JOURNAL OF COMPUTATIONAL AND GRAPHICAL STATISTICS
2022, VOL. 31, NO. 3, 753–765
https://doi.org/10.1080/10618600.2021.2000873

Graphical Influence Diagnostics for Changepoint Models

Ines Wilms[a], Rebecca Killick[b], and David S. Matteson[c]

[a]Department of Quantitative Economics, Maastricht University, Maastricht, Netherlands; [b]Department of Mathematics and Statistics, Lancaster University, Lancaster, UK; [c]Department of Statistics and Data Science, Cornell University, Ithaca, NY

ABSTRACT

Changepoint models enjoy a wide appeal in a variety of disciplines to model the heterogeneity of ordered data. Graphical influence diagnostics to characterize the influence of single observations on changepoint models are, however, lacking. We address this gap by developing a framework for investigating instabilities in changepoint segmentations and assessing the influence of single observations on various outputs of a changepoint analysis. We construct graphical diagnostic plots that allow practitioners to assess whether instabilities occur; how and where they occur; and to detect influential individual observations triggering instability. We analyze well-log data to illustrate how such influence diagnostic plots can be used in practice to reveal features of the data that may otherwise remain hidden. Supplementary materials for this article are available online.

ARTICLE HISTORY
Received February 2021
Revised July 2021

KEYWORDS
Change point; Influential data; Segmentation; Statistical graphics; Structural change; Visual diagnostics

1. Introduction

Detecting changes in the distributional properties of data are a common problem that arises in many application areas including; anomaly detection (e.g., Rubin-Delanchy, Lawson, and Heard 2016), bioinformatics (Erdman and Emerson 2008), economics (Spokoiny 2009), genetics (Hocking et al. 2013), network traffic analysis (Kwon et al. 2006), and oceanography (Leeson et al. 2017). The first work on such changepoint problems date back to Page (1954). Since then, changepoint models have been actively investigated (see Eckley, Fearnhead, and Killick 2011 for an overview) with most studies focusing on the development of changepoint detection algorithms (e.g., the PELT method of Killick, Fearnhead, and Eckley 2012 or wild binary segmentation of Fryzlewicz 2014) and inference methods (e.g., Wu and Matteson 2020 for univariate changepoint detection in the presence of local outliers, Grundy, Killick, and Mihaylov 2020 for a recent multivariate change in mean and variance method).

Recent literature is starting to consider issues that arise when applying offline changepoint techniques in practice such as the impact of where changepoints are incorporated into the inference pipeline; pre-modelling as a data cleaning process, within the main modelling framework or as a post-modelling diagnostic on residuals from a fitted model (Chapman and Killick 2020). However, influence diagnostics—as an integral part of any data analysis—have been overlooked for offline changepoint analyses. Yet, diagnostic work is vital to enable analysts and practitioners to detect (i) potential problems with the changepoint model, (ii) how and where they occur and (iii) what may trigger these to occur. Providing such diagnostic tools is crucial to ensure the potential of changepoint models to be fully realized outside of the academic domain and to help practitioners develop intuition, discover features of the data that may otherwise remain hidden and make, in the end, more informed decisions (Rajaratnam et al. 2019).

In this article, we present a unified influence framework for offline changepoint models that is fully aligned with this articulated need. The graphical influence diagnostic tools we develop are the first to highlight instabilities in changepoint models and assess the influence of single observations on the stability of the changepoint segmentation and corresponding segment parameters.

We devise these plots through automated procedures, available on CRAN in the R (R Core Team 2017) package `changepoint.influence`, to bring the importance of influence diagnosis to the attention of researchers in the changepoint community and to stimulate their widespread usage amongst practitioners.

Model instabilities are well-known statistical problems and influence diagnostics are essential to detect them and to investigate the role of single observations that give rise to these instabilities. We call observations whose alteration changes the resulting changepoint segmentation, and thus give rise to model instabilities, *influential*. Influence diagnostics have a long-standing history in regression analysis, see early studies by, among others, Cook (1979) and Belsley, Kuh, and Welsch (1980) who assessed the effect of single observations on coefficient estimates in low-dimensional settings or more recent work by Hellton, Lingjærde, and De Bin (2019), Rajaratnam et al. (2019), and Zhao et al. (2019) who adapted diagnostic tools to high-dimensional regression settings. Developing influence diagnostics for changepoint analysis is arguably even more compelling

CONTACT Rebecca Killick r.killick@lancs.ac.uk Fylde College, Lancaster University, Lancaster LA1 4YW, UK.
Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JCGS.

than it is for regression analysis since influential observations can not only affect parameter estimates (for instance, segment means) but also the entire changepoint segmentation. In the same vein, several recent studies have developed influence diagnostics for variable selection procedures (e.g., De Bin, Boulesteix, and Sauerbrei 2017 for resampling-based methods or Rajaratnam et al. 2019 for the lasso) thereby addressing instabilities in both parameter estimation and model/variable selection.

Our contributions to the changepoint literature are twofold. First, we introduce a new framework for diagnosing influential observations within changepoint models. We propose two types of influence diagnostics. Both types *alter* individual data points and evaluate how, and to what extent, such alterations induce differences in various outputs of the changepoint analysis. They differ in the way the data are altered. In one extreme case, we alter data points by deleting them, one by one, thereby following the intuitive and popular deletion diagnostics (Belsley, Kuh, and Welsch 1980) for regression analysis. In the other extreme case, we alter data points by contaminating them such that each point forms a segment on its own, thereby building on the idea of empirical influence functions used in robust statistics (e.g., Hampel et al. 2011 for an overview or Pison and Van Aelst 2004 for diagnostic plots). In Section 3, we will see that the two proposed types of influence diagnostics provide complementary views. Second, we equip researchers, analysts and practitioners working with changepoint models with a set of diagnostic plots. These plots help to visualize the output of the influence diagnostics and identify whether the original segmentation is vulnerable to instabilities. If so, in-depth plots clearly depict how and where these instabilities manifest. More detailed follow-up visual tools then aim to identify single observations that trigger these instabilities to arise and assess their influencing role.

The remainder of this article is structured as follows. In Section 2, we present a motivating example for the development of changepoint influence diagnostics. We introduce the framework for diagnosing changepoint models in Section 3. In Section 4, we present the influence diagnostic plots that guide practitioners in answering various diagnostic questions. We demonstrate the usage of our graphical influence diagnostics on an application to well-log data in Section 5. Finally, in Section 6, we summarize our contributions and propose several directions for future work.

2. Motivating Example

We present a motivating example to illustrate that changepoint segmentations can be highly sensitive to individual data points, thereby calling for appropriate influence diagnostics to identify and assess these various sources of instability and data influence on them.

2.1. Well-Log Data

We consider the problem of detecting changes in well-log data (Ruanaidh and Fitzgerald 2012). Figure 1(a) displays $n = 1000$ measurements from a probe that is lowered into a bore-hole. The probe takes measurements of the nuclear-magnetic response of the rock that it is passing through. Abrupt changes occur in the measurements as the probe moves from one type of rock strata to another. Changepoint analysis is used to detect these rock strata. While online changepoint detection can be used to modify the settings of the drill in (near-)real time, we focus on influence diagnostics that are suitable for offline changepoint analysis. The well-log data are particularly suited for this purpose as they exhibit several interesting features for influence diagnosis, as discussed below.

2.2. Changepoint Segmentation

Several changepoint methods have been used to detect changes in the well-log data (Fearnhead 2006, and Ruanaidh and Fitzgerald 2012). Following Fearnhead and Rigaill (2019), we focus–throughout the article–on the minimum penalized cost approach to detect changes in the mean and use a Normal likelihood test statistic and the Pruned Exact Linear Time (PELT) algorithm (Killick, Fearnhead, and Eckley 2012), available in the `R` package `changepoint` (Killick and Eckley 2014), to detect these changes. Details about this approach are included in Appendix A of the supplemental material. This is merely an illustrative example of a changepoint model as our framework is more broadly applicable, as will be discussed in Section 6.

In Figure 1(a), we visualize in dashed (blue) lines the changepoints detected from a change in mean model: 19 changes are detected. The segments vary considerably in length, ranging from segments containing as many as 171 observations (last segment) to single observation segments (e.g., two segments in between observations 219 and 221). The latter are very low measurements that occur due to malfunctioning of the probe and can be highly influential. Indeed, if a data point is sufficiently extreme compared to its neighbors, it occurs in a segment of its own (see Fearnhead and Rigaill 2019 and Proposition B.1 in the Appendix).

2.3. Changepoint Stability and Data Influence

While the well-log data have been extensively analyzed through various changepoint methods, the stability of the obtained changepoint segmentation and influence of single data observations on it is less well understood. To illustrate this, consider the influence of two types of data alterations on the obtained segmentation.

First, in Figure 1(b), we *delete* data point 685. This data alteration has a drastic local impact: 17 instead of 19 changes in the mean are detected, the original changepoints at positions 687 and 695 no longer arise and all observations from position 684 to 866 are placed in a single segment instead of the original three. To reiterate, deleting a single observation removes two non-adjacent changepoints—this clearly calls into question any inference regarding those inferred changepoints and the segment means.

Second, in Figure 1(c), we *contaminate* data point 326 (in the middle of a segment) by adding twice the range of the data to its value. Two additional changes around the contaminated data point occur since "outlying" data points are placed

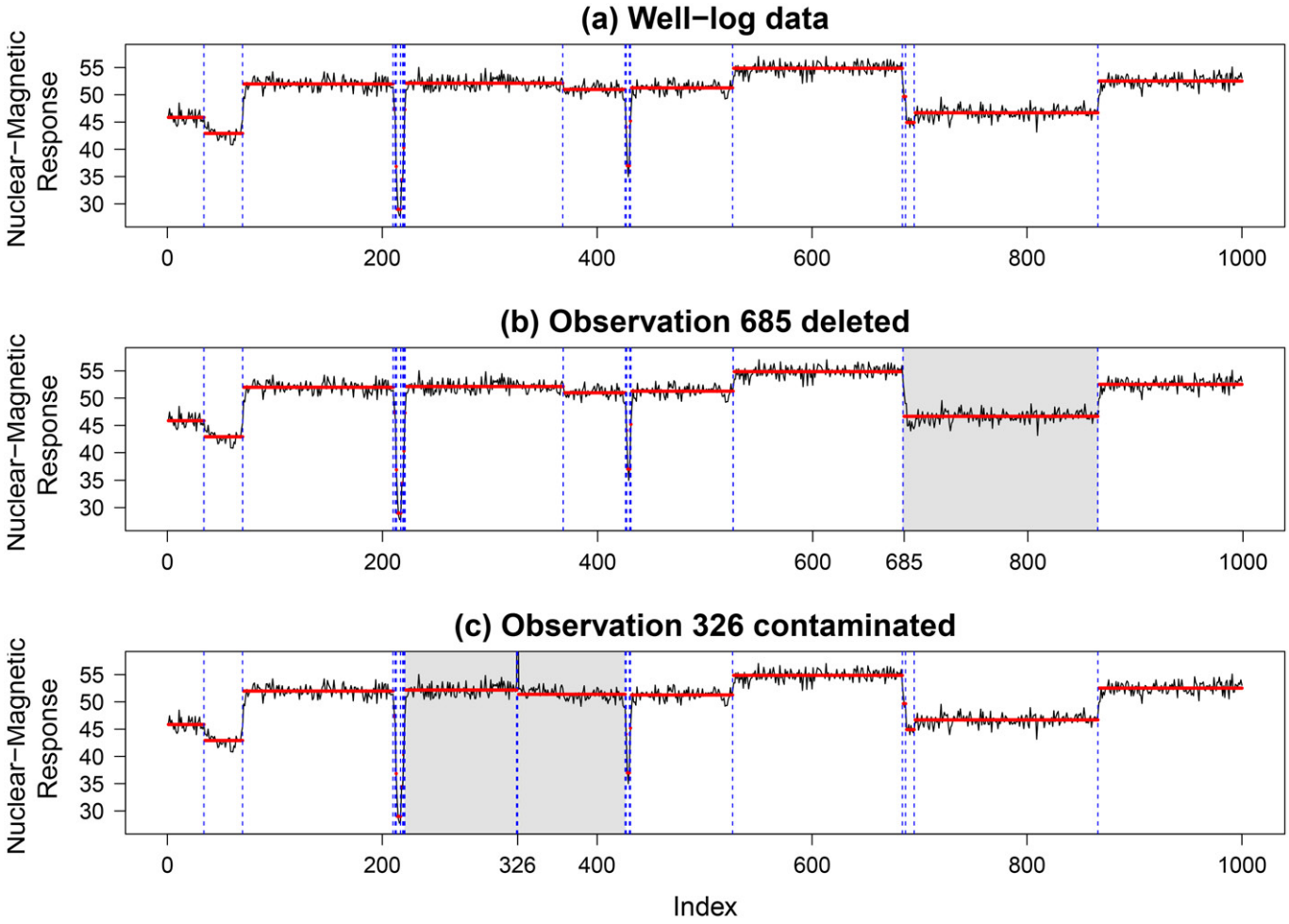

Figure 1. (a) Well-log data with 19 changepoints (vertical dashed blue lines) and segment means (horizontal solid red lines). (b) Segmentation when deleting observation 685. (c) Segmentation when contaminating observation 326. The gray background in panels (b) and (c) highlights the span of changes to the segmentation compared to panel (a).

in their own segment—this is to be expected (see Appendix B.2). Additionally and unexpectedly, the original changepoint at location 368 no longer occurs. Hence, when observations 327–387 are not jointly considered with observations 221–326, the former are not sufficiently different from the latter observations. Again the inference regarding the changepoint at location 368 is affected. These two examples illustrate the dramatic impact slight data alterations or measurement errors might have on the changepoint analysis.

2.4. Diagnostic Questions

While these motivating examples are deliberately chosen to emphasize the potential dramatic influence a *single* observation can have on the output from a changepoint analysis, they raise several general diagnostic questions practitioners might be concerned with. We present three main diagnostic questions, each motivating the need for a particular type of influence diagnostic, as will be discussed in Section 4:

Q1. Is the output of the changepoint analysis stable or vulnerable to data instabilities?
Q2. If vulnerable, how and where do the instabilities manifest?
Q3. Which single influential observations trigger these instabilities to arise and how so?

The first two questions aim to assess the stability of various outputs of the changepoint analysis: For instance, which changepoints are sensitive to the data at hand and does this sensitivity raise questions on their occurrence and/or their location? The third question digs deeper into the influential role of single observations on the various output measures. An important remark that needs to be made here is that influential observations need not to be seen as harmful in the analysis, in the sense of measurement errors or extreme/atypical data points, but can be seen as data points that are highly relevant for obtaining the segmentation at hand (Serneels et al. 2005).

3. Framework for Diagnosing Changepoint Models

In this article, we consider observed sequences of data, $y_1, \ldots, y_n$, and assume a changepoint analysis has been performed on such a sequence resulting in $\hat{m}$ identified changepoints at ordered locations $\hat{\tau}_1, \ldots, \hat{\tau}_{\hat{m}}$. This segmentation splits the data into $\hat{m}+1$ independent segments, the ith of which contains the data points $y_{\tau_i+1}, \ldots, y_{\tau_{i+1}}$ using the convention that $\tau_0 = 0$ and $\tau_{\hat{m}+1} = n$. Our goal is to develop a general

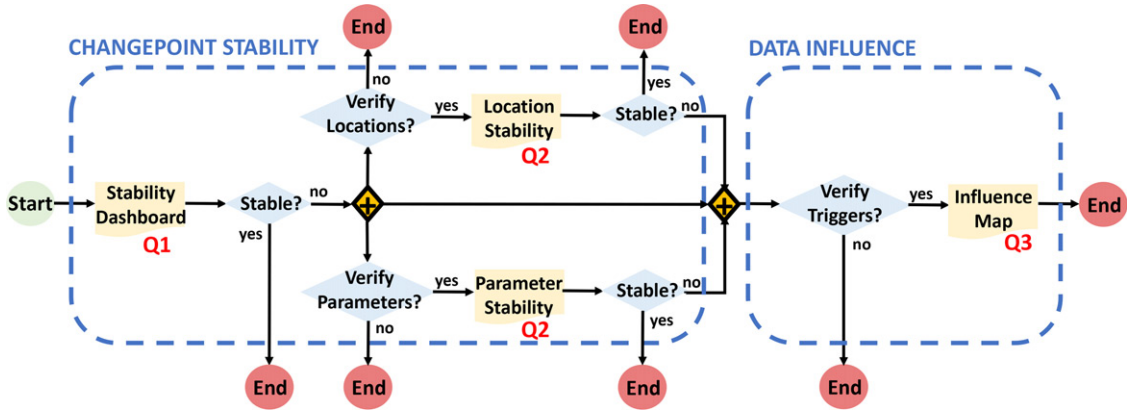

Figure 2. Workflow for assessing a segmentation's stability and identifying influence of single data points on the segmentation obtained.

framework for assessing a segmentation's (in)stability, and understanding the role of each observation y_t, $t = 1, \ldots, n$, on the estimated segmentation. The framework is presented in this section. In Section 4, we explain how to use the newly constructed influence diagnostic plots for these purposes.

The proposed framework allows us to identify and analyze both global and local (observation-specific) instabilities. The global diagnosis involves the assessment of the (in)stability in three pertinent outputs of a changepoint analysis: the number of changepoints, changepoint locations and segment parameters. Monitoring changes in each of these outputs will be a useful guide toward assessing a changepoint model's (in)stability. The more detailed diagnosis then involves the identification of single influential observations that trigger these changepoint instabilities.

Figure 2 visualizes the typical workflow of a changepoint diagnostic analysis, thereby linking both diagnostic objectives (dashed boxes) to the three diagnostic questions (Q1–3 in Section 2) and the four graphical tools as indispensable workflow documents to guide practitioners in answering these questions. The diagnostic questions and corresponding graphics are hierarchically structured from global toward detailed diagnosis. While the Stability Dashboard is useful to address Q1, the location and parameter stability plots (additionally) address Q2, and the Influence Map is useful to (additionally) address Q3.

To detect changepoint (in)stabilities and quantify the effect of a single data point on them, we follow the procedure of consecutively *rolling* through all data points and at each time either *deleting* or *contaminating* a particular observation. This way, we identify (i) whether various output measures of the changepoint analysis are stable or differ substantially after altering individual data points; (ii) how and where instabilities arise: locally, only affecting the segment the deleted data point is within, or global, thereby affecting other segments as well; and (iii) the individual influential data points triggering these instabilities.

3.1. Rolling Procedure

For each data point $t = 1, \ldots, n$ in this rolling procedure, two new segmentations are obtained. The first new segmentation, we call the "Observed Segmentation." Here, we simply re-run the changepoint method on the *altered* (deleted or contaminated) data and record the "Observed Segmentation" obtained. The second segmentation, we call the "Expected Segmentation." This segmentation corresponds to our expectation regarding the change in a particular output measure of the changepoint analysis when either deleting a data point or contaminating it. Detailed results for the expected segmentations under a penalized cost approach for a change in mean are provided in Appendix B.

While all influence diagnostic plots are constructed from the observed segmentation, only the Location Stability plot and the Influence Map rely on the comparison between the observed and expected segmentation. The idea is that these plots should directly highlight unusual behavior rather than changes to the segmentation we expect to see due to the data alteration. We therefore do not display the expected changes but instead compare what is observed to what is expected under a penalized cost approach such that only changes beyond the expected ones are displayed. As such, we draw a practitioner's attention to insightful discrepancies that help to assess how and where instabilities manifest (location stability plot) and quantify the influence of individual data points on a changepoint's (in)stability (Influence Map). The difference between both segmentations presents evidence of a single data point's influence beyond what is to be theoretically expected, which a practitioner can interpret as influence.

3.2. Deleting Observations

Deletion diagnostics have been the subject of extensive research in the context of regression analysis and date back to Cook's distance (Cook 1979), which measures the influence of single observations on various aspects of the fitted regression model, including and excluding the observation in question. For changepoint models, a diagnostic analysis based on deleting data points is complicated by the fact that (i) individual data points can, in addition to parameter estimates as in traditional regression analysis, affect the entire changepoint model through the number of changepoints and their location. (ii) While the total number of observations n might be considerable, each segment contains only a (sometimes small) fraction of the total sample size, hence individual observations can not only have a potentially tremendous local influence on the segment to which it belongs but this influence might also spill over

globally to other segments. Hence, this calls for the need of new deletion diagnostics for detecting changepoint (in)stabilities and understanding the influential role of single data points on them.

Inspired by these intuitive and popular deletion diagnostics, we alter data points by deleting them, one by one, and assess the relative change in various outputs of the changepoint analysis (i.e., number of changepoints, changepoint location, segment parameters) to demonstrate the influence, or not, of the deleted data points. We show in Appendix B that the segmentation expected under a single data point deletion remains the same as the original one unless the data point belonged to a segment of length one. In practice, we implement the deletion approach as follows. By way of example, consider a changepoint segmentation, 1 1 1 2 3 3 3. When we leave out the first observation the expected segmentation then becomes NA 1 1 2 3 3 3. However, when we leave out the fourth observation, the expected segmentation is 1 1 1 NA 2 2 2. Note that we re-number the segments to ensure that two neighboring segments differ in their numbering by one.

3.3. Contaminating Observations

As an alternative to deletion diagnostics, empirical influence functions are commonly used in robust statistics to determine, on a sample-specific basis, the influence of each data point on parameter estimation or prediction (Hampel et al. 2011). To this end, the effect of an infinitesimal contamination at a certain data point on a statistical functional of interest is measured and used as a diagnostic tool to assess its influence. It is hereby crucial to stress the discrepancy between influence and extremeness. While both properties coincide in the detection of influential outliers (i.e., atypical data points), in general, non-outlying influential data points as well as non-influential outliers do exist (Serneels et al. 2005).

Inspired by techniques from robust statistics, we alter data points one by one such that each point is made atypical/outlying and assess the relative change in various changepoint outputs of this contaminated data point. Fearnhead and Rigaill (2019) showed that the segmentation expected under this data alteration corresponds to the segmentation obtained on the original data with two extra changes added before and at the contaminated position. Being different from the bulk of the data, a contaminated point thus warrants its own segment. However, only one extra change occurs if we are close to an original changepoint. Coming back to our earlier example (i.e., segmentation 1 1 1 2 3 3 3), when we contaminate the first observation the expected segmentation becomes 1 2 2 3 4 4 4; hence, there are four segments in total instead of three. However, when we contaminate the second observation, the expected segmentation is 1 2 3 4 5 5 5, thereby including two additional changes.

These two ways of altering observations provide two extreme perspectives: on the one hand, when deleted, the segmentation reveals what would have happened had the observation not been observed. On the other hand, when contaminated, the point is simultaneously both maximally influential (it has its own segment) and minimally influential (in its own segment it does not directly contribute to other segments). Curiously, the contaminated point also forces a shortening of the segments on either side, thus allowing one to identify if a change is sensitive to the length of the segment it is within. Thereby, both present complementary views on the stability of changepoint segmentations and allow us to better grasp the overall influence of single observations.

4. Influence Diagnostic Plots

We create a *set of four diagnostic plots* which range from coarse level to detailed, namely the "Stability Dashboard," "Segment Location Stability," and "Segment Parameter Stability" plots, and finally the "Influence Map." The different plots each aim to tackle a specific diagnostic question (see Figure 2), making the choice of an appropriate plot crucial for highlighting a particular aspect of the influence diagnosis for changepoint models. Practitioners should choose the most appropriate level of detail for the data set and question they are considering. All plots rely on the rolling procedure discussed in Section 3 and are constructed for both the case where data points are consecutively *deleted* and *contaminated*. The influence diagnostic plots can be created via the R package `changepoint.influence`.

To illustrate the usage of the plots and provide guidance on how to interpret their various features, we make use of a simulated data example. We generate an ordered sequence of length $n = 200$ with four changes in mean: the first 50 data points are generated from a standard Normal distribution. Data points 51 to 100 as well as data points 102 to 150 are drawn from a Normal distribution with mean five and unit variance. An atypical data point, drawn from a Normal distribution with mean 15 (10 standard deviations above the mean either side) is included at 101. Finally, the last 50 data points are drawn from a Normal distribution with mean four and unit variance, giving rise to a relatively small change in mean. The simulated data sequence is shown in Figure 3 together with the four changepoints (at positions 50, 100, 101, 145) detected by the Normal likelihood test statistic with the PELT search method.

4.1. Stability Dashboard

At the coarsest level, we have the "*Stability Dashboard*": it presents the original data with the original changepoints depicted as vertical lines at the changepoint locations. The lines are now displayed either dashed (green), dot-dashed (orange) or dotted (red), in line with the extent to which each is vulnerable to data instabilities. We use dashed green for unaffected changepoints (i.e., unaffected when all data points other than itself are altered), dot-dashed orange when a changepoint moves or is deleted (for at least one altered point) and dotted red when a changepoint forms a segment on its own. This plot thus provides a coarse depiction of the results to directly address the first diagnostic question.

The Stability Dashboards for the simulated data example are presented in Figure 4. Both the deletion and outlier method give the same insights (though deviate as we delve deeper): the first changepoint is stable (dash green), the last is somewhat unstable (dot-dash orange), and changepoints at positions 100 and 101 are bounding an outlier (dot red) Whether a practitioner con-

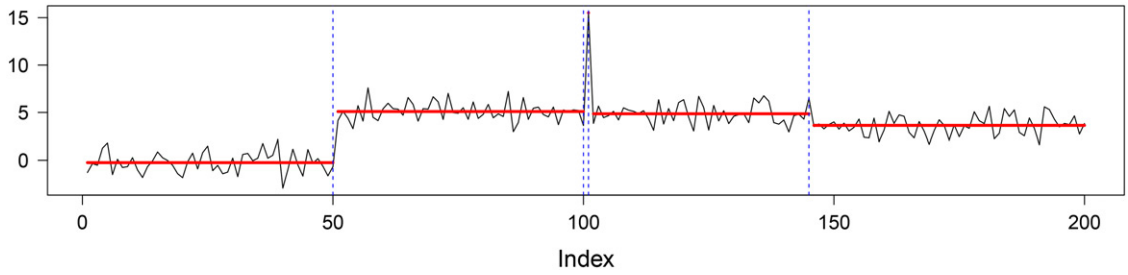

Figure 3. Simulated data example with four changepoints (vertical dashed lines). The horizontal solid lines are the segment means.

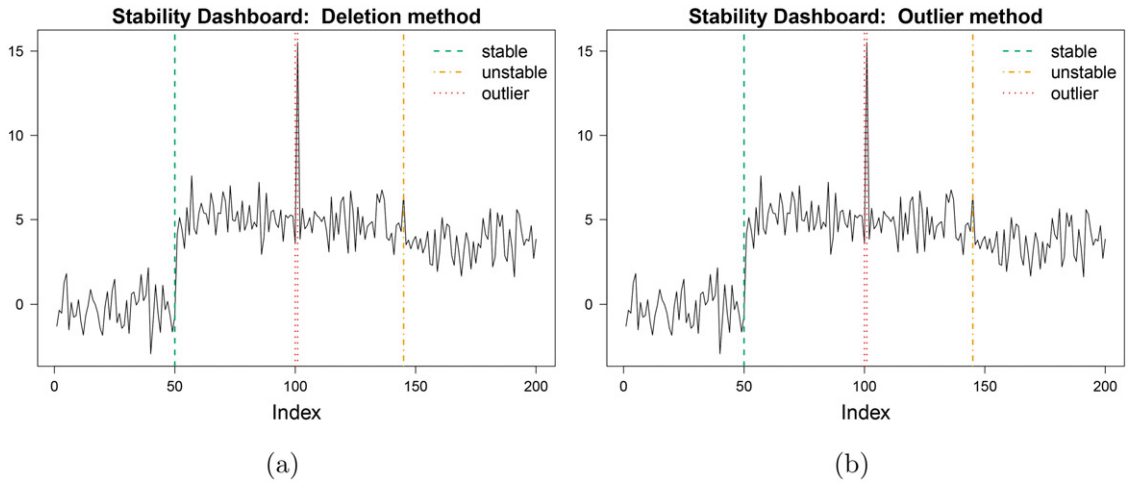

Figure 4. Stability Dashboard when (a) deleting and (b) contaminating observations.

tinues or not with inspecting the more detailed diagnostic plots depends on the outcome of the Stability Dashboard. If all the changepoints are stable (green), then the segmentation seems stable and there may be no need to delve further into the results. In contrast, if some changepoints are unstable or outlying then the practitioner may wish to further investigate how and where these instabilities manifest.

4.2. Segment Location Stability

We next move to the second level to assess how and where instabilities in the location of the changepoints occur. To this end, a Location Stability plot of the changepoint locations across the n altered data points can be used.

We record the number of times a changepoint alteration occurs as well as the location of any moved or additional changepoints. This plot thus allows practitioners to assess how and where instabilities manifest, thereby (partially) addressing the second diagnostic question.

For a segmentation which does not vary across altered data points, we expect each of the original changepoints to stay in place when all data points other than itself are altered. For ease of use, we directly display the discrepancy in number of changepoint occurrences from this expected maximum. This way only unstable (dot-dashed orange) or outlying (dotted red) original changepoints may enter the plot with a negative difference. The latter indicates that the changepoint no longer occurs at its original location for some instances. Either the changepoint moves to another location or it disappears completely. When a changepoint moves, it will be offset and depicted by a positive difference at another location, thereby leading toward a net balance of zero. Disappearing changepoints, on the other hand, do not appear in the plot but can be deduced from net negative balances. It can also occur that the original changepoints remain but that additional changepoints occur due to the alterations (net positive) but this is much less common in our experience.

The Location Stability plots for the simulated data example are presented in Figure 5. When deleting observations one by one, changepoint 145 turns out to be slightly unstable as it moves earlier for one instance, as can be seen from the negative dot-dashed (orange) line of height one at the changepoint location which is offset by the positive solid (black) line of the same height at location 137, see Figure 5(a). When contaminating observations one by one, changepoint 145 shows more instability as it displays differences from what is expected on more than 25 instances. On six instances, the changepoint is moved earlier, as can be seen from the black positive line of height six in Figure 5(b). The net negative balance indicates that the changepoint disappeared completely for several data contaminations. The negative red line of length one at location 101 (Figure 5(a)) indicates that the changepoint induced by the outlying data point at this location disappears completely when it is deleted. These plots give an indication of *what* the stability looks like but not *which* data points influence this behaviour.

4.3. Segment Parameter Stability

This plot complements the previous in tackling the second diagnostic question by considering instabilities in the segment

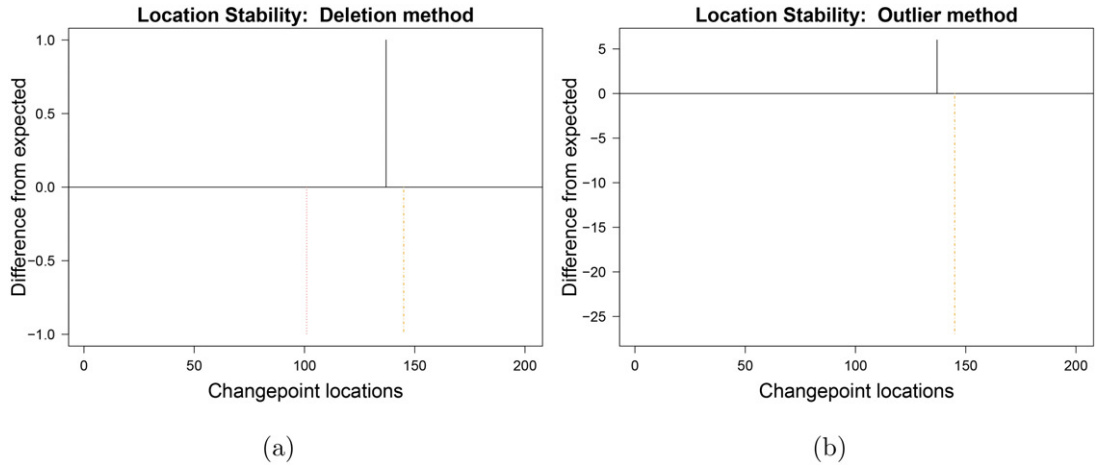

Figure 5. Location stability plot of the simulated data when (a) deleting and (b) contaminating observations.

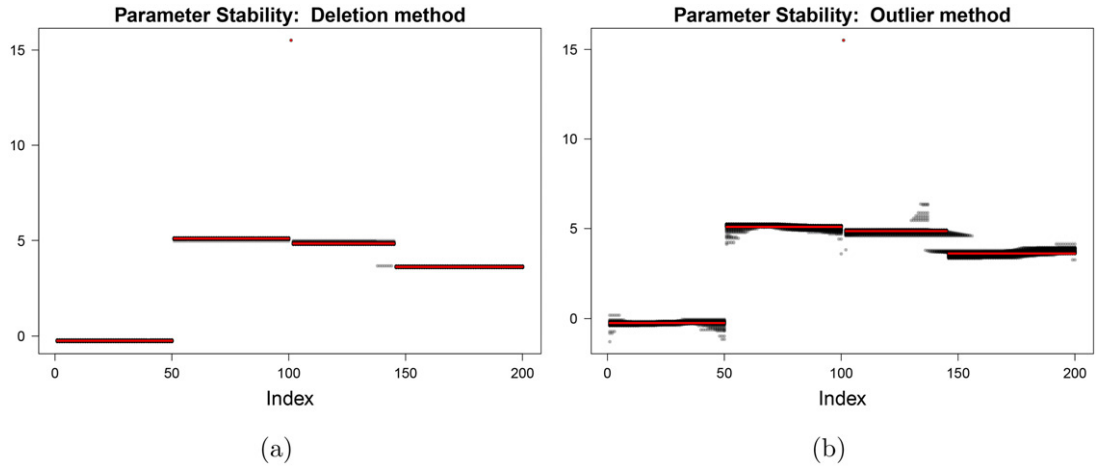

Figure 6. Parameter stability plot of the simulated data when (a) deleting and (b) contaminating observations.

parameters, such as the mean. It is important to investigate the segment parameters separately as the changepoint locations may vary but for a small or uncertain changepoint, the segment parameters may not vary considerably. If one is only interested in inference on the segment parameters and not the changepoint locations then this is important information.

To construct our diagnostic plot we start by depicting the original segment parameters, such as the mean in our example, by solid (red) lines, which correspond to the ones from Figure 3. On the same plot, we add the mean of each data point across the n data alterations over time. For many segmentations where the altered point is far away, there will be no difference in the mean values. We thus take the unique values of the segment means and plot them in shades of gray, scaled by the frequency of occurrence. Thus, common values across many iterations of the rolling procedure appear darker than those across few iterations. The original parameter estimate is intentionally thick on the plot to ensure that any black seen around this is meaningful. The wider the dark area (vertically) that can be seen around the original segment means (red), the more evidence of instability.

The parameter stability plots for the simulated data example are presented in Figure 6. For the deletion method (Figure 6(a)), the means across the data deletions appear very tightly around the original means, thereby supporting stability of the segment means. Only for data points 137–144, a minor instability is observed by the additional dark (lower) line which occurs at the height of the next segment's mean and is caused by changepoint 145 being moved earlier.

For the outlier method (Figure 6(b)), the darker areas are typically larger in size, especially toward the edges of a segment. This is in line with our expectation, since the data contamination induces two additional changes thereby triggering directly surrounding segments to be smaller in size. The additional variability then arises due to the fewer data points available for estimation of the segment parameters. This results in a "bleeding" effect at the edges of the segments. The most pronounced instability in segment means is again observed for the data points around the last changepoint. Coupled with the Stability Dashboard, it is clear that the changepoint moves both earlier (producing a lower mean) and disappears (producing a higher mean after 145) although again it gives no information as to *which* observations are responsible for this.

4.4. Influence Map

At the most detailed level, we have an *Influence Map* of salient differences between the observed and expected segmentation from the deleted or contaminated data points. This final plot is

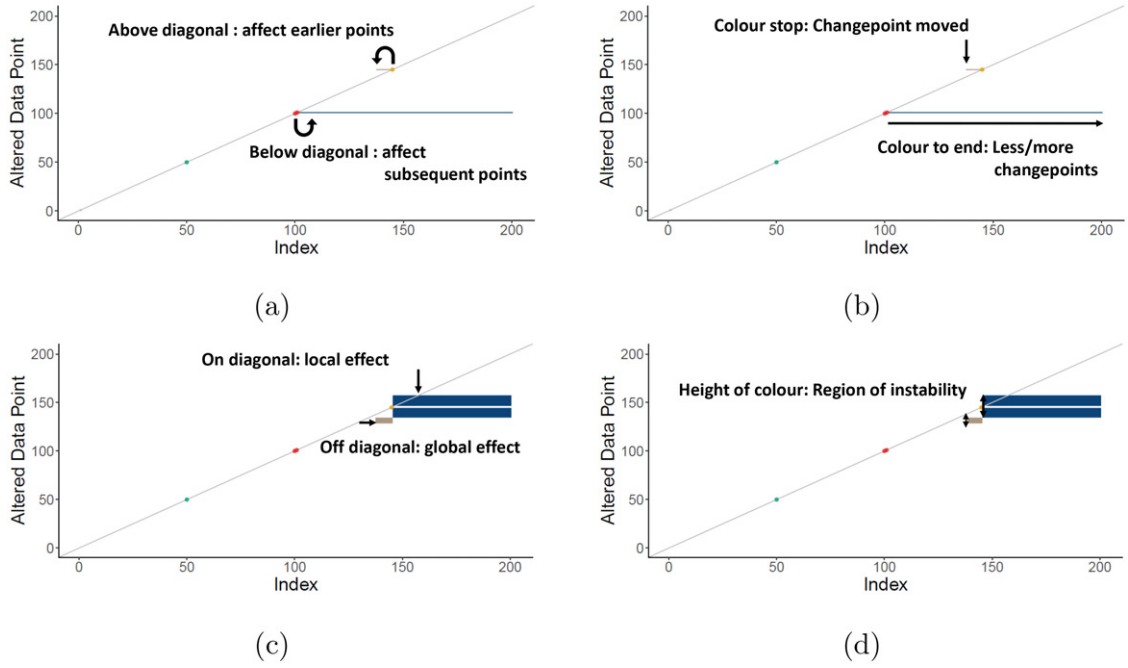

Figure 7. Main features of the Influence Map.

a heat map which identifies single, influential observations that trigger changepoint instabilities, thereby addressing the third diagnostic question. The heat map depicts the difference in segment number between the observed and expected segmentations across each of the altered data points. Analogous influence maps can be made for other outputs of the changepoint analysis.

The horizontal axis of the Influence Map is the standard time index of the original data (1 to n). The vertical-axis indexes the altered data point (1 to n). Each colored (taupe or blue) pixel marks the difference between the observed and expected segmentation at the specific (x, y) coordinate. The data point on the vertical-axis should be understood as the influential data point whose alteration leads to changes in the affected data points on the horizontal axis if any coloring appears. We color zero difference in the heat map as white, increases in segment number as taupe and decreases as blue. Hence, data points on the vertical axis without a single colored co-ordinate on the horizontal axis can be considered as noninfluential since they do not trigger any changepoint instability. Rows with colored pixels correspond to data points which are instability triggers. The intensity of the color then signifies the magnitude of the instability (namely the increase or decrease in segment number). Coloured areas are expected to occur around unstable or outlying (orange or red) changepoints, which are depicted as colored circles on the diagonal.

Before discussing the Influence Maps for the simulated data example, we describe its important features to aid practitioners in studying the influential role of the individual data points through these maps. These features are summarized in Figure 7. (i) Figure 7(a) highlights the role of the *diagonal*: coloring above the diagonal indicates that an alteration of the corresponding data point (on the vertical axis) affects earlier data points, coloring below the diagonal indicates that subsequent data points are affected. (ii) Figure 7(b) concerns the *horizontal span* of the coloring: a stop in coloring indicates that changepoints have moved, while a continuation of coloring to the last data point indicates that, in total, fewer or additional changepoints are detected. (iii) Figure 7(c) zooms in on the discrepancy between *local versus global* effects. Most coloring originates on the diagonal, thereby indicating that a data point's alteration mainly affects neighboring data points that most often belong to the same segment. By contrast, in some cases a colored pixel may originate away from the diagonal, thereby exercising global influence. (iv) Finally, Figure 7(d) zooms in on the *height* of the coloring. All data points (on the vertical axis) that appear in the colored area are influential and assert influence over the corresponding data points on the horizontal axis. The height can be seen as the extent to which instability arises in this influential region.

Relying on these features, we are ready to discuss the Influence Maps for the simulated data example, as presented in Figure 8. Across both maps, few colored areas appear, each of them characterizing some form of instability. All of them occur around the originally detected unstable or outlying changepoints. The instability triggers (i.e., data points on vertical axis with coloring) are observations 101 and 129-157; their influential role will be detailed below. Note that most colored areas are blue, thereby indicating that a particular data point (on the horizontal axis) has a lower segment number in the observed segmentation than expected; in other words less expected changepoints occur. We subsequently discuss the Influence Maps according to their main features.

(i) In this example, we see that influential data points have a tendency to affect subsequent data points rather than preceding ones since most coloring occurs below the diagonal. Consider the blue coloring in Figure 8(a). When deleting outlying data point 101 (i.e., the instability trigger on the vertical axis), the changepoint induced by it disappears, and thus all subsequent observations (horizontal axis) are affected by having a lower segment number than expected. The Influence Map highlights

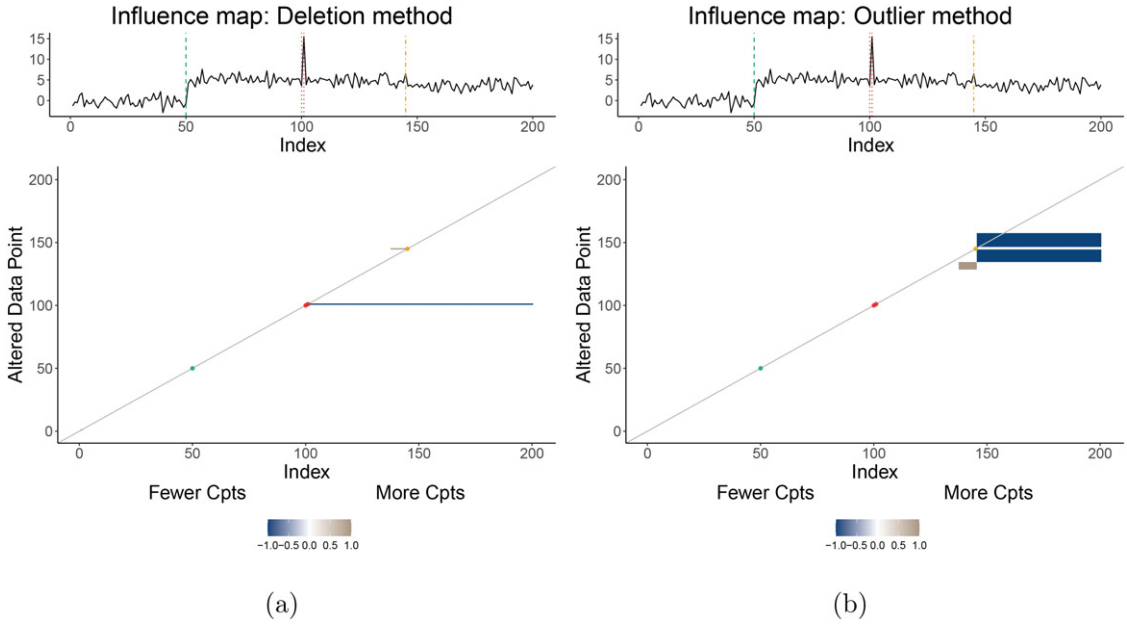

Figure 8. Influence Map of the simulated data when (a) deleting and (b) contaminating observations.

this data instability through the blue coloring of all pixels until the last data point and is in line with the negative dotted red line of height one in the location stability plot (Figure 5(a)).

(ii) We find evidence of influential data points moving original changepoints as well as triggering some to disappear. To this end, we zoom in on the instability of changepoint 145. When deleting this data point, Figure 8(a) (taupe pixels above diagonal line) shows that the changepoint gets moved earlier toward position 137; in line with the black positive height in Figure 5(a). The same change in changepoint location occurs when contaminating observations 129–135, as can be seen from the taupe below-diagonal coloring in Figure 8(b). By contrast, the changepoint disappears when observations 136–157 (vertical axis) are contaminated (blue zone).

(iii) Almost all coloring originates on the diagonal, thereby indicating that a data point's alteration mainly affects its neighbors of the same segment. An exception is the contamination of observations 129–135 (right panel) which makes observations 138 to 144 move from its original fourth segment to the fifth segment.

(iv) Finally, two influential regions appear in Figure 8(b). The blue influential region is the most outspoken one: the contamination of no less than 21 data points (i.e., instability triggers 136–157 on vertical axis) all trigger changepoint 145 to disappear; in line with the net negative balance for changepoint 145 in Figure 5(b). The smaller taupe region indicates that the contamination of six data points (i.e., instability triggers 129–135 on vertical axis) causes the changepoint at location 145 to move earlier; in line with the black positive difference of height six at position 137 in the Location Stability plot (Figure 5(b)).

5. Well-log Application

We now return to the well-log data, presented in Section 2, and address our main diagnostic questions one by one.

5.1. Stability of the Changepoint Analysis

We start by tackling our general diagnostic question "Is the output of changepoint analysis stable or vulnerable to data instabilities?" through the Stability Dashboards, presented in Figure 9. Out of the original 19 changepoints, 15 are depicted as potentially unstable (orange or red) by the deletion method (Figure 9(a)) while all but one (change at location 217) are colored orange or red by the outlier method (Figure 9(b)).

While such pronounced results will not arise for each changepoint analysis, they do illustrate that influence diagnostics should not be overlooked but rather considered as a much needed natural successor to any changepoint analysis. The mere visualization of one single additional graphic, the Stability Dashboard, can either re-assure practitioners on the stability of their performed analysis or warn them for the occurrence of instabilities. In the latter case, a more detailed influence diagnosis can be performed through our other diagnostic tools which are discussed next.

5.2. Manifestation of the Instabilities

Next, we address the question "How and where do the instabilities manifest?" First, consider the stability of the changepoint *locations*, as visualized in the Location Stability plots of Figure 10. For the deletion method, the few positive short (black) heights (Figure 10(a)) immediately highlight that only a minority of location instabilities occur. Hence, while many changes are labeled as potentially unstable (orange dot-dashed lines in Figure 10(a)); these instabilities only manifest themselves in rare cases. For the outlier method, by contrast, especially changepoints 368 and 695 are prone to more severe instability as can be observed from the long negative heights at their locations in Figure 10(b). The positive solid (black) heights show that when the changepoints move, they do not move far from existing changepoint locations. The net negative balance shows

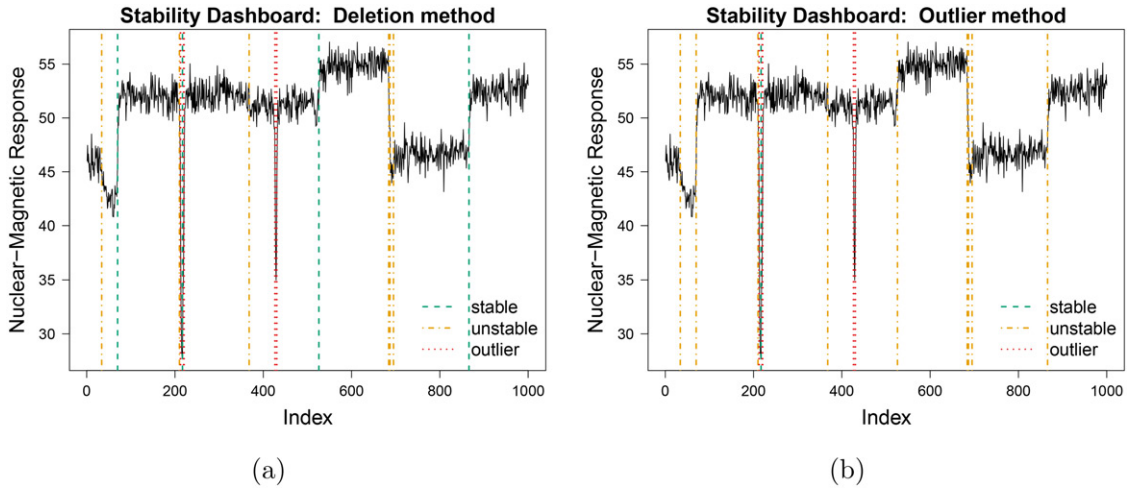

Figure 9. Stability Dashboards for the Well-log data.

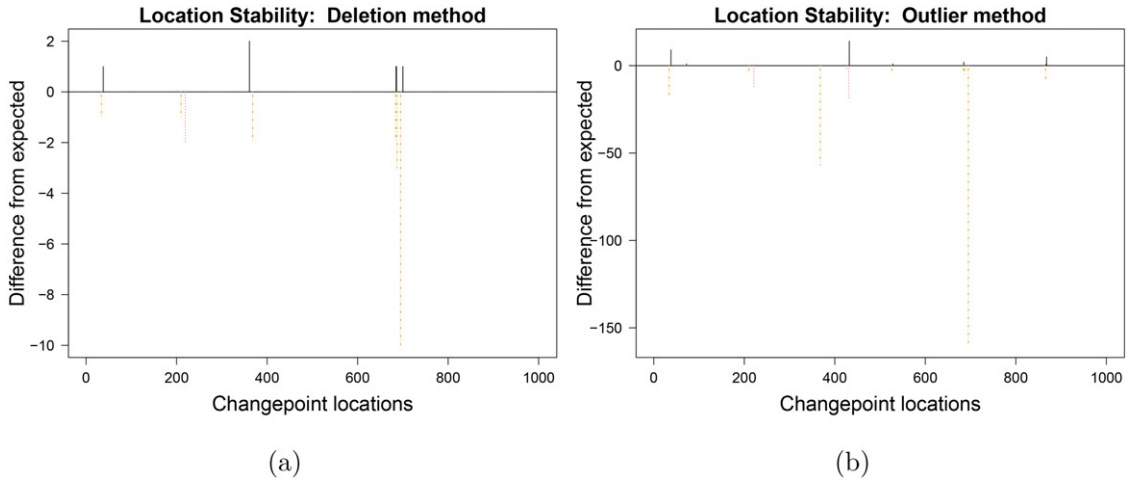

Figure 10. Location Stability plots for the Well-log data.

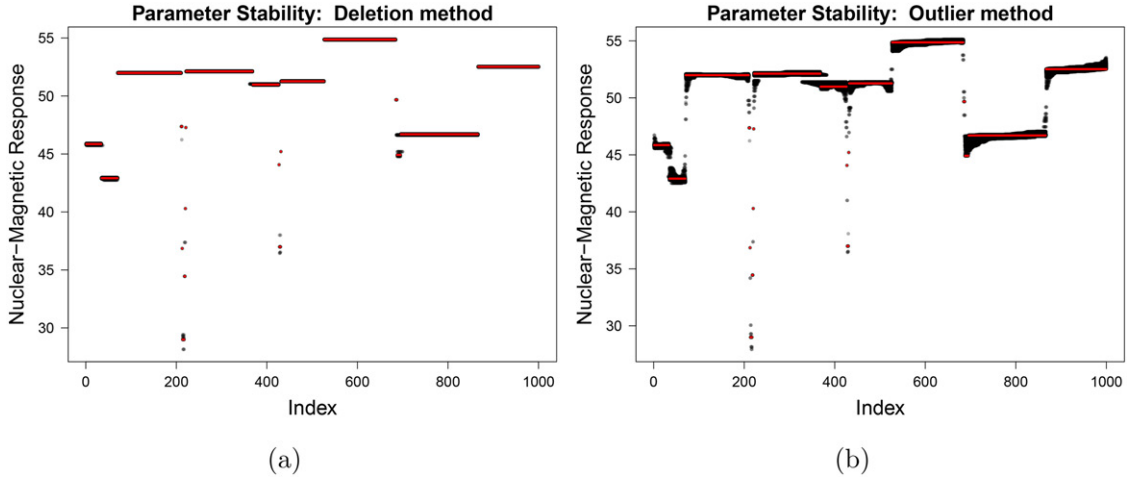

Figure 11. Parameter Stability plots for the Well-log data.

that overall changepoints are deleted rather than moved. The analysis of the well-log data is, however, more complex than the simulation data example, which makes it harder to directly associate the changepoint moves (black positive lines) to the original changepoints (colored negative lines) in the location stability plots. Practitioners are therefore advised to consult the more detailed Influence Map to match how various data perturbations affect the original data.

Second, consider the stability of the segment *parameters*, namely the mean, as visualized in the Parameter Stability plots of

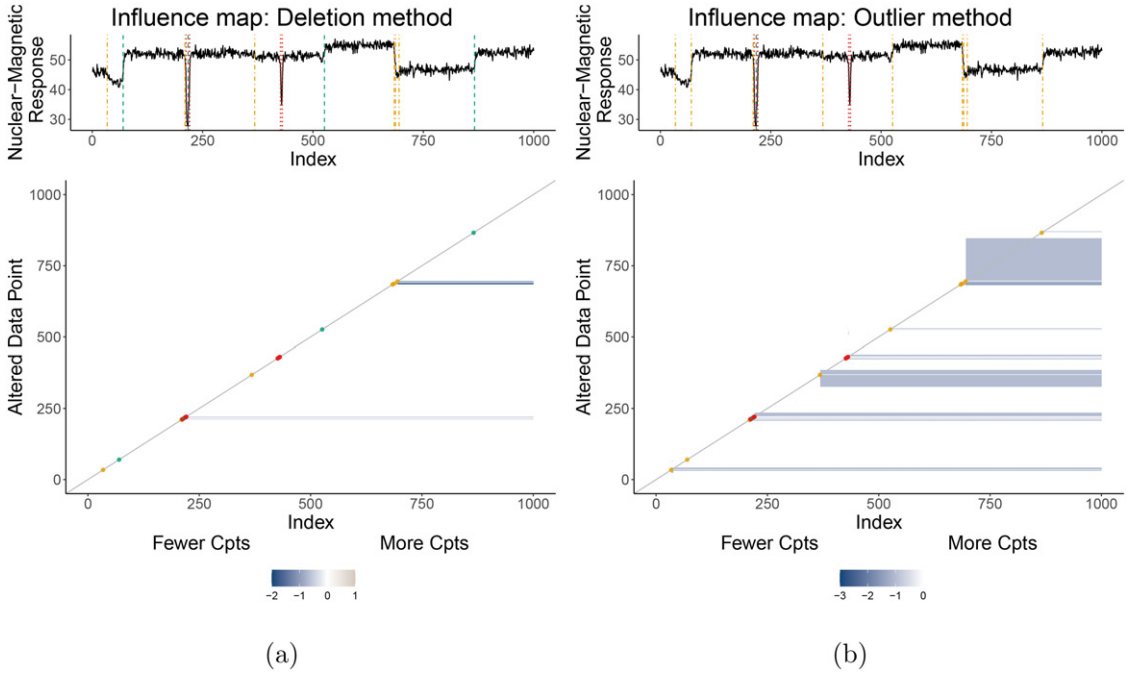

Figure 12. Influence Maps for the Well-log data.

Figure 11. Due to the (minor) evidence of changepoint location instability for the deletion method, the vast majority of segment means appears very stable in Figure 11(a). Some instability can be observed for observations around changepoint 368 and 687 by the additional dark lines occurring to their left. This indicates that these changepoints are (somewhat) unstable and cause their preceding observations to have lower segment means when the changepoint is moved earlier. These same instabilities in the segment mean appear even stronger in the Parameter Stability plot of the outlier method (Figure 11(b)). Furthermore, around observation 200, both parameter stability plots show some instabilities corresponding to the malfunctioning of the probe. When only two observations form a segment, the deletion or contamination of one of them causes the mean to be the other data point, thereby giving rise to the light gray areas of instability. Alongside this there are further instabilities in the 300–450 range. Jointly using the parameter stability and the location stability plots we can see that this instability is driven by the instances where the change before 400 does not manifest along with the movement of the change after 400. Anyone inspecting the original segmentation would be unlikely to think that a single data point would have such a profound and far-reaching effect on the resulting segment means.

5.3. Sources of the Instabilities

Finally, we consider our more detailed influence Diagnostic Objective, namely *"Which single influential observations trigger these instabilities to arise and how so?"*, through the lens of the Influence Maps (Figure 12).

We first discuss the results for the deletion method (Figure 12(a)). Several potentially unstable (orange) changepoints–such as the one at location 34–hardly have any (clearly) visible colored pixels of instability surrounding them. This is due to the larger dataset size $n = 1000$ than our previous example. We recommend to consider these changepoints as sufficiently stable. A handful of observations are found to assert notable influence. Recall that within the data just after time points 200 and 400 there are malfunctions in the observations recorded. Rather than affect single observations these result in a quick degradation and restoration of the signal. Most, but not all, of these are isolated as individual segments in the original segmentation. The deletion of irregular data points 212, 218, and 219 each triggers the changepoint induced by their outlyingness to disappear. Similarly, almost all observations in the range 685–698 are highly influential: their deletion triggers the segment 687–695 to no longer arise.

Secondly, consider the results for the outlier method (Figure 12(b)). The largest area of influential observations (vertical axis) concerns the data points in the range 681–845. Their contamination triggers changepoint 695 to disappear; the first five additionally trigger the removal of changepoints 684 and 687. This is interesting as it signifies that if the original segment were shorter, the changepoint at 695 would no longer arise. Care must be taken about inference made for this changepoint.

Interestingly, the same phenomenon occurs for the instability triggers 326–384: their contamination causes changepoint 368 to disappear. Other similar, though less outspoken, influential regions occur in Figure 12(b). All of these are blue, indicating a changepoint removal and the majority of them affect subsequent observations from the same segment (since the colored pixels start below-diagonal and continue until the end of the sample).

6. Conclusion

Motivated by questions from practitioners in applying commonly used changepoint methods, this article has presented the first approach to considering influence of the observed data

points on changepoint segmentations. We provide a framework for two methods to characterize influence; deletion and contamination. Alongside the framework three levels of graphics were introduced. The Stability Dashboard provides an overview of the results which indicates if there are any locations of concern. The location and parameter stability plots provide the second granularity of detail indicating how the segmentations are affected. The most detailed level is the influence map which includes *which* observations are influential and *how* they influence the segmentation.

A challenging aspect of the proposed approach is to characterize what a "no problem" situation looks like. The simplest answer is that if all changepoints are stable (dashed green) then there is no problem. However, in reality this is unlikely to be the case as short segments, small changepoints and clustered changepoints as seen in our well-log example, are common. We have deliberately not addressed this subjective issue of when a point is "influential enough" to compromise an analysis, as the deliberation of this depends on the downstream pipeline of decision to be made based upon the original segmentation. We prefer to leave this evaluation to the sensibility of the practitioner.

We illustrated our general approach using the change in Normal mean test statistic coupled with the PELT search method but we stress that our approach can be applied to *all* changepoint methods. Furthermore, the only aspect of this article which is specific to the change in Normal mean test statistic is our justification of the expected alterations to the segmentations that feed into the location stability and influence map. For other test statistics, these either need to be calculated or to plot the altered segmentations rather than the difference from the expected. This is an important consideration as if one was using a robust test statistic such as that provided in Fearnhead and Rigaill (2019), then the outlier method would not guarantee the creation of two new changepoints. It is still interesting to consider the influence of the data in this robust setting but we leave this for future work. Our aim in this article is to provide a framework for assessing influence in a general sense; utilizing a common test statistic and search method purely as an example. In the future research, it would be interesting to explore whether different test statistic and search method combinations may be more/less prone to instabilities than others.

Finally, one may consider that the influence plots characterize information about uncertainty in the changepoint segmentation. Whilst this is true, we are not aiming to provide confidence intervals or similar measures of uncertainty quantification. Akin to regression analyses, there are questions best answered by confidence intervals and others by measures of influence. Similarly, we have advocated questions here that practitioners may wish to answer for which a measure of influence for changepoint segmentations is required.

Supplemental Materials

R-code: The supplemental files for this article include R-code to obtain the influence diagnostic plots with the R-package `changepoint.influence`, version 1.0. We include an R-script that generates all the plots for the simulated data example and the well-log data example. Please consult the file README contained in the zip file for more details. (Rcode.zip, zip archive)

Appendix: The Appendix contains background to changepoint methods used and details on the expected segmentation (Appendix.pdf)

Acknowledgments

We thank to the editor, associate editor and reviewers for their thorough review and highly appreciate their constructive comments which substantially improved the quality of the article.

Funding

Wilms gratefully acknowledges funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 832671. Killick gratefully acknowledges funding from EP/R01860X/1, EP/T014105/1, NE/T012307/1 and NE/T006102/1. Matteson gratefully acknowledges funding from a Xerox PARC Faculty Research Award, National Science Foundation Awards 1455172, 1934985, 1940124, and 1940276, USAID, and Cornell University Atkinson Center for a Sustainable Future.

References

Belsley, D. A., Kuh, E., and Welsch, R. E. (1980), *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, Vol. 571. New York: Wiley. [753,754]

Chapman, J.-L., and Killick, R. (2020), "An Assessment of Practitioners Approaches to Forecasting in the Presence of Changepoints," *Quality and Reliability Engineering International*, 36, 2676–2687. [753]

Cook, R. D. (1979), "Influential Observations in Linear Regression," *Journal of the American Statistical Association*, 74, 169–174. [753,756]

De Bin, R., Boulesteix, A.-L., and Sauerbrei, W. (2017), "Detection of Influential Points as a Byproduct of Resampling-Based Variable Selection Procedures," *Computational Statistics & Data Analysis*, 116, 19–31. [754]

Eckley, I. A., Fearnhead, P., and Killick, R. (2011), "Analysis of Changepoint Models," in *Bayesian Time Series Models*, chapter 1, eds. D. Barber, T. Cemgil, and S. Chiappa. Cambridge: Cambridge University Press, pp. 1–24. [753]

Erdman, C., and Emerson, J. W. (2008), "A Fast Bayesian Change Point Analysis for the Segmentation of Microarray Data," *Bioinformatics*, 24, 2143–2148. [753]

Fearnhead, P. (2006), "Exact and Efficient Bayesian Inference for Multiple Changepoint Problems," *Statistics and Computing*, 16, 203–213. [754]

Fearnhead, P., and Rigaill, G. (2019), "Changepoint Detection in the Presence of Outliers," *Journal of the American Statistical Association*, 114, 169–183. [754,757,764]

Fryzlewicz, P. (2014), "Wild Binary Segmentation for Multiple Change-Point Detection," *The Annals of Statistics*, 42, 2243–2281. [753]

Grundy, T., Killick, R., and Mihaylov, G. (2020), "High-Dimensional Changepoint Detection Via a Geometrically Inspired Mapping," *Statistics and Computing*, 30, 1155–1166. [753]

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (2011), *Robust Statistics: The Approach Based on Influence Functions*, Vol. 196, Wiley. [754,757]

Hellton, K. H., Lingjærde, C., and De Bin, R. (2019), "Influence of Single Observations on the Choice of the Penalty Parameter in Ridge Regression," arXiv:1911.03662. [753]

Hocking, T. D., Schleiermacher, G., Janoueix-Lerosey, I., Boeva, V., Cappo, J., Delattre, O., Bach, F., and Vert, J.-P. (2013), "Learning Smoothing Models of Copy Number Profiles Using Breakpoint Annotations," *BMC Bioinformatics*, 14, 164. [753]

Killick, R., and Eckley, I. (2014), "changepoint: An R Package for Changepoint Analysis," *Journal of Statistical Software*, 58, 1–19. [754]

Killick, R., Fearnhead, P., and Eckley, I. A. (2012), "Optimal Detection of Changepoints With a Linear Computational Cost," *Journal of the American Statistical Association*, 107, 1590–1598. [753,754]

Kwon, D., Ko, K., Vannucci, M., Reddy, A. N., and Kim, S. (2006), "Wavelet Methods for the Detection of Anomalies and Their Application to Network Traffic Analysis," *Quality and Reliability Engineering International*, 22, 953–969. [753]

Leeson, A. A., Van Wessem, J. M., Lightenberg, S. R. M., Shepherd, A., Van Den Broeke, M. R., Killick, R., Skvarca, P., Marinsek, S., and Colwell, S. (2017), "Regional Climate of the Larsen B embayment 1980–2014," *Journal of Glaciology*, 63, 683–690. [753]

Page, E. S. (1954), "Continuous Inspection Schemes," *Biometrika*, 41, 100–115. [753]

Pison, G., and Van Aelst, S. (2004), "Diagnostic Plots for Robust Multivariate Methods," *Journal of Computational and Graphical Statistics*, 13, 310–329. [754]

R Core Team. (2017), *R: A Language and Environment for Statistical Computing*, Vienna, Austria: R Foundation for Statistical Computing. [753]

Rajaratnam, B., Roberts, S., Sparks, D., and Yu, H. (2019), "Influence Diagnostics for High-Dimensional Lasso Regression," *Journal of Computational and Graphical Statistics*, 28, 877–890. [753,754]

Ruanaidh, J. J. O., and Fitzgerald, W. J. (2012), *Numerical Bayesian Methods Applied to Signal Processing*, Springer Science & Business Media, New York: Springer. [754]

Rubin-Delanchy, P., Lawson, D. J., and Heard, N. A. (2016), "Anomaly Detection for Cyber Security Applications," in *Dynamic Networks and Cyber-Security*, World Scientific, pp. 137–156. [753]

Serneels, S., Geladi, P., Moens, M., Blockhuys, F., and Van Espen, P. J. (2005), "Influence Properties of Trilinear Partial Least Squares Regression," *Journal of Chemometrics: A Journal of the Chemometrics Society*, 19, 405–411. [755,757]

Spokoiny, V. (2009), "Multiscale Local Change Point Detection With Applications to Value-at-Risk," *The Annals of Statistics*, 37, 1405–1436. [753]

Wu, H., and Matteson, D. S. (2020), "Adaptive Bayesian Changepoint Analysis and Local Outlier Scoring," arXiv:2011.09437. [753]

Zhao, J., Liu, C., Niu, L., and Leng, C. (2019), "Multiple Influential Point Detection in High Dimensional Regression Spaces," *Journal of the Royal Statistical Society*, Series B, 81, 385–408. [753]