptance and two ask for a further minor revision. The remaining work is in the code that reproduces the paper. The uploaded script fits the group model from the wrong starting object, two printed functions give different intervals for the same slope, and the package build that produced the printed output is not on file.

DECISION LETTER

Thank you for the revised version of your manuscript. This is round 2, and we are inviting a further minor revision. This round was read against the list you answered, and the editor has read your response against it in full. What follows is only what remains. An item carried over from the previous round says so.

This paper shows, with one real and one simulated dataset, how to fit and visualise segmented mixed models with random changepoints using segmented.lme in the R package segmented. It matters to applied analysts of longitudinal and clustered data who need subject-specific breakpoints estimated in a likelihood framework. We would be glad to see the revised version.

How to read this letter. The decision rests on the items marked required. If a required item seems wrong to you, or cannot be done within this paper, decline it on the revision page and give the reason. A reasoned decline is a legitimate answer, and the editor weighs it. Reviewers can be mistaken. Items marked recommended or optional are advice from the reviewers. Adopt them, decline them with a sentence, or leave them for later work. None of that counts against the paper. The reviewer reports below carry the same marks. The editor who signs this decision may weigh an item differently in the accompanying message, and that message prevails.

Reproducibility items marked required are the journal's minimum for a paper that ships code with its results: code and a command that runs it, and by acceptance a captured computational environment (dependencies/lockfile/container) and a signed determinism attestation for the entry-point command. They document what you already have. They do not ask for new analysis.

REQUIRED (the decision rests on these):

- [consistency-check] The confint(os1) interval for the baseline left slope ("time" row, [−1.6293, −1.0747]) is not centred on the point estimate −1.2796 shown in summary(os1), and it disagrees with slope(os1), which reports [−1.5569, −1.0023] for the same baseline left slope. The breakpoint rows (G0, G.groupObese) in confint do match their point estimates, so the mismatch is specific to the non-breakpoint parameters. Please reconcile the confint output with summary and slope, or add one sentence stating why confint and slope produce different intervals for the same quantity.
- [methods-soundness] The new simulation block that produces Table 1 cannot run as printed. The storage matrices are sized to B with `B=3` (`V=matrix(,B,4)`, `SE<-E<-matrix(,B,7)`), while the loop

is hard-coded `for(i in 1:300)`, so assignment to `V[i,]` fails once `i` exceeds 3 with a subscript-out-of-bounds error. The text states the table rests on 300 runs. Set `B=300` and index the loop as `for(i in 1:B)` so the printed snippet reproduces the table, or state that the manuscript snippet is a reduced illustration and the full run sits in the uploaded script. I did not open Rcode.R, so the script may already carry the corrected form.

- [software-quality] The self-contained script `Rcode.R` fits the group model with the wrong starting object. The line `os1<-segmented(o, time, z.psi= group, x.diff= group, random=subj 1+time+U+G0)` passes `o` (from `lme(phos time, ...)`), but the manuscript builds `os1` from `o1 <- lme(phos time*group, ...)`. Starting from `o` drops the `groupObese` and `time:groupObese` fixed effects, so the script cannot produce the `summary(os1)`, `confint(os1)`, Figure 1 or Figure 2 shown in Section 3.1. Change `o` to `o1` on that line.
- [software-quality] Two divergent copies of `Rcode.R` are shipped (one inside `package+Rscript.zip`, one standalone) and neither reproduces the paper end to end. The zip copy calls `sim.seg(N, nj=12, ...)`, but the argument is `ni` (per `args(sim.seg)` in the paper), so that call errors; it also uses the old `n.plot=c(3,11)` layout. The standalone copy carries the `os1` bug above. Please ship one authoritative script that matches the printed output, or state which file is the reproduction artifact.

RECOMMENDED:

- [software-quality] The package source extracted for this review contained only `Rcode.R`, not the `segmented` help pages or any `tests/` directory, so I cannot confirm that `segmented.lme`, `plot.segmented.lme`, `sim.seg`, `fixef`, `getVarCov`, `slope`, and `breakpoint` are exercised by automated tests. The earlier request (75ae2e1b) asked the paper to state whether these functions are covered by the package test suite and where those tests live; the revised text still does not say. One sentence pointing readers to the tests would close this.
- [software-quality] The reproducibility of the printed estimates rests on `segmented.lme()` deriving a seed from the response values, as the text now states. The script does not set `seg.control(seed=...)` explicitly anywhere, so reproduction depends entirely on that implicit behaviour and on the exact `segmented` version. The `sessionInfo()` now pins `segmented_2.3-0`, which helps; consider setting the seed explicitly in the script so the path is fixed independently of the internal default.
- [statistical-rigor] The new Table 1 supports the recovery claim with bias and a comparison of the mean model-based standard error to the empirical sampling sd over 300 replicates. It does not report the empirical coverage of the Wald intervals, which is the quantity that would directly back the sentence "we can trust the Wald statistic to build confidence intervals." Add the coverage from the replicates already run.
- [statistical-rigor] For $\psi 2$ the mean standard error (0.020) is below the empirical sampling sd (0.025), and the mean estimate (-0.186) sits off the true -0.2. The summary "model-based standard errors quantifying rather appropriately the distribution standard deviations" overstates agreement for that term. Name the terms where the model-based SE understates spread.
- [ethics-and-integrity] The competing-interests point from the prior round [b6d11824] was declined because the request was unclear, so no statement was added. To be plain about what is meant: you wrote and maintain the segmented package that this paper is a tutorial for, and that is a competing interest a reader should see stated in the paper, separate from the maintainer field on CRAN. Add one line, for example "The author develops and maintains the segmented R package", to the Acknowledgements or a Competing Interests section.
- [revision-response-verifier] Carry the unit phrasing into the Figure 2 caption and either pin an exact segmented version in Section 1 or note that "2.3-0 or higher" is intentional for the printed outputs.
- [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. This was asked in the previous round and is still open.

OPTIONAL:

- [visualisation] One small residual, newly visible now that the grid is denser: each of the 33 panels in Figure 1 still repeats the base-plot "time"/"phos" labels inside the panel, while the units sit only on the

outer edges. The panels would read more cleanly with the per-panel "time"/"phos" text suppressed so only the edge labels with units remain. The figure stands as drawn.

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

- [consistency-check] In Section 3.1 the text reads "a less steep left slope (−1.2796 + 0.5461)", but the summary(os1) table gives time:groupObese = 0.5460. The text value should read 0.5460.
- [consistency-check] The Section 1 roadmap states "section 3 reports some analyses on the simulated dataset", but Section 3 contains both Example 1 (real data, glucoseData) and Example 2 (simulated data). The sentence should name both.
- [consistency-check] In Section 3.2 the additional profiles are named "the 6rd, the 5th and the 10th". The subject numbers match the code `plot(oseg1, id=c(4,28,6,5,10), ...)` and Figure 3, so only the ordinal is wrong: "6rd" should read "6th".
- [writing-quality] In Section 3 the sentence "even under the the standard Gaussian assumptions for the ϵ's" repeats "the". This is a duplicated word that survives into the revision, distinct from the Section 2 instance that was fixed.
- [writing-quality] In the Acknowledgements, "Teh author also thanks dr. V. Genoa" should read "The author". A transposition, not an extraction artefact.
- [writing-quality] In Example 2, "Starting the analysis with a simple structure, e.e a diagonal matrix" should read "e.g.".
- [revision-response-verifier] (production) The consistency fix for Example 2 introduced "the 6rd", which should read "6th".

REVIEWS

Report Reviewer 1, recommends minor

Written by the journal's review engine in the role of the application reviewer, and read by the editor before the decision. How the engine works is on the peer review page.

Of the one required point and the recommended and optional points I raised, all are met or declined with a reason I accept. Example 2 no longer carries the validation framing and now shows the parameter recovery directly, through fixef() and getVarCov() tables that place the true values beside the estimates. That is a cleaner answer than the one I asked for and it settles the point. The near-zero random-effect standard deviations are now discussed and tied to the parsimony and model-selection reasoning, the blank U and G0 entries in the summary are explained in the text after the first summary() call, and sessionInfo() is printed. The decline on a clinical reading of the changepoint difference is reasonable: the author is a statistician and the paper does not need a medical interpretation to stand.

One issue is new in this version and I flag it as new. The small simulation study added at the end of Section 3.2 is a genuine improvement and the coverage-style summary in Table 1 is exactly the kind of evidence the recovery claim needed. The printed code for it will not run, though. The matrices V, E and SE are created with B rows where B is set to 3, and the loop then runs for(i in 1:300). In R a matrix does not grow on assignment, so V[i,] fails once i exceeds 3. A reader who copies the block to reproduce Table 1 gets a subscript error. The fix is small: set B=300 and index the loop and the matrices by B. The uploaded source file may already be correct, but the manuscript code is what a reader of a tutorial copies, so the printed snippet should be the one that runs.

Three small textual errors are new and can be settled in proof: "6rd" for "6th" in the Section 3.2 profile list, "Teh author" in the Acknowledgements, and "e.e" for "e.g." in Section 3.2.

Numbered suggestions to the authors:

1. REQUIRED Repair the Section 3.2 simulation snippet so it runs: set B=300 and size V, E and SE and the loop by B (for(i in 1:B)), so the code reproduces Table 1 rather than erroring at the fourth iteration.
2. [editor_fix] Correct "the 6rd" to "the 6th" in Section 3.2 to match the plot call and Figure 3.
3. [editor_fix] Correct "Teh author" to "The author" in the Acknowledgements and "e.e a diagonal matrix" to "e.g. a diagonal matrix" in Section 3.2.

Report Reviewer 2, recommends accept

Written by the journal's review engine in the role of the methodology reviewer, and read by the editor before the decision. How the engine works is on the peer review page.

Of the one required point and five lighter ones I raised, all are met. The validation framing in Example 2 is now "assess", and the example displays the fixed effects against the sim.seg truth via fixef() and the random-effect standard deviations via getVarCov(), which is exactly the evidence the stated purpose needed. The summary output now carries a clear paragraph on why U drops its p-value and why both the t- and p-value are dropped for the changepoint, tied to the non-regular inference. The "approx" log-likelihood is explained with a pointer to Muggeo et al. (2014). The Conclusions give a usable rule for the four covariance specifications, start simple and build up. The new repeated-simulation study (Table 1, 300 replicates) reports means, medians and sampling standard deviations against the model-based standard errors, which answers the bias-and-coverage request in substance even though coverage is not tabulated as a percentage. The decline on residual and random-effects diagnostics is reasoned and reasonable for a tutorial of this length, and I am content to leave it out of scope.

What remains is a single code inconsistency in the new material, which I could not have raised before because the simulation did not exist in the prior version. The loop dimensions V, SE and E with B rows after setting B=3, then iterates over 1:300. As printed it will fail at the fourth iteration with a subscript error, so the code does not reproduce Table 1. The fix is to set B to 300 before the matrices are created, and to confirm the uploaded script carries the same value.

The remaining items are small demonstrable mismatches between the text and the outputs it quotes, which production handles and you confirm in the proof.

Numbered suggestions to the authors:

1. [editor_fix] In the simulation code, set B to 300 before dimensioning V, SE and E, so the loop over 1:300 runs and produces Table 1, and align the uploaded script.
2. [editor_fix] Remove the duplicated word in "even under the the standard Gaussian assumptions" (Section 3).
3. [editor_fix] Change "the 6rd" to "the 6th" (Section 3.2) to match id=c(4,28,6,5,10) and Figure 3.
4. [editor_fix] Reconcile "−1.2796 + 0.5461" in Section 3.1 with the summary coefficient 0.5460.
5. [editor_fix] Change "Teh author" to "The author" in the acknowledgements.

Report Reviewer 3, recommends minor

Written by the journal's review engine in the role of the software reviewer, and read by the editor before the decision. How the engine works is on the peer review page.

All five points from my prior report are met. The Example 1 reproducibility gap is resolved cleanly: the text now explains that segmented.lme() derives its seed from the response values, so the bootstrap-restart path is deterministic, and it points to seg.control(seed=) for an explicit alternative. The validation framing is softened to "assess" and backed by a direct true-versus-estimate table for both fixed effects (via fixef) and the random-effect standard deviations (via getVarCov), which is exactly the comparison I asked for. The sessionInfo() block pins R 4.6.1, the OS, and segmented_2.3-0 and nlme_3.1-169. Example 2 now shows head(d,3) with the Xdif, Zpsi1, and Zpsi2 columns named in the text. The duplicated "the the 'plot' method" is gone.

One new item is material and newly visible in this version. The simulation study added at the end of Section 3.2 is a genuine improvement and the Table 1 recovery looks sound, but the displayed code will not run. V, E, and SE are allocated with matrix(,B,4) and matrix(,B,7) while B is set to 3, yet the loop iterates for(i in 1:300). Assigning V[i,] for i greater than 3 raises "subscript out of bounds" in R, so the block errors at the fourth iteration and never fills the matrices Table 1 reports. The caption and surrounding text both say the results are based on 300 runs, so the intent is clear. Set B=300 and index both the allocations and the loop by B (or 1:B), so the code that produces Table 1 is the code the paper prints. For a software article whose value rests on the printed numbers being the code's output, this one should be correct.

The three small text errors I list as editor_fix all sit in newly added or newly edited sentences and the paper itself fixes their values.

Numbered suggestions to the authors:

1. REQUIRED Make the simulation block run as printed: set B=300 (or allocate V, E, SE with nrow=300) and iterate the loop by 1:B, so the code matches Table 1′s stated 300 replicates.
2. [editor_fix] Correct three text errors in the new material: “under the the standard Gaussian assumptions” (duplicated “the”), “Teh author” in the Acknowledgements, and “the 6rd” in Section 3.2 (should read “6th”).

REVISION, RECEIVED 2026-11-03

Authors confirm each required point as done or give their reason for declining it. Further answers are voluntary, because the review compares every revision with the previous version and the next round’s reports say what remains open.

> REQUIRED The confint(os1) interval for the baseline left slope (“time” row, [−1.6293, −1.0747]) is not centred on the point estimate −1.2796 shown in summary(os1), and it disagrees with slope(os1), which reports [−1.5569, −1.0023] for the same baseline left slope. The breakpoint rows (G0, G.groupObese) in confint do match their point estimates, so the mismatch is specific to the non-breakpoint parameters. Please reconcile the confint output with summary and slope, or add one sentence stating why confint and slope produce different intervals for the same quantity.

Done. Thanks for noting that. I have corrected the function

> REQUIRED The new simulation block that produces Table 1 cannot run as printed. The storage matrices are sized to B with `B=3` (`V=matrix(,B,4)`, `SE<-E<-matrix(,B,7)`), while the loop is hard-coded `for(i in 1:300)`, so assignment to `V[i,]` fails once `i` exceeds 3 with a subscript-out-of-bounds error. The text states the table rests on 300 runs. Set `B=300` and index the loop as `for(i in 1:B)` so the printed snippet reproduces the table, or state that the manuscript snippet is a reduced illustration and the full run sits in the uploaded script. I did not open Rcode.R, so the script may already carry the corrected form.

Done. corrected, thanks.

> REQUIRED The self-contained script `Rcode.R` fits the group model with the wrong starting object. The line `os1<-segmented(o, time, z.psi= group, x.diff= group, random=subj 1+time+U+G0)` passes `o` (from `lme(phos time, ...)`), but the manuscript builds `os1` from `o1 <- lme(phos time*group, ...)`. Starting from `o` drops the `groupObese` and `time:groupObese` fixed effects, so the script cannot produce the `summary(os1)`, `confint(os1)`, Figure 1 or Figure 2 shown in Section 3.1. Change `o` to `o1` on that line.

Done. Yes, of course. Thanks for noting that. I have corrected that

> REQUIRED Two divergent copies of `Rcode.R` are shipped (one inside `package+Rscript.zip`, one standalone) and neither reproduces the paper end to end. The zip copy calls `sim.seg(N, nj=12, ...)`, but the argument is `ni` (per `args(sim.seg)` in the paper), so that call errors; it also uses the old `n.plot=c(3,11)` layout. The standalone copy carries the `os1` bug above. Please ship one authoritative script that matches the printed output, or state which file is the reproduction artifact.

Done. I’m sorry for the inconvenience. I got confused. I am uploading the file Rcode+package.zip including i)the new version of the package segmented_2.3-0.tar.gz, and ii)the Rcode.R to replicate the analyses in the paper

> RECOMMENDED The package source extracted for this review contained only `Rcode.R`, not the `segmented` help pages or any `tests/` directory, so I cannot confirm that `segmented.lme`, `plot.segmented.lme`, `sim.seg`, `fixef`, `getVarCov`, `slope`, and `breakpoint` are exercised by automated tests. The earlier request (75ae2e1b) asked the paper to state whether these functions are covered by the package test suite and where those tests live; the revised text still does not say. One sentence pointing readers to the tests would close this.

Done. I apologize for the inconvenience. I got confused. I am uploading the file Rcode+package.zip including i)the new version of the package segmented_2.3-0.tar.gz, and ii)the Rcode.R to replicate the analyses in the paper

> RECOMMENDED The reproducibility of the printed estimates rests on `segmented.lme()` deriving a seed from the response values, as the text now states. The script does not set `seg.control(seed=...)` explicitly anywhere, so reproduction depends entirely on that implicit behaviour and on the exact `segmented` version. The `sessionInfo()` now pins `segmented_2.3-0`, which helps; consider setting the seed explicitly in the script so the path is fixed independently of the internal default.

Declined. I have commented in the text the behaviour of segmented and the possibility of using the ‘seg.control(seed=..)’. I think it is enough. I hope you are fine with my choice

RECOMMENDED The new Table 1 supports the recovery claim with bias and a comparison of the mean model-based standard error to the empirical sampling sd over 300 replicates. It does not report the empirical coverage of the Wald intervals, which is the quantity that would directly back the sentence “we can trust the Wald statistic to build confidence intervals.” Add the coverage from the replicates already run.

Done. Table 1 now includes the coverage levels (last column)

RECOMMENDED For $\kappa 2$ the mean standard error (0.020) is below the empirical sampling sd (0.025), and the mean estimate (-0.186) sits off the true −0.2. The summary “model-based standard errors quantifying rather appropriately the distribution standard deviations” overstates agreement for that term. Name the terms where the model-based SE understates spread.

Done. I have modified the comments, by underlining the behaviour for kappa1 (and kappa2 too).

RECOMMENDED The competing-interests point from the prior round [b6d11824] was declined because the request was unclear, so no statement was added. To be plain about what is meant: you wrote and maintain the segmented package that this paper is a tutorial for, and that is a competing interest a reader should see stated in the paper, separate from the maintainer field on CRAN. Add one line, for example “The author develops and maintains the segmented R package”, to the Acknowledgements or a Competing Interests section.

Done. Added in the Acknowledgements

RECOMMENDED 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. This was asked in the previous round and is still open.

Declined. Yes, it was asked, and in fact I followed such suggestion by specifying the labels in the x- and y-axis (by also adding the units as you suggested) of Figure 1. In Figure 2 , in order to keep simple the relevant code, the labels are the variable names.

OPTIONAL One small residual, newly visible now that the grid is denser: each of the 33 panels in Figure 1 still repeats the base-plot “time”/“phos” labels inside the panel, while the units sit only on the outer edges. The panels would read more cleanly with the per-panel “time”/“phos” text suppressed so only the edge labels with units remain. The figure stands as drawn.

Declined. I am sorry, but I see no “time”/“phos” labels inside the 33 panels in Figure 1.

RECOMMENDED 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.

Done. I am uploading the file Rcode+package.zip including i)the new version of the package segmented_2.3-0.tar.gz, and ii)the Rcode.R to replicate the analyses in the paper

RECOMMENDED 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.

Done. In the caption of Figure 2

RECOMMENDED 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.

Done. At the end of section 3, I report the output of sessionInfo()

Round 3: Minor revision

SIGNED DECISION

Confirmed and signed by Matthias Templ on 2026-11-03 · ORCID 0000-0002-8638-5276

The editor re-weighted 2 points.

DECISION LETTER

Thank you for the revised version of your manuscript. This is round 3, and we are inviting a further minor revision. This round was read against the list you answered, and the editor has read your response against it in full. What follows is only what remains. An item carried over from the previous round says so.

This paper shows, with one real and one simulated dataset, how to fit and visualise segmented mixed models with random changepoints using segmented.lme in the R package segmented. It matters to applied analysts of longitudinal and clustered data who need subject-specific breakpoints estimated in a likelihood framework. We would be glad to see the revised version.

How to read this letter. The decision rests on the items marked required. If a required item seems wrong to you, or cannot be done within this paper, decline it on the revision page and give the reason. A reasoned decline is a legitimate answer, and the editor weighs it. Reviewers can be mistaken. Items marked recommended or optional are advice from the reviewers. Adopt them, decline them with a sentence, or leave them for later work. None of that counts against the paper. The reviewer reports below carry the same marks. The editor who signs this decision may weigh an item differently in the accompanying message, and that message prevails.

RECOMMENDED:

- [software-quality] The submission still carries two stale copies of `Rcode.R` that do not match the paper: `upload: Rcode.R` (standalone) still fits `os1<-segmented(o, ...)` with the wrong starting object the prior round flagged, and `package+Rscript.zip: Rcode.R` still calls `sim.seg(N, nj=12, ...)` with the wrong argument `nj` and the old `n.plot=c(3,11)` layout. The author has named `Rcode+package.zip` as the reproduction artifact, which resolves the ambiguity, but removing or clearly marking the two superseded copies would stop a reader from running the wrong file.
- [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. This was asked in the previous round and is still open.

OPTIONAL:

- [software-quality] The declared reproduction artifact `Rcode+package.zip: Rcode.R` sets the simulation covariance as `S=diag(c(.12,.05,.3,.02)^2)` in the small simulation loop, so the breakpoint random-effect sd is 0.02. The paper's printed snippet uses `S=diag(c(.12,.05,.3,.1)^2)` and Table 1 reports σk true 0.10 with mean 0.090 and sd 0.011. A run with σk=0.02 cannot produce a mean of 0.090, so the uploaded authoritative script would not reproduce the σk row of Table 1. Set the fourth variance in `S` to `.1` in the uploaded script, or correct the Table 1 σk entries, so the script and the table agree.

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

- [consistency-check] The Section 3.1 text still reads "a less steep left slope (−1.2796 + 0.5461)", while summary(os1) gives time:groupObese = 0.5460. The text should read 0.5460. This was raised last round (faa3a12a) and the manuscript still carries 0.5461.
- [consistency-check] The Section 1 roadmap still states "section 3 reports some analyses on the simulated dataset", but Section 3 contains Example 1 (real glucoseData) and Example 2 (simulated data). The sentence should name both. Raised last round (5436d34f) and still open.
- [consistency-check] In Section 3.2 the profiles are still named "the 6rd, the 5th and the 10th". The code plot(oseg1, id=c(4,28,6,5,10)) and Figure 3 confirm subject 6, so the ordinal should read "6th". Raised last round (e3ab272b/f99c9341) and still present.
- [methods-soundness] In Section 3.2 the phrase "values quite close to 0.95%" carries a stray percent sign; the nominal level is 0.95 or 95%. The paper itself determines the correct value.
- [revision-response-verifier] Several production corrections from the prior round remain unmade in the text, the "0.5461" value (should be 0.5460), the Section 1 roadmap naming both examples, "6rd"

(should be 6th), the duplicated "the", and "e.e" (should be "e.g."), and should be carried to the proof list.
- [Reviewer 1] Section 3.2 still reads "e.e a diagonal matrix"; this should read "e.g. a diagonal matrix".
- [Reviewer 2] Section 3 still reads "even under the the standard Gaussian assumptions"; the duplicated "the" remains.
- [revision-response-verifier] (production) The consistency fix for Example 2 introduced "the 6rd", which should read "6th". This was asked in the previous round and is still open.

REVIEWS

Report Reviewer 1, recommends accept

Written by the journal's review engine in the role of the application reviewer, and read by the editor before the decision. How the engine works is on the peer review page.

The one required point is met. The simulation snippet at the end of Section 3.2 now runs: the loop is for(i in 1:B) and V, E and SE are sized by B, so the subscript error at i=4 is gone. The author kept B=3 in the printed block for speed and states in the text that Table 1 rests on 300 runs rather than 3. That is a sound choice for a tutorial, the printed code executes without error, and a reader who wants the table is told to raise B. The uploaded Rcode.R and segmented_2.3-0 archive settle the reproducibility artifact as well. I am content with the resolution.

Two of the three textual corrections I flagged remain. The Acknowledgements now read "The author". The profile-list typo "6rd" and the "e.e" abbreviation in Section 3.2 are still present in this version. Both are editor_fix items whose correct values the text itself fixes, so they belong on the proof production list rather than holding the paper.

Numbered suggestions to the authors:

1. [editor_fix] Correct "the 6rd" to "the 6th" in Section 3.2 to match the plot call and Figure 3.
2. [editor_fix] Correct "e.e a diagonal matrix" to "e.g. a diagonal matrix" in Section 3.2.

Report Reviewer 2, recommends accept

Written by the journal's review engine in the role of the methodology reviewer, and read by the editor before the decision. How the engine works is on the peer review page.

Of the five production points I raised last round, two are settled and three remain. The acknowledgements now read "The author", so that one is done. The simulation code no longer crashes: the loop runs over 1:B rather than 1:300, so the subscript error is gone, and the text now records that Table 1 rests on 300 runs while the printed snippet keeps B=3 for speed. That is a reasonable way to leave it, and with the uploaded Rcode.R carrying the full run I am content. A reader who copies the printed snippet verbatim will get a 3-run table rather than the one shown, but the text says so.

Three small demonstrable mismatches between the text and its own outputs are still in the manuscript: the duplicated "the the" in Section 3, "6rd" where "6th" is meant in Section 3.2, and "0.5461" in Section 3.1 against the table value 0.5460. None of these touch a claim or a result. They are production fixes you confirm in the proof, and they do not hold the paper. The methodology, the worked examples, the summary output explanation for the dropped p-values, and the simulation with coverage levels in Table 1 all stand as accepted last round.

Numbered suggestions to the authors:

1. [editor_fix] Remove the duplicated word in "even under the the standard Gaussian assumptions" (Section 3).
2. [editor_fix] Change "the 6rd" to "the 6th" (Section 3.2) to match id=c(4,28,6,5,10) and Figure 3.
3. [editor_fix] Reconcile "−1.2796 + 0.5461" in Section 3.1 with the summary coefficient 0.5460.

Report Reviewer 3, recommends accept

Written by the journal's review engine in the role of the software reviewer, and read by the editor before the decision. How the engine works is on the peer review page.

The required point is met. The simulation block is now internally consistent and runs as printed: B is set to 3, V, E and SE are sized to B, and the loop iterates 1:B, so there is no subscript error. The author

resolved the 3-versus-300 mismatch by keeping the quick B=3 demo in the printed listing and stating in the surrounding text that Table 1 is based on 300 runs. A reader reproducing Table 1 sets B=300, and the provided Rcode.R carries the full run. For a software illustration this is a reasonable resolution, and the defect that made the block error out is gone.

Two of the three editor_fix text errors remain in the current version. Section 3 still reads “even under the the standard Gaussian assumptions”, and Section 3.2 still reads “the 6rd” where “6th” is meant. Both carry values the paper itself fixes, so they belong on the production list for the proof. “Teh author” is corrected.

The seed decline is fine. Commenting the segmented.lme() seed behaviour and pointing to seg.control(seed=) in the text covers the replicability question without a code change, and the Example 1 narrative already explains the response-derived seed.

Numbered suggestions to the authors:

1. [editor_fix] Correct the two remaining text errors for the proof: “under the the standard Gaussian assumptions” (duplicated “the”) in Section 3, and “the 6rd” (should read “6th”) in Section 3.2.

REVISION, RECEIVED 2026-11-06

Authors confirm each required point as done or give their reason for declining it. Further answers are voluntary, because the review compares every revision with the previous version and the next round’s reports say what remains open.

> RECOMMENDED The submission still carries two stale copies of `Rcode.R` that do not match the paper: `upload: Rcode.R` (standalone) still fits `os1<-segmented(o, ...)` with the wrong starting object the prior round flagged, and `package+Rscript.zip: Rcode.R` still calls `sim.seg(N, nj=12, ...)` with the wrong argument `nj` and the old `n.plot=c(3,11)` layout. The author has named `Rcode+package.zip` as the reproduction artifact, which resolves the ambiguity, but removing or clearly marking the two superseded copies would stop a reader from running the wrong file.
>
> **Declined.** You clearly refer to the Rcode.R of the 1st submission. I really do not understand. I really can’t see multiple copies of Rcode.R. When I go to submit a revision, I upload the zip file (paperJReprStatOK.zip) including the paper source files and a zip file (Rcode+package.zip) including the new Rcode.R along with the (updated) segmented package. I can’t see (and can’t cancel) the previous The other Rcode.R

> RECOMMENDED 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.
>
> **Declined.** I do not understand. When I go for a new the revision, I upload the zip file (paperJReprStatOK.zip) including the paper source files and a zip file (Rcode+package.zip) including the new Rcode.R along with the (updated) segmented package. Note that referee 1 writes “The uploaded Rcode.R and segmented_2.3-0 archive settle the reproducibility artifact as well. I am content with the resolution”. Do it seems (s)he is able to run the example.

> OPTIONAL The declared reproduction artifact `Rcode+package.zip: Rcode.R` sets the simulation covariance as `S=diag(c(.12,.05,.3,.02)^2)` in the small simulation loop, so the breakpoint random-effect sd is 0.02. The paper’s printed snippet uses `S=diag(c(.12,.05,.3,.1)^2)` and Table 1 reports σk true 0.10 with mean 0.090 and sd 0.011. A run with σk=0.02 cannot produce a mean of 0.090, so the uploaded authoritative script would not reproduce the σk row of Table 1. Set the fourth variance in `S` to `.1` in the uploaded script, or correct the Table 1 σk entries, so the script and the table agree.
>
> **Done.** done, thanks for reporting that.

> RECOMMENDED The package source extracted for this review contained only `Rcode.R`, not the `segmented` help pages or any `tests/` directory, so I cannot confirm that `segmented.lme`, `plot.segmented.lme`, `sim.seg`, `fixef`, `getVarCov`, `slope`, and `breakpoint` are exercised by automated tests. The earlier request (75ae2e1b) asked the paper to state whether these functions are covered by the package test suite and where those tests live; the revised text still does not say. One sentence pointing readers to the tests would close this.
>
> **Declined.** I do not understand. When I go for a new the revision, I upload the zip file (paperJReprStatOK.zip) including the paper source files and a zip file (Rcode+package.zip) including

> the new Rcode.R along with the (updated) segmented package. Note that referee 1 writes "The uploaded Rcode.R and segmented_2.3-0 archive settle the reproducibility artifact as well. I am content with the resolution". Do it seems (s)he is able to run the example.

Round 4: Accepted

SIGNED DECISION

Signed by Matthias Templ on 2026-11-18 · ORCID 0000-0002-8638-5276

Accepted. All three reviewers recommend acceptance. The author corrected the simulation so that each replicate draws its own data set, and the journal ran his code with his package: Table 1 reproduces exactly from 300 replicates, and the two examples agree.

DECISION LETTER

Congratulations, your manuscript has been accepted for publication in JRepStat. Production begins now.

Where it is not on file yet, we still need the following for production.

- a valid ORCID iD for every author
- a captured computational environment (dependencies/lockfile/container)
- a signed determinism attestation for the entry-point command

REVIEWS

Report Reviewer 1, recommends accept

Written by the journal's review engine in the role of the application reviewer, and read by the editor before the decision. How the engine works is on the peer review page.

Both points I raised are met. Section 3.2 now reads "(the 6th, the 5th and the 10th, say)", matching plot(oseg1, id=c(4,28,6,5,10)) and Figure 3, and the diagonal-matrix sentence now reads "e.g. a diagonal matrix". Nothing from my report is still open.

The three reproducibility points the author declined belong to other referees, not to me, and the author's lines there report confusion about which Rcode.R copy is attached rather than a refusal to supply the artifact. On the version in front of me the uploaded Rcode.R and segmented_2.3-0 archive run as I reported last round, so from the application and reproducibility angle I am satisfied. If the submission system is carrying a stale Rcode.R from the first round alongside the current one, that is a matter for the editorial office to resolve at acceptance so the recorded artifact is the one that matches this manuscript.

Numbered suggestions to the authors: (none)

Report Reviewer 2, recommends accept

Written by the journal's review engine in the role of the methodology reviewer, and read by the editor before the decision. How the engine works is on the peer review page.

All three points I left open are settled in this version. Section 3 now reads "even under the standard Gaussian assumptions". Section 3.2 reads "the 6th, the 5th and the 10th", matching id=c(4,28,6,5,10) and Figure 3. Section 3.1 now reads "(−1.2796 + 0.5460)", in agreement with the summary coefficient 0.5460.

The three reproducibility points the author declined were raised by another reviewer, not by me, and concern the upload mechanics for Rcode.R and the segmented archive rather than any methodology question. They do not bear on my assessment. The methodology, the worked examples, the summary-output explanation for the dropped p-values, and the simulation with coverage levels in Table 1 all stand as accepted.

Nothing is open on my lens. I recommend acceptance.

Numbered suggestions to the authors:

Report Reviewer 3, recommends accept

Written by the journal's review engine in the role of the software reviewer, and read by the editor before the decision. How the engine works is on the peer review page.

The one item I left open is met. Section 3 now reads "even under the standard Gaussian assumptions for the ϵ's" with the duplicated "the" removed, and Section 3.2 now reads "(the 6th, the 5th and the 10th, say)", which matches the plot call id=c(4,28,6,5,10). Both production corrections are in the current version.

On the reproducibility declines the author recorded: the response explains that each revision upload replaces the prior Rcode.R inside Rcode+package.zip, and that an earlier reference to multiple copies reflected a stale artifact the author cannot see or remove on their end. That reading is consistent with how the submission system versions attachments, and referee 1 confirms the uploaded Rcode.R and segmented_2.3-0 archive run. Those points were recommended, and the explanation convinces me, so I take them as settled.

Numbered suggestions to the authors: (none)

APPENDIX A

The authors' full response letter to round 1

As uploaded with the revision received 2026-10-30. Its own pages follow.

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.