Authors' Response

We would like to thank the Special Issue Editors and the Reviewers for their time and effort in providing very constructive and helpful comments that have led to improving clarity and quality of the paper. We have tried to incorporate all the suggestions, and present here the point-by-point response to the reviewers' comments and questions.

Reviewer 1

I encourage the authors to share their R code with readers/community on a Github or similar cloud site (or as an R library) for this reason, in addition to indicating they can be contacted for such supplementary material.

Thank you very much for your comments and suggestions. We prepared an R function `mcusum.test` for the developed bootstrap testing procedure that we shared within R package *funtimes* via the Comprehensive R Archive Network (CRAN). With the suggestion of the Reviewer 2 on using kernel smoothing in the sieve bootstrap, we added this option in version 6.1 of the package (Lyubchich and Gel, 2019). We changed the data statement in the manuscript as follows (lines 509–511):

"The code for the method is available from `mcusum.test` function in R package *funtimes* (Lyubchich and Gel, 2019); the data that support the findings of this study are available from the corresponding author upon request."

Reviewer 2

1. I'm confused about the number of changepoints: is it exactly equal to m or in $\{0, 1, \ldots, m\}$? The quantity $K_{T,m}$ would indicate the latter; however, the test statistic in (5) and (6) always has m changepoints. This needs to be clarified. Does the stat M_T in (6) for ℓ changepoints increase in ℓ if the changepoint configurations are subsets of one and other? What am I missing?

In response to this comment and Comment 6, we clarified the tested hypotheses (lines 109–111):

"The H_0 of no changes in the coefficients is rejected in all cases from at least one change point to at most m change points being present, where the m change point locations are specified before applying the test."

We also added the following clarification about (5) and (6) (lines 121–129):

"The statistic M_T is the maximum of statistics M calculated over all combinations of one to m change points $k_1, \ldots, k_m$ in $\mathfrak{M}$. For example, for given $m = 3$ and respective change point locations $\mathfrak{M} = \{1 \leqslant k_1 \leqslant k_2 \leqslant k_3 < T\}$, the seven possible combinations of the change point locations are explored:

$$\mathfrak{M}^* = \{(k_1), (k_2), (k_3), (k_1, k_2), (k_1, k_3), (k_2, k_3), (k_1, k_2, k_3)\},$$
$$M_T = \max\{M(\mathfrak{M}_1^*), \ldots, M(\mathfrak{M}_7^*)\},$$

and the final change point locations are those corresponding to the M_T:

$$\arg\max_{\mathfrak{M}^*}\{M(\mathfrak{M}_1^*), \ldots, M(\mathfrak{M}_7^*)\}."$$

2. One of the major findings in the paper is that the methods have better power than those that adjust CUSUM limit laws for correlation. This is surely because the authors are using the v_t terms, which are the one-step-ahead predictions and are independent, rather than the correlated ϵ_t terms. This issue is key for positively correlated series and is discussed in detail in Robbins et al (2011). I suggest reiterating it here as its important.

We added the following discussion in Section 2.1 (lines 205–214):

"The residuals $\hat{v}_t$ calculated at Step 3 of our procedure are the differences between the potentially autocorrelated observed values $\hat{\varepsilon}_t$ and the one-step-ahead predictions obtained using the autoregressive model (3). That is, $\hat{v}_t$ are one-step-ahead prediction residuals. Robbins et al (2011) provide a summary of using such residuals in CUSUM methods when considering a problem of detecting a single mean shift in an autoregressive moving average (ARMA) process. Robbins et al (2011) note that under certain regularity conditions the CUSUM-based inferences are asymptotically the same, whether based on the raw ARMA series or the uncorrelated one-step-ahead prediction residuals. In our sieve bootstrap approach, however, the bootstrap counterparts of residuals $\hat{v}_t$ are plugged back into the autoregressive model to generate new autoregressive series $\hat{\varepsilon}_t^*$."

3. I think a bit more on a comparative discussion to penalized likelihood techniques like the MDL is warranted. One could easily penalize the CUSUM terms in (5) for each segment as a means of selecting the number of changepoints.

We added the following Remark 1 (lines 133–142):

"An alternative to trying all combinations of change points in $\mathfrak{M}^*$ could be an approach utilizing penalized likelihood, such as Akaike information criterion (AIC), Bayesian information criterion (BIC), or risk inflation criterion (RIC); see Stine (2004) and references therein for comparative assessment of these criteria in variable selection. The approach using criterion-based stepwise elimination of terms, starting from the total of $m+1$ terms (one for each subperiod), may be faster than using the $\mathfrak{M}^*$, however, both these approaches require the candidate change point locations $k_1, \ldots, k_m$ to start with. If $k_1, \ldots, k_m$ are unknown, they can be pre-selected from the data, which usually constitutes a more computationally demanding task (see Section 2.2) than exploring the combinations of m terms in $\mathfrak{M}^*$."

4. There are a lot of article abuses and other faux pas with short words in the paper that need to be fixed. I think it is The Chesapeake Bay, for example.

We reviewed the text and corrected the typos.

5. The bottom 3 equations on page 5 can be written as one equation over all $m+1$ regimes when the appropriate boundary conditions are made.

Due to the difference in standardization for the subperiods in the ends and in the center of the time series $\hat{\varepsilon}_t$, and also to maintain a closer correspondence with the original paper of Horváth et al (2017), we would prefer to keep the separate equations.

6. The notation in (4) is bad: $\beta^{(j)}$ is the jth regime's regression parameters, which to this point have not been defined, not the β vector at time j. By the way, I don't think the alternative conveys that there are at most m changepoints. Rather, it says there is at least one changepoint.

We defined subperiods (regimes) and added an explanation on the number of change points that lead to rejection of the null hypothesis (lines 104–111; $\boldsymbol{\beta}_t$ has been also defined in the manuscript, in response to Comment 12, lines 91–93):

"Our testing procedure is based on the flexible framework for detecting at-most-m changes, by Horváth et al (2017):

$$\begin{aligned} &H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \ldots = \boldsymbol{\beta}_T \\ &H_a : \boldsymbol{\beta}^{(j)} \neq \boldsymbol{\beta}^{(l)} \text{ for some } 1 \leqslant j, l \leqslant m+1, \end{aligned} \tag{1}$$

where $\boldsymbol{\beta}^{(i)}$ $(i = 1, \ldots, m+1)$ are regression coefficients in the ith subperiod; $k_1, \ldots, k_m$ are the locations of m change points $(1 \leqslant k_1 \leqslant k_2 \leqslant \cdots \leqslant k_m < T)$; $k_{i-1} < t \leqslant k_i$ ($k_0 = 0$ and $k_{m+1} = T$). The H_0 of no changes in the coefficients is rejected in all cases from at least one change point to at most m change points being present, where the m change point locations are specified before applying the test."

7. Please specify $\hat{\phi}_1, \ldots, \hat{\phi}_p$ rather than $\hat{\phi}$ in the steps on page 6.

We replaced the two instances of $\hat{\phi}$ with $\hat{\phi}_1, \ldots, \hat{\phi}_p$ (lines 149 and 157).

8. I would have generated the $\hat{v}_t^*$ from a kernel smoothing of the $\hat{v}_t$ for extra mixing.

Thank you for the suggestion. We added the option of kernel smoothing in the updated version of the function `mcusum.test` in R package *funtimes* (Lyubchich and Gel, 2019) and modified the text as follows (lines 155–156):

"Let $\hat{v}_t^*$ be a sample with replacement (i.e., bootstrap sample) from $\hat{v}_t$. Alternatively, $\hat{v}_t^*$ can be generated from a kernel smoothing of the $\hat{v}_t$."

In an additional short simulation study, however, the test with smoothing showed a deteriorated performance for the normal case (see Table A; cf. Table 1 in the manuscript).

Table A: Empirical size of the bootstrapped test under different specifications of the models and error processes ($\delta = 0$, $\alpha = 0.05$, MC = 1000, Gaussian kernel smoothing applied)

		Error process ε_t		
Model	T	i.i.d. $N(0,1)$	GARCH(1,1)	AR(1)
Model I, $m = 1$ and $\theta = 0.5$	30	0.030	0.042	0.121
	100	0.028	0.050	0.062
	400	0.051	0.048	0.049

9. Top of page 5: the OLS estimators are not only unbiased under correlated errors, they are also usually asymptotically most efficient: Lee and Lund (2004) and the references within.

Thank you for the reference. We modified the text in parentheses as follows (lines 96–99):

"We assume possible autoregressive dependence in the errors (under which the OLS estimates are still unbiased; see Lee and Lund, 2004 and references therein on asymptotic efficiency of the OLS estimates under common cases of autocorrelated regression errors)..."

10. The dynamic programming reference on line 168 seems like a bum steer: one cannot use this technique because the time series parameters apply to all regimes and are hence estimated from data from all regimes.

We would prefer to keep the reference for the readers to be familiar with the broad selection of relevant methods. We have modified the text as follows (lines 197–200):

"Other available approaches use the principle of dynamic programming (see Zeileis et al, 2003; Antoch and Jarušková, 2013, and references therein) and genetic algorithms (e.g., Li and Lund, 2012) to find a solution optimizing some likelihood function, which is usually the residual sum of squares, AIC, or BIC."

11. What's with the GARCH simulations in an environmental study? This seems misplaced.

This comment seems to be incomplete. To further clarify: we did not apply GARCH models in the Chesapeake Bay case study because of the type of the data and relatively small sample size for this type of models.

12. Line 90. Design vector (not matrix). And don't you need a transpose on x_t in the equations?

We used Section 3-Appendix in Chatterjee and Hadi (2006) to revise the matrix notations in model (1) and specified each element in the text explicitly (such format of the design matrix is convenient because it is close to data representation in statistical software, such as `data.frame` structure in R, and transpose is not needed; lines 89–94):

"

$$\mathbf{Y}_t = \mathbf{X}_t\boldsymbol{\beta}_t + \boldsymbol{\varepsilon}_t,$$

where $t = 1, \ldots, T$; T is the sample size (i.e., length of the time series); $\mathbf{Y}_t$ is the dependent variable, $\mathbf{X}_t$ is the design matrix with d regressors; $\boldsymbol{\beta}_t$ is a vector of regression coefficients, and $\boldsymbol{\varepsilon}_t$ are regression errors:

$$\mathbf{Y}_t = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix}, \quad \mathbf{X}_t = \begin{bmatrix} x_{10} & x_{11} & \ldots & x_{1d} \\ x_{20} & x_{21} & \ldots & x_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ x_{T0} & x_{T1} & \ldots & x_{Td} \end{bmatrix}, \quad \boldsymbol{\beta}_t = \begin{bmatrix} \beta_{0t} \\ \beta_{1t} \\ \vdots \\ \beta_{dt} \end{bmatrix}, \quad \boldsymbol{\varepsilon}_t = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{bmatrix},$$

and $x_{t0} = 1$ for all t."

References

Antoch J, Jarušková D (2013) Testing for multiple change points. Computational Statistics 28(5):2161–2183, DOI 10.1007/s00180-013-0401-1

Chatterjee S, Hadi AS (2006) Regression Analysis by Example. John Wiley & Sons, Hoboken

Horváth L, Pouliot W, Wang S (2017) Detecting at-most-m changes in linear regression models. Journal of Time Series Analysis 38:552–590, DOI 10.1111/jtsa.12228

Lee J, Lund RB (2004) Revisiting simple linear regression with autocorrelated errors. Biometrika 91(1):240–245, DOI 10.1093/biomet/91.1.240

Li S, Lund RB (2012) Multiple changepoint detection via genetic algorithms. Journal of Climate 25(2):674–686, DOI 10.1175/2011JCLI4055.1

Lyubchich V, Gel YR (2019) funtimes: Functions for Time Series Analysis. URL https://CRAN.R-project.org/package=funtimes, R package version 6.1

Robbins M, Gallagher C, Lund RB, Aue A (2011) Mean shift testing in correlated data. Journal of Time Series Analysis 32(5):498–511, DOI 10.1111/j.1467-9892.2010.00707.x

Stine RA (2004) Model selection using information theory and the MDL principle. Sociological Methods & Research 33(2):230–260, DOI 10.1177/0049124103262064

Zeileis A, Kleiber C, Krämer W, Hornik K (2003) Testing and dating of structural changes in practice. Computational Statistics & Data Analysis 44(1-2):109–123, DOI 10.1016/S0167-9473(03)00030-6