

1 Exploratory Statistics for Break Detection

A break or structural change in a time series is considered to have taken place if a certain characteristic of the time series is altered thereafter. Changes can be persistent over time or transient (i.e. two structural changes of the same amount or type but opposite directions). Typical causes for air quality data are changes in measurement instrumentation or in environmental conditions. Such changes can apply to the mean or minimum values, to the percentage of missing values or zero measurements and the like. We considered the latter ones in our descriptive analyses in ???. In the following we focus on abrupt changes in the running median.

Regarding structural change detection, results in [?] are clear, but not as convincing as for the outlier case. The recommended method is the 'moving average filter', which we follow up by testing its performance and suitability on AirBase [?] data for both an extended period of 2000 to 2010 and a series of pollutants. Furthermore the aim is to proof its robustness and to refine and determine the optimum set of parameter values when applied on the various pollutants.

1.1 Method and Parameterisation

The method we consider is the Kolmogorov-Zurbenko adaptive filter, introduced by [?], enhancing work by [?] in order to more easily detect abrupt changes and more accurately estimate the time of these events.

The idea behind the Kolmogorov-Zurbenko filter ([?]) is to smooth out short term variations from a time series by iteratively applying moving average filters. The smoothed time series reflects seasonal patterns, trends and potentially breaks, i.e. structural changes. Break points are characterized by increased variability in the surrounding window. Thus, for identifying break locations local maxima in variance of the smoothed time series can be investigated. However, abrupt discontinuities tend to be smoothed too, making it hard to identify them and to precisely determine their time of introduction. The procedure was thus refined by [?] allowing for adaptive window width in dependence on the rate of change leading to a sharpening at break locations.

The method depends on two parameters: the half window width q and the number of filtering iterations k . As, especially for smaller window sizes, the method tends to over detection, a third parameter (threshold b) is introduced. This threshold is typically a quantile of the variance time series corresponding to the adaptively smoothed time series. Only local maxima in variance above this threshold are considered to be indicators of suspected inhomogeneities.

The subsequent steps of change detection by the Kolmogorov-Zurbenko (adaptive) filter are illustrated in Figure 1.1. The first two columns show the filters applied to an original time series, the columns to the right show the filters applied to the same time series after removing the seasonal component. In each case, figures on the left show computation of the Kolmogorov-

Zurbenko (kz) break points, while figures in the second column show respective plots for the adaptive version (kza). From top to bottom, first the raw time series is shown with the filter added in red, followed by a pure display of the filter, the variance of the filter where local maxima have been highlighted, and finally the raw time series with identified breaks.

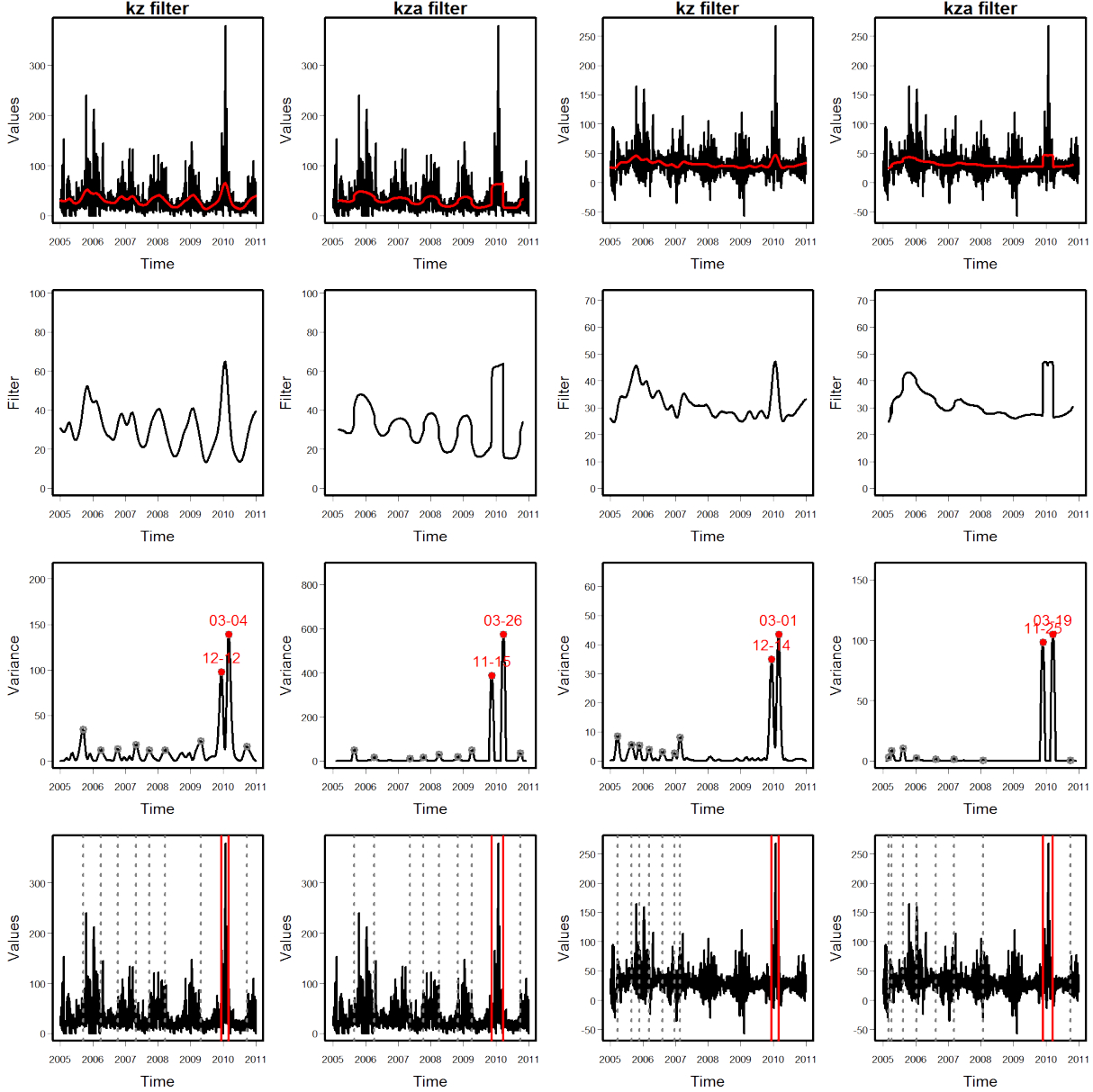


Figure 1.1: Change detection with the Kolmogorov-Zurbenko (adaptive) filter. From top to bottom: time series with filter, pure filter, filter's variance and time series with detected break points.

The first two columns show the filters applied to the original time series from station CZ0MBEL, the columns to the right show the filters applied to the same time series after removing the seasonal component.

The example time series of daily PM_{10} measurements was taken from station CZ0MBEL. Parameters were a quantile threshold of 0.95 (and 0.75), window width of two months and three iterations for each of the filters.

In our implementation we made use of functions `kz` and `kza` from the R package `kza` [?], as

well as function `runsd` from package `caTools` [?].

Parameter fine-tuning is complicated by the fact that ground truth data cannot be provided. For hourly data [?] recommend a single iteration for the Kolmogorov-Zurbenko filter (the adaptive version had not been considered for computational reasons), a half window width of 18 or 24 hours and restricting local maxima in variance to pass the 0.975 quantile. However, performance was not overly convincing, and parameter recommendation and testing was based on only a single AirBase time series, respectively, where furthermore breaks (and non-breaks) had been identified based on subjective judgement.

In the following paragraphs we first list the parameter values tested with selected time series from the Netherlands and Czech Republic where we had some information about events that potentially affected the time series. Finally we present exemplary results and draw conclusions.

1.2 Approach

Our approach is similar to the preceding section: We combine selected thresholds b with a number of half window widths q and the additional parameters of filtering iterations k . Again we are lacking ground truth data but were provided with some potential break dates for a set of stations where changes of the emission structure, of location in microscale or of the measuring device had taken place as outlined in [?]. Thus, at these points in time and for the given station a break can be suspected and is hypothesized (but does not need to have taken place). We do not have any information whether any further breaks have happened at any other time points.

In our analysis we considered the following parameter values:

Choice of threshold b [?] specify the expected number of detections $E(D)$ in a clean time series without any breaks in dependence on the number of observations (length of time series) n and the method's parameters half window width q and number of iterations k :

$$E(D) = \frac{n}{2q\sqrt{k}}$$

For a given tolerance of false positives and a given length of the time series, method's parameters could be chosen accordingly. Decreasing the tolerance, however, leads to more computationally expensive iterations or increased window widths such that breaks within the windows become hard to distinguish and accuracy of break location deteriorates. Instead of choosing suboptimal parameters based on a tolerance for false detections, only the most probable, i.e. the largest, most evident breaks could be investigated by introducing threshold parameter b . We chose the local maxima in variance to pass the 0.75, 0.95, 0.975 and 0.995 quantiles.

Choice of half window size q The choice of the window width parameter should depend on the minimum size of breaks of interest, expressed as a multiple of the time series' standard deviation, (the smaller the breaks the wider the window has to be chosen) or of the desired accuracy of locating the break (the more accurate the narrower the window has to be chosen). As we are both interested in obvious breaks and higher accuracy, the half window width was chosen to equal six month at most for daily and two months for hourly data. A window width of two months in the case of hourly data is 24 times longer than in the case of daily data (as is the time series in all). Short-time variability is smoothed out, while accuracy of locating a break is worsened by this factor, but is still within the same adequate range of days. Furthermore we can

expect the same amount of false positive detections. The minimum window width for daily data was set to equal two weeks. For hourly data we decided to consider also windows of seven days, but no narrower as we were less concerned about short-time discontinuities than about breaks that are rather persistent and of relevance for some longer time period or the entire time series.

In detail, for daily data the q -values considered are 7, 15, 30, 60 and 90, leading to window widths ranging from two weeks up to a maximum window of half a year. For hourly data the q -values considered are $3 \cdot 24$, $7 \cdot 24$, $15 \cdot 24$, $30 \cdot 24$, leading to window widths ranging from one week up to a maximum window of two months.

Choice of iteration k The smoothing effect of additional runs of the filters was investigated by choosing and combining iteration parameters k . We considered 3 and 10 iterations for the Kolmogorov-Zurbenko filter and 1, 3 and 10 iterations for the adaptive Kolmogorov-Zurbenko filter.

1.3 Results

We show results for daily PM_{10} measurements from selected stations in Czech Republic. ?? contains more figures on break detection for selected stations from Czech Republic. Complete results can easily be reproduced and investigated using our scripts that are available online and can be found at http://ifgi.uni-muenster.de/~epebe_01/ETC-ACM/subtask_1.0.1.2-5b.

The example data were presented in ?. Black triangles are used to mark a priori suspected breaks. In the time series plot in ? a change is clearly visible only at the suspected break time point in station CZ0UTEM/CZ0UTPM, and another one might be hypothesized at station CZ0TOPO. All other suspected breaks are rather questionable. When investigating deseasonalized time series only the suspected break for station CZ0UTEM/CZ0UTPM would probably get marked as such by merely looking at the time series (i.e. subjective judgement). In fact, applying Kolmogorov-Zurbenko adaptive filters do not reveal any breaks at the hypothesized time points but do so for the one in station CZ0UTEM/CZ0UTPM. This one is a special case in the sense that it is the start of a series of null measurements of the kind that were investigated in ?. We see these runs of equal (null) values as a special type of true transient breaks. Thus, we hope to find and precisely locate these break points with the Kolmogorov-Zurbenko adaptive filter.

First, we comment on the rational and the effect of deseasonalized data rather than original time series for use with Kolmogorov-Zurbenko (adaptive) filters. ?] considered time series spanning multiple decades. To prevent from overdetection large window sizes had to be chosen, such that as a side effect annual seasonality was smoothed out in their examples. On the other hand the large window widths resulted in poor accuracy of locating the breaks. The AirBase data we are exploring here span eleven years at most and more accurate results are desired. Recall that the window width had been restricted to a maximum length of half a year for exactly this reason. Figure 1.1 not only serves to demonstrate the steps in break detection but also shows differences in applying the method to original and deseasonalized time series.

We see that seasonality is preserved (together with trend and potential breaks) in the filter from the original time series. We observe a much higher amplitude in the winter 2009/10 in comparison with other years which in both cases gets marked as a transient break. Considering less obvious breaks (above 0.75 quantile threshold, depicted by grey dotted lines) with original data some of the winter periods of other years get marked, too, which is not an issue for deseasonalized data.

sonalized data. Instead some other minor discontinuities become visible here.

Even though the 0.75 threshold in combination with the other parameters investigated seems not to be high enough to prevent from overdetection (and for this reason and the sake of lucidity was excluded from any further plots) we recommend the use of deseasonalized data. We show another comparison in the end of this section when examining the impact of window widths.

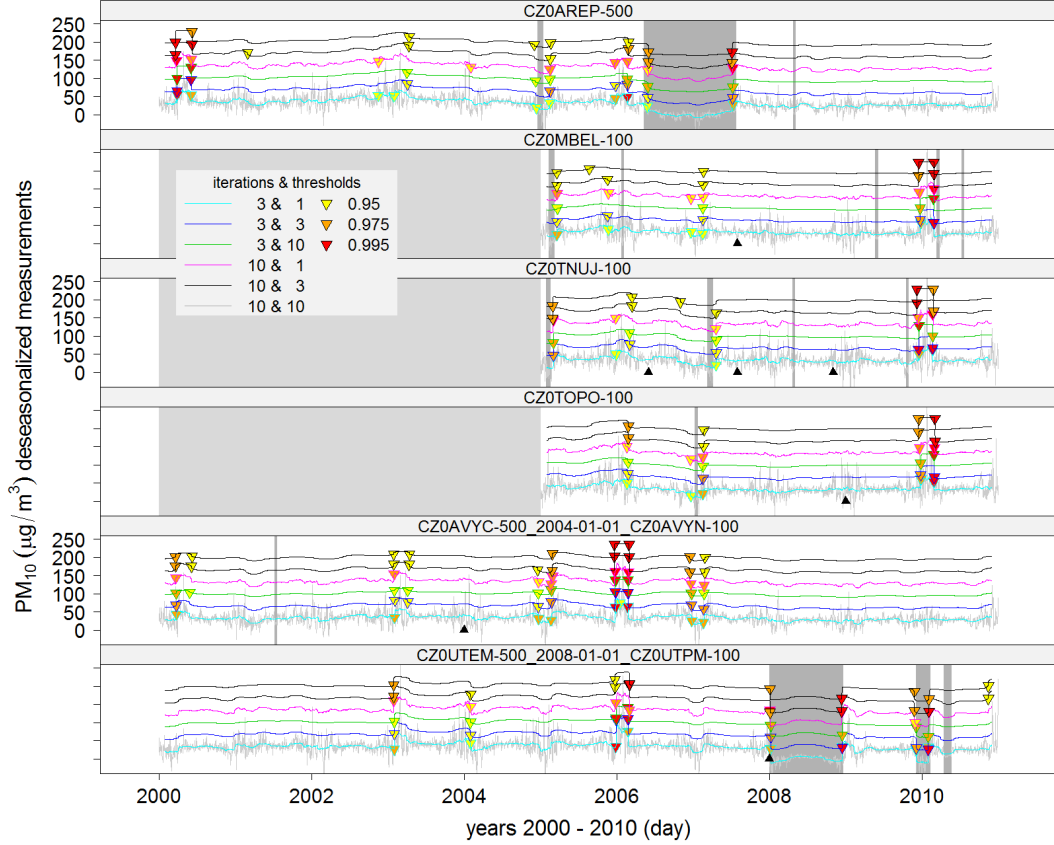


Figure 1.2: Findings of the Kolmogorov-Zurbenko adaptive filter for varying numbers of iterations, applied to deseasonalized data and with a window width of one month.

Next, we present results of comparing the Kolmogorov-Zurbenko adaptive filter with differing numbers of iterations both for the initial (kz) and the adaptive runs (kza). For a given window width of one month ($q = 15$) in Figure 1.2 the deseasonalized time series of Czech example stations are overlaid with the filter using three kz and one kza runs. All other combinations considered are depicted with additional offsets for the seek of graph readability. Identified break points are shown by triangles and are color coded with respect to thresholds b . Clearly, the higher the threshold the more evident is the suggested break.

More iterations are expected to result in lower precision of locating the breaks. This effect seems to be too small to be observed in Figure 1.2. The smoothing effect of additional runs can be seen both for the kz and especially for the kza runs when comparing the filters. It should lead to less false positives but might also dilute the power of detecting true positives to some extent. Without ground truth data we cannot judge on this. However, we observe that differing numbers of iterations seem not to result in big differences in the detected breaks. At least for

the most evident breaks, results coincide quite well, indicating robustness of the method with regard to the number of iterations.

We recommend to use three iterations of filtering runs such that some smoothing is applied to the data and breaks are sharpened while false positives still can be accounted for either by window width or even better through the threshold parameter. As an advantageous side effect less iterations finally minimize computing time.

Last, we show results of investigating the effect of varying window widths. For given iteration parameters ($k_{kz} = 3$ and $k_{kza} = 3$) Figure 1.3 and Figure 1.4 show the original and the deseasonalized time series of Czech example stations, respectively. Again they are overlaid with the filters, using window widths of two weeks up to six months, in this case. Identified break points are shown by triangles and color coded with respect to the thresholds.

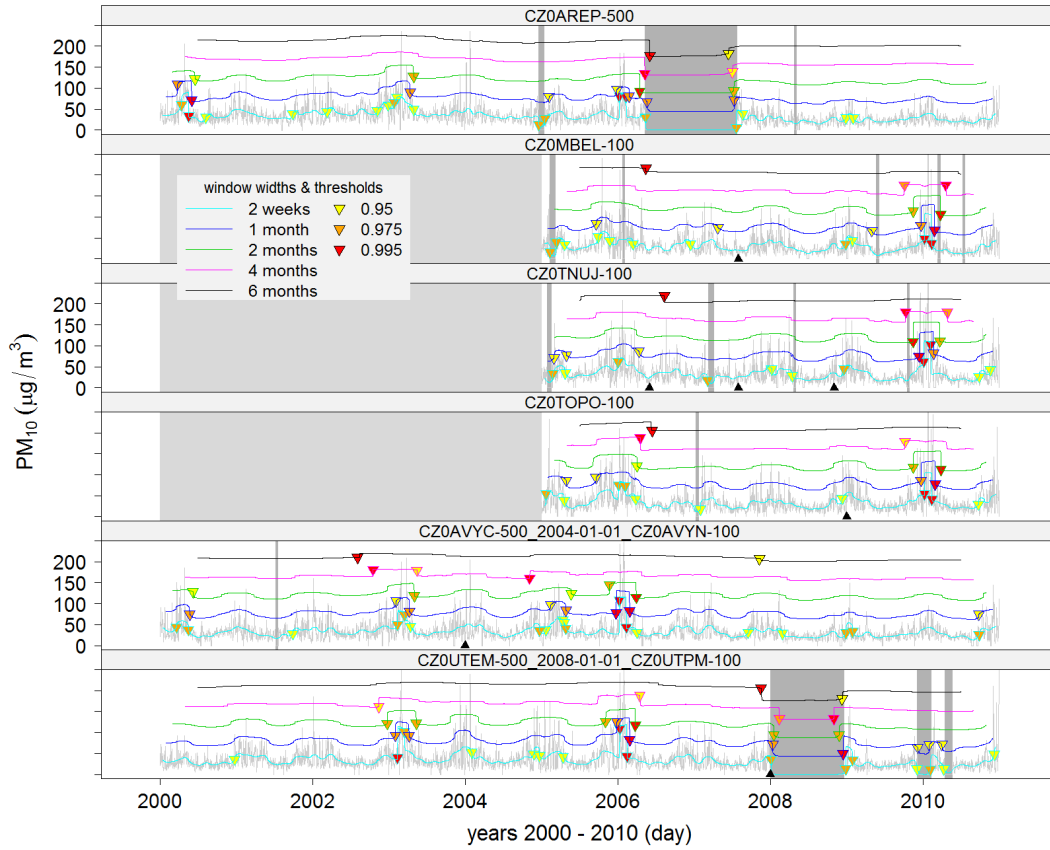


Figure 1.3: Findings of the Kolmogorov-Zurbenko adaptive filter for varying window widths.

Narrower window widths are expected to result in higher precision of locating the breaks on the one hand and in more false positive detections on the other hand. Both these effects are clearly visible from Figure 1.3 and Figure 1.4: considerably more detections occur for narrower windows as well as better localisation of transient breaks. We thus recommend to use narrow window widths in combination with high threshold values.

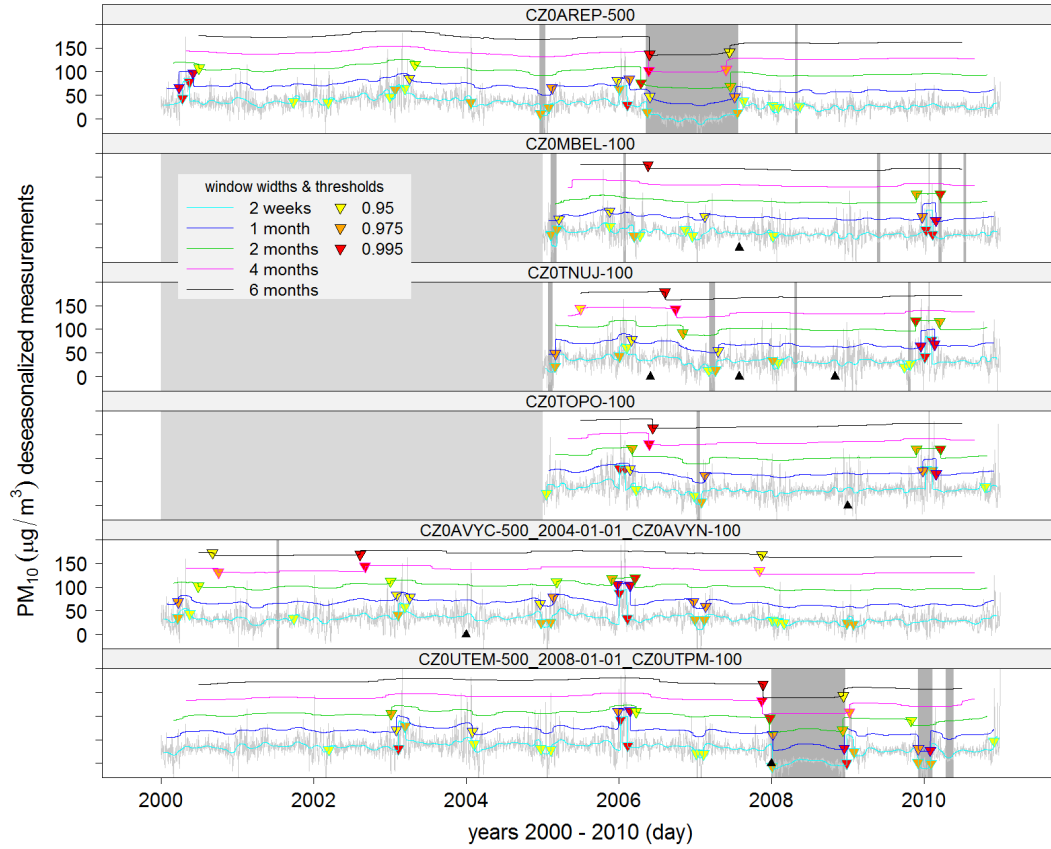


Figure 1.4: Findings of the Kolmogorov-Zurbenko adaptive filter for varying window widths when applied to deseasonalized data.

Window width comparisons for hourly data (PM_{10} , SO_2 and O_3 measurements from Czech stations) can be found in ???. One week filters look rather rough and several (short transient) breaks are identified at least with low thresholds. At times it seems that for very narrow windows the break detection rather becomes an outlier detection method. For PM_{10} and SO_2 data we regularly observed breaks across stations at the beginning of a year, especially for 2003 and 2006. For both pollutants, window widths of two weeks or one month seem to be adequate, just as for daily PM_{10} data. Ozone measurements exhibit very pronounced seasonality. We therefore highly recommend the use of deseasonalized data for this pollutant.

Summarizing our suggestions regarding parameter choice, we recommend to use three iterations each for the initial Kolmogorov-Zurbenko filter and the actual adaptive filter. We do not recommend to use window width longer than two months, but rather to choose narrow windows of widths up to e.g. one month and use rather high threshold values like the 0.975 and 0.995 quantiles when worried about false positives. We recommend the use of two or three threshold values to get an idea about how the relevance of detected breaks could be ranked.

When applying the Kolmogorov-Zurbenko (adaptive) filter with window widths that do not span the largest periodicity within the data, deseasonalized measurements should be used in order to reduce false positive break detections. Even with deseasonalized data some breaks, especially transient changes within periods of typically higher or lower measurements, might rather be due to physical reasons (e.g. particularly cold winters or particularly hot and dry summer seasons) than due to errors or changes in the measuring device. Transient changes in general can also be

seen as an accumulation of (potentially reasonable) outliers. In order to keep such false detections away, instead of deseasonalizing the data, prior to analysis, we recommend to adjust data with respect to the variables that seasonality is a substitute for (e.g. weather conditions). Consequently, natural expected breaks are accounted for and would not get marked as breaks any more. Another possibility would be a comparison between neighboring stations or all stations in a region with similar (weather) conditions. Figures [1.2](#), [1.3](#) and [1.4](#) show quite a number of such coincidences across stations, questioning their classification as undesirable breaks.