Article

Machine Learning Method for Changepoint Detection in Short Time Series Data

Veronika Smejkalová [1], Radovan Šomplák [1,*], Martin Rosecký [2] and Kristína Šramková [3]

1 Faculty of Mechanical Engineering, Institute of Process Engineering, Brno University of Technology, Technická 2896/2, 616 69 Brno, Czech Republic; veronika.smejkalova1@vutbr.cz
2 Czech Math, a.s., Šumavská 416/15, 602 00 Brno, Czech Republic; martin.rosecky@czechmath.com
3 Faculty of Mechanical Engineering, Institute of Mathematics, Brno University of Technology, Technická 2896/2, 616 69 Brno, Czech Republic; kristina.sramkova@vutbr.cz
* Correspondence: radovan.somplak@vutbr.cz

Abstract: Analysis of data is crucial in waste management to improve effective planning from both short- and long-term perspectives. Real-world data often presents anomalies, but in the waste management sector, anomaly detection is seldom performed. The main goal and contribution of this paper is a proposal of a complex machine learning framework for changepoint detection in a large number of short time series from waste management. In such a case, it is not possible to use only an expert-based approach due to the time-consuming nature of this process and subjectivity. The proposed framework consists of two steps: (1) outlier detection via outlier test for trend-adjusted data, and (2) changepoints are identified via comparison of linear model parameters. In order to use the proposed method, it is necessary to have a sufficient number of experts' assessments of the presence of anomalies in time series. The proposed framework is demonstrated on waste management data from the Czech Republic. It is observed that certain waste categories in specific regions frequently exhibit changepoints. On the micro-regional level, approximately 31.1% of time series contain at least one outlier and 16.4% exhibit changepoints. Certain groups of waste are more prone to the occurrence of anomalies. The results indicate that even in the case of aggregated data, anomalies are not rare, and their presence should always be checked.

Keywords: machine learning for time series; waste generation; short time series; anomaly detection; outlier; changepoint

1. Introduction

Theoretical procedures for making forecasts are already known and described in detail. An example is the extensive theoretical and practical overview given by Petropoulos et al. [1]. However, each forecast must be approached individually with regard to the character of the data (scope, probability distribution, etc.). A crucial part of the forecast is data pre-processing [2]. It is appropriate to consider a possible transformation of the data with respect to heteroskedasticity [3]. Another important point of the forecast is the identification of anomalies, i.e., general strangeness in the data series.

An example of anomalies in data is shown in Figure 1; it is chosen to resemble waste generation datasets. Obviously, the data in Figure 1 show signs of anomalies. When an outlier is identified, this point is removed for further work with data. In the case of step changepoint and trend changepoint in the data, respectively (see Section 1.2), only a part of the time series after this anomaly is considered further. When comparing the common linear trend, significant differences in future forecasts can be seen in Figure 1 (blue: 2010–2018, red: 2014–2018, and green: 2015–2018) according to the identified anomalies. Identifying outliers in the first step is crucial but not always straightforward. Subjectively, it can be determined that there is an outlier value in 2014 and that the production trend has changed since 2015 (green). It can be assumed that the trend change is a part of cyclicality, and a

decline may be followed by an increase in the future. In the case of cyclicality, the data would oscillate around the trend shown by the blue line. However, cyclicality cannot be clearly demonstrated in such a short time series. If the value in 2014 was not evaluated as an outlier, the changepoint between 2013 and 2014 can be identified (red). The authors are convinced that the most probable upcoming trend is the green line (data 2015–2018). This hypothesis of the authors was subsequently tested in Section 3 after the introduction of new methods of identification of anomalies in data. However, this opinion may not prove to be correct in the future, and it is necessary to revise it with new data. This is a major contribution of this contribution.
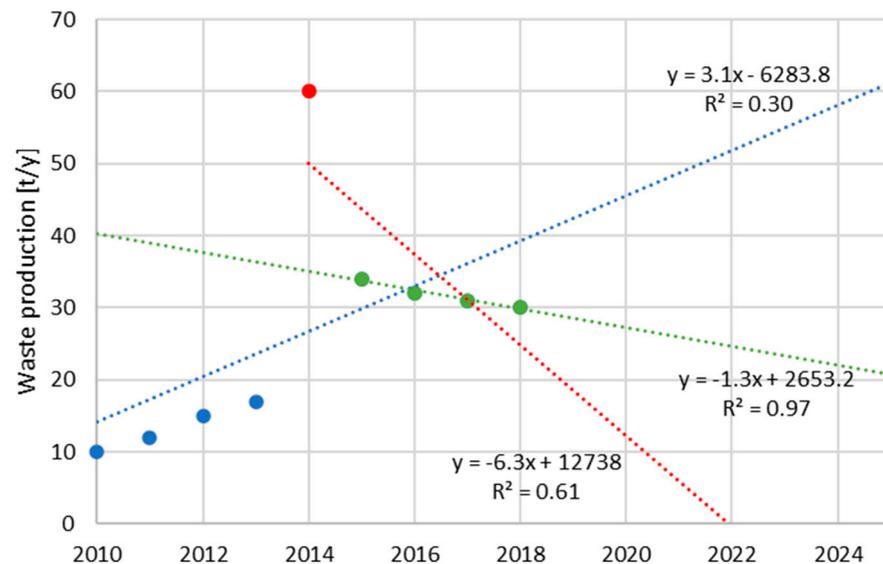
Figure 1. Example of problematic changepoint detection due to outlier presence.

The example in Figure 1 shows a situation based on actual waste generation data. As mentioned above, the individual colors of the linear trends correspond to different approaches to anomalies in the data. It is therefore evident that the approach to anomalies can significantly influence the forecast of the generation of waste production. This type of forecasting is essential for planning in waste management [4]. To be able to handle waste adequately and fulfill the legislation set by the EU, it is necessary to build the necessary infrastructure [5]. In the case of a low-quality forecast, there is a risk that the infrastructure will be insufficient or excessive capacity. The direct consequence is then financial losses and non-fulfillment of the legislative goal. The purpose of the analysis of anomalies in the historical waste management data is the pre-processing of the data so that quality forecasts can be made.

Different kinds of waste are analyzed quite often to support decision-making in waste management (WM) and subsequent use of waste. In some cases, WM data are available on a daily or weekly basis, mainly for larger cities like Zagreb [6], Helsinki [7], or Tehran and Mashad [8]. However, annual data are still more common, especially for higher territorial levels. In such a case, the number of data points is low, e.g., for the countries of the European Union, data was available for 23 years [9]. However, longer time series are often not adequate as data collection methodologies change. Small datasets are usually difficult to deal with, as common statistical tools aim for a sample size of 30 or more. In all cases, accurate forecasts and understanding of waste generation are crucial for further steps and their real-world implementation. Study [6] focused on short-term forecasts, while the main goal of Kannangara et al. [10] or Niska and Serkkola [7] was to understand the socio-economic impacts on waste generation. However, even in studies like [11] or [12], when waste generation is not analyzed directly, WM data quality is important to achieve reliable results, e.g., when e-waste is of interest [13], it is easy to imagine that rapid development in a number of e-waste devices can introduce anomalies into data. The same is even more

important outside of municipal solid waste (MSW), where waste generation can behave much more irregularly across the years. An example of such a case is agriculture-related waste processing and optimization [14]. Even more stable kinds of waste, like soil waste [15], can be affected by mine closures and openings. It is known that anomalies (outliers or changepoints) are present in WM datasets, at least in some cases. Causes of these problems are such as wrong entry (outlier) or structural change (changepoint). Structural changes can be caused by legislative or other changes in the WM system [16]. Although changepoint detection in time series is quite a common topic in the scientific literature, see, e.g., [17], no paper dealing with this issue in the WM field was found.

No automatic anomaly detection method can outperform an expert, especially In the case of short time series. An expert can even find the reason for anomaly due to domain knowledge. However, a large number of waste types that need special treatment exist (e.g., various kinds of hazardous waste). In case multiple types of waste are analyzed for many territories, the number of such short time series is growing fast. It is not possible to examine, e.g., tens of thousands of time series by an expert. Automatic detection of anomalies is thus a needed but complicated task. In order to identify changepoints (i.e., step changes and trend changes) in the data, it is advisable to remove the outliers first. There are a number of methods for outlier analysis. However, data on waste production have their specific character. Based on a review of existing methods, suitable methods were selected and tested on representative samples. The aim is to select an appropriate procedure for identifying outliers, considering this particular data type, Section 2. In the next step, and this is the main contribution of this paper, the available approaches for identifying changepoints are assessed (Section 2.3). The existing approaches were compared with the approach proposed by the authors of this paper, and their performance was evaluated. The new approach was further applied in the case study to all wastes produced in the Czech Republic for the period 2010–2018 (Section 3).

1.1. Outliers

Even before anomaly detection, some forms of data pre-processing should be considered [1]. These include the removal of external effects (e.g., the COVID pandemic is likely to introduce a significant increase in the amount of packaging waste), aggregation of waste types/territories (e.g., when various waste types are treated as the same or only higher territorial unit is of interest), redefinition of variable of interest (e.g., amount of separated waste to ratio of the separated to total amount of waste) or data transformation (e.g., logarithmic). These can both reduce the number of anomalies and help to answer the right questions. However, these adjustments are not trivial since, e.g., a suitable proxy for external effect is needed, and it is not commonly available (or it is available only for higher territorial units).

Outlier detection is a topic that has attracted a lot of interest in recent years in general, e.g., [18], as well as in a time series context. Only general time series works will be discussed here since no paper focused specifically on short-time series was found. Braei and Wagner [19] focus on univariate time series only and offer a classification of methods into three groups: (1) statistical methods, (2) classical machine learning methods, and (3) deep learning methods.

Choi et al. [20] aimed toward multivariate time series, which are more common nowadays. It mentions the challenges of classical approaches, namely the lack of labels and complexity of data and explains the benefits of deep learning approaches. These include the possibility of establishing relationships between variables and modeling of temporal context. This paper mentions three possible ways to obtain an anomaly score from deep learning methods: (1) reconstruction error, (2) predictive error, and (3) dissimilarity.

Blázquez-García et al. [21] deals with both univariate and multivariate case. The mentioned univariate and multivariate methods are basically in agreement with already mentioned reviews [19,20]. However, the usage of univariate methods for multivariate cases is also discussed. The main direction in how to leverage well-developed univariate methods

is to first apply a dimensionality reduction technique (e.g., PCA—principal component analysis) and then one of the traditional methods. This method is able to compare the municipalities in individual years and identify outliers. It is, therefore, not a question of evaluating outliers in a time series but rather of identifying localities with anomalous behavior. This approach is not suitable for the application required in this study, where outlier removal is to precede changepoint detection.

Most of the current research in outlier detection in temporal data aims to solve problems involving multivariate datasets of enormous sizes. In such a case, data are collected automatically, and sometimes real-time processing is needed. Chalapathy and Chawla [22] provide examples of these applications, e.g., intrusion detection, fraud detection, malware detection, medical anomaly detection or industrial anomalies detection. Deep learning approaches are usually successful in these cases. However, automatic (and high-frequency) data collection is still not common in WM. Insufficient lengths of individual time series strongly restrict the possible range of methods in such a case.

A different view is offered by an extensive review [23], which recognizes these groups of approaches: classification, nearest neighbor, clustering, statistical techniques, information–theoretical and spectral techniques. Statistical techniques use standard statistical tests (e.g., Grubbs's, Dixon's or Rosner's test, z-score, boxplot rule) for stationary time series. For non-stationary data, the trend-removing model is fitted, and the residuals of this model are subsequently analyzed. To detect outliers in residuals, some arbitrary threshold or model-related measures like the Akaike information criterion are commonly used. The last group of statistical approaches can be called distribution estimation-related. These methods assume either that a sample is a composition of multiple known parametric distributions (multiple "normal" distributions or "normal" and "anomalous" mixtures) or try to estimate the distribution based on the data using non-parametric methods (histograms or kernel density estimation). Information–theoretical approaches (using, e.g., Kolmogorov complexity, entropy, relative entropy) use a sliding window. Spectral techniques use methods like principal component analysis, and they try to capture the bulk of variability in the data by combining the data attributes.

Outlier detection techniques used in WM include the boxplot rule [10], exclusion of top and bottom deciles [16] and z-scores [24]. However, Kannangara et al. [10] do not take time development into account and the usage of deciles [16] is likely to be too rough. Rybová et al. [24] deal only with one year of annual data (so the time component is not relevant). For the detection of outliers, an approach combining common methods is presented in Section 2.1.

1.2. Changepoints

The term changepoint is used in this paper to refer to two types of anomalies: (1) step changepoints; and (2) trend changepoints.

Step changepoints are sudden changes at a certain point in the time series. These changes may include changes in the average or other parameters of the process that create the time series. Thanks to the detection of step changepoints in the data, it is possible to find out when the structure of the time series changes and thus better understand the analyzed problem. Various methods make it possible to find out whether there are significant step changepoints in the time series and also determine their position. Incorrect identification of step changepoints in the data can lead to erroneous conclusions and inaccurate prediction models. The example of the step changepoint is shown in Figure 2a; the position of the step changepoint is between the years 2012 and 2013, whereas the year 2012 will be called the beginning of the step changepoint and the year 2013 as the end of the step changepoint.

Trend changepoints are sudden changes in the behavior and properties of the observed series. Specifically, a trend changepoint indicates a situation with a significant change in the tangent. An example of the trend changepoint is shown in Figure 2b, where the trend changepoint is observed in 2013.
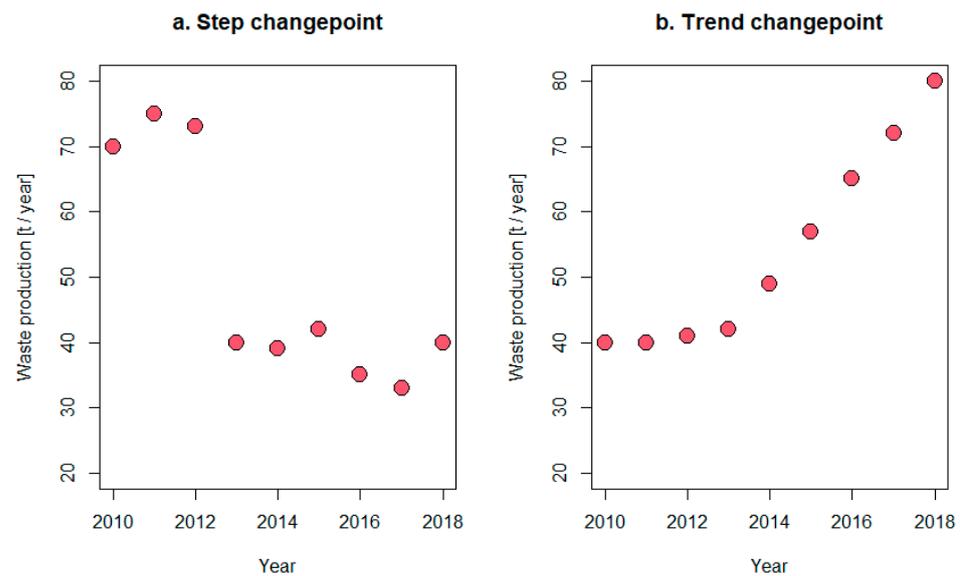
Figure 2. Examples of changepoints: (**a**) step changepoint; (**b**) trend changepoint.

In the literature, these two kinds of changepoints are sometimes analyzed separately. According to Truong et al. [25], methods for changepoint detection can be divided into two main branches: (1) online; and (2) offline. Online methods are based on real-time detection of changes. Offline methods detect changes in the entire dataset at once, so they retrospectively identify the location of abrupt changes. For the purposes of this paper, a more detailed description of offline methods with machine learning principles will be sufficient.

1.3. Machine Learning in Changepoint Detection

According to the extensive overview of methods by Aminikhanghahi and Cook [17], machine learning algorithms can be divided into (1) supervised; and (2) unsupervised.

Supervised methods are machine learning tasks where the algorithm analyses the training dataset, which consists of pairs of input and output variables. The goal of the algorithm is to learn the mapping function from the input to the output. When the supervised methods are applied to the problem of changepoint detection, algorithms are trained as classifiers (binary or multi-state). After determining the number of states and state boundaries, the methods work on the principle of a sliding window passing through the time series and looking for the occurrence of changepoints [17]. A summary of possible multi-class classifiers: decision trees, nearest neighbor, support vector machine, Naïve Bayes, Bayesian net, hidden Markov Model, conditional random field and Gaussian Mixture Model can be found in the study [17]. Also, an overview of binary class classifiers is offered [17]: support vector machine, Naïve Bayes and logistic regression. A more detailed review of online changepoint detection methods is offered in the study [17,26]. Supervised machine learning approaches are usually fairly accurate models with simple creation. The main disadvantage is the dependence on the quality of the training data.

Unsupervised methods are algorithms that discover hidden patterns based on statistical features, not data labeling. The typical approach is to consider probability distributions from which data in the past and present are generated. These two distributions are statistically tested to determine whether they are equal or significantly different. This kind of approach is based on the likelihood ratio. This is the ratio between two probability densities calculated in two consecutive intervals [27]. The second line of approach is subspace identification; this is based on the analysis of subspaces in which time series sequences are constrained [28]. The last approach is called probabilistic methods, which is divided into two lines: (1) Bayesian; and (2) Gaussian. The assumption of Bayesian methods is that a time series may be divided into non-overlapping state partitions, and the

data within each state are from some probability distribution. Compared to the likelihood ratio methods, Bayesian methods consider not only pairs of consecutive intervals but all previous intervals [29]. In the Gaussian methods (also called the Gaussian process), time series observations are defined as a noisy version of Gaussian distribution function values. Gaussian process function is used to make a normal distribution prediction at time t using observations available through time (t − 1). Then, the *p*-value is calculated for the actual observation under the reference distribution. The α-threshold is used to evaluate the *p*-value, and an algorithm determines whether the actual observation does not follow the predictive distribution, which indicates a changepoint [30]. A more detailed review of offline changepoint detection methods is offered in studies [17,25].

It should be noted that changepoint detection is usually done on much longer time series (hundreds or thousands of data points). The use of common methods for annual data (e.g., in WM) is therefore very limited.

1.4. Summary of Literature Review and Novelty

No paper dealing with changepoint detection in WM was found. Thus, this paper aims to propose a complex framework for changepoint detection in WM time series.

Generally, there is little to no attention to outlier detection in short time series (say, length < 30). Outliers are made using a combination of known methods. Previous studies have not addressed this approach, and this is the first point of novelty in this study. Different approaches to outliers were tested, and the solution that achieved the best results was recommended (Section 2.1).

A suitable method was not found for the detection of changepoints on WM data, which is also proven by testing in Section 2.5. With regard to the inappropriateness of using existing approaches to the WM data, a completely new approach to the analysis of changepoints is presented, taking into account the specific nature of data in the WM area (Section 2.3). The main goal is to develop an approach that allows (1) automated updating of results of changepoint detection and (2) learning of the model with the help of new expert evaluations. The proposed framework for changepoint detection can be helpful to everybody dealing with WM data. Waste generation data are essential for WM infrastructure, capacity allocation and shift to the circular economy.

2. Material and Methods

This section introduces methods used for both outlier and changepoint detection. The methods were tested on waste production data from the Czech Republic, dataset includes annual data for the period 2010–2018. Waste production in the Czech Republic is registered under catalog numbers, which are classified into 20 groups according to the origin of the waste. Waste group 3 (waste from wood treatment) and waste group 20 (municipal solid waste) were chosen for testing. Waste group 20 consists of 15,320 time series and represents waste with a relatively stable trend. Conversely, group 3 consists of 737 time series and represents highly variable waste production (see Appendix A). Since the proposed methods incorporate expert judgment, smaller subsets were used for testing data. As already mentioned in Section 1, little to no work was done in anomaly detection in WM. Thus, the changepoint detection framework is a very important tool for data analysis in WM.

Data transformations can even improve the properties of the time series under examination. On the other hand, commonly used transformations like logarithms can bring in some problems if zero waste generation is possible for a particular waste type. In such a case, zeroes need to be replaced, but the selection of the right value for replacement can be a difficult task with a big impact on subsequent analysis. Thus, none of these adjustments was used here. This problem remains an option for future research.

The proposed methodology for anomaly detection consists of two subsequent steps:

1. Outlier detection (Sections 2.1 and 2.2);
2. Changepoint detection (Sections 2.3–2.5).

Outlier detection is a necessary prerequisite for subsequent changepoint detection. A combination of common approaches to data processing is used for outlier detection. A new method, including a machine learning approach, was developed for changepoint detection; therefore, a significantly larger space is devoted to this part.

Unless otherwise stated, all of the computations, data manipulation and visualization were conducted via R software (R Core Team, 2021, [31]) with default parameter setting.

2.1. Outlier Detection—Methods Description

This section provides a short description of methods selected for testing based on review. Some types of models for detecting outliers are suitable for a specific type of data. We can mention, for example, the widely used semantic models that cannot be directly used for the issue of time series [32]. The first group of considered methods for short time series consists of fitting a simple model and subsequent analysis of residuals. The Holt method was selected as suitable for trend removal. Since it is generally slightly more flexible than, e.g., linear fit, but not as flexible as polynomial or nonlinear fits. Such a balance is needed to avoid significant under or overfitting. Note that in this step, the main goal is not to fit the data perfectly but rather to estimate and remove general trends from the data to use for stationary data methods. Three methods were selected for the analysis of residuals, namely the Dixon test, Grubbs test and z-score. The elected tests (Grubbs and Dixon) should be able to deal even with small datasets. Z-score is a common method to compare residuals with a normal distribution, which uses quantiles of normal distribution to assess possible outliers. "Far away" usually means two or three standard deviations from the mean, which corresponds to the probability of observing at least such a distant point with a probability of 4.6 and 0.27%, respectively (based on the assumption that data come from normal distribution). However, this method is more of an unwritten rule than an exact test. In contrast, both the Grubbs and Dixon tests are exact tests used for outlier detection. These tests should be suitable even for small datasets, which are very beneficial in our problem setup, see [33,34].

Other common methods selected for testing (LOF, GLOSH and kNN distance) are more or less connected to the notion of density. The kNN distance technique for outlier detection consists of first creating a kNN graph and then. The anomaly score is defined as the distance to the k-th nearest neighbor. The main idea of LOF is also based on the kNN technique (Breunig et al., 2000) [35]. It uses reachability distance, which is the actual distance of two points (e.g., A and B). However, if the points are close enough (i.e., B lies within a radius defined by distance from A to its k-th nearest neighbor), the distance to k-th nearest neighbor (so-called k-th distance) is used instead of the actual distance. LOF measures local reachability density (i.e., the inverse of the average reachability distance of the object from its neighbors). Then, the LOF is the average local reachability density of the neighbors divided by the object's own local reachability density.

Finally, GLOSH is the method that unifies both the global and local flavors of the outlier detection problem into a single definition of an outlier detection measure [35]. Starting from the usual assumption that there are one or more data-generating processes deemed nonsuspicious and noticing that clusters are natural candidates to model such generator(s). The scope of the reference set can be adjusted for each object based on the closest cluster (in a density-based perspective) within the density-based hierarchy. Therefore, hierarchical DBSCAN (density-based spatial clustering of applications with noise) or HDBSCAN (hierarchical DBSCAN) clustering is done first. Then, $\epsilon(x_i)$ is the lowest radius at which x_i still belongs to its cluster (and below which x_i is labeled as noise). $\epsilon_{max}(x_i)$ is the lowest radius at which this cluster or any of its subclusters still exist (and below which all its objects are labeled as noise). The GLOSH score is then defined as follows:

$$GLOSH(x_i) = 1 - \frac{\epsilon_{max}(x_i)}{\epsilon(x_i)} \tag{1}$$

Both time and waste production were rescaled to [0,1] intervals for every time series to allow a general setting of thresholds for kNN distance, LOF and GLOSH. As in the case of the z-score, these techniques are not statistical tests, so a suitable threshold needs to be found for each specific problem. The general recommendation for the kNN distance threshold is quite difficult, so it is recommended to explore the histogram of values and iteratively adjust the threshold. For LOF, points with a score "larger" than 1 can be considered outliers. However, in some cases, 1.2 is large enough; in other cases, 3 is not. Scores are from the range [0,1] in the case of GLOSH, while values close to 1 are suspicious.

Threshold identification

The Dixon and Grubbs tests are used in the usual way with a set significance level (a common value of 0.05 is used here). For other methods, a suitable threshold has to be determined. It is necessary to find a suitable threshold for other considered approaches (z-score, kNNd, LOF, and GLOSH). There are general recommendations for setting values, but no specific values are available. For the z-score, the limit value of 3 is most often used, sometimes 2, but these values were not suitable for the tested data. So, the limit values were adjusted for the z-score as well. First, outliers are detected using the selected algorithm. The following procedure was applied to verify the algorithm (A) and find a suitable parameter setting (B):

A. Verification of algorithmic solution based on visual assessment:

  1. Representative examples were selected for individual types of waste for expert assessment. For each type of waste, 10 suspicious series were selected (if available), which, according to the authors, represented the largest possible spectrum of anomalous cases. In total, 230 time series were assessed. This procedure should contribute to a more general parameter setting;

  2. Five experts independently evaluated the outlying severity on a scale of 1 (certainly not)–4 (certainly yes). The proposed procedure can be specified in the future by greater involvement of expert knowledge and more sensitive treatment of individual waste fractions.

B. Setting the parameters of algorithms considering expert judgment:

  1. The expert evaluation from point 2 was used to set the limit value of each approach, from which the observation is already considered outlying.

The threshold was set as the sample for the z-score, as shown in Figure 3. First, the median of the expert evaluation was calculated for each evaluated series. Due to the odd number of experts, categories 1–4 were maintained. Subsequently, the median rating was compared with the criterion value of the suspected point. A graphic representation is provided in Figure 3. The resulting limit value of 2.17 was determined as the median of the z-score for the median expert evaluation (e_med = 3). It is also easy to see from Figure 3 that the differences in medians between originators are typically not very large; therefore, it is possible to use a universal parameter for both originators.

2.2. Outlier Detection—Performance Evaluation

As already mentioned, rather simple methods were selected for outlier detection with regard to the length of the investigated time series. These include the procedure consisting of data fitting by the Holt method (trend) and the subsequent analysis of residuals (Dixon test, Grubbs test, z-score). These steps are repeated iteratively if an outlier was found in the previous iteration. Other common techniques like LOF, GLOSH and kNN distance were also considered. A total of six variants of methods will be tested. Subsequently, a visual assessment of the success of the approach for outlier search was performed.
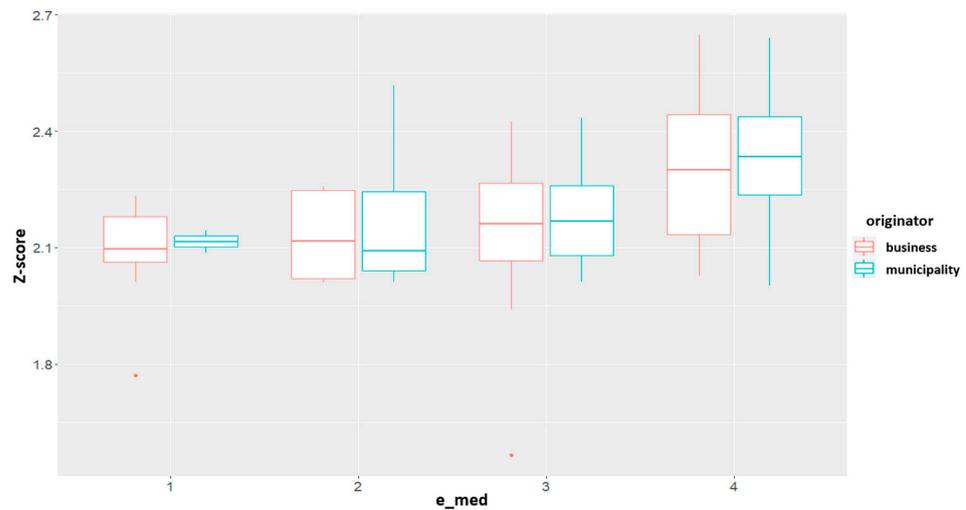
Figure 3. Boxplot of z-scores for the Holt method by the median of expert evaluation and originator. Remark: Dots indicate outliers.

For the evaluation of the applied methods, 100 time series for group 3 and 200 time series for group 20 were selected at random. The presence of outliers was evaluated by visual assessment and then compared with individual methods. The results are presented pointwise since pointwise outlier detection is of interest here (see Table 1). Otherwise, cases like the wrong location (correctly identifying that time series contains outlier but at the wrong place) or partial success (either not all of the present outliers are identified, or a higher number of outliers is predicted) would occur. In total, 1100 points of group 3 and 2200 points of group 20 were assessed (11 points in each time series). The points were divided into four groups:

- True positives (TP)—the outlier was correctly identified;
- False positives (FP)—the outlier has been identified, but based on visual assessment, it does not occur,
- False negatives (FN)—the outlier has not been identified but based on visual assessment, it occurs;
- True negatives (TN)—the outlier has not been identified, nor does it occur in the data,

Table 1. Accuracy of methods for outlier detection sampled from the set of time series.

		TP	FP	FN	TN	Precision	Recall	F1	GM	Jaccard
	Holt + Grubbs test	**45**	**18**	**15**	**1022**	**0.71**	**0.75**	**0.73**	**0.84**	**0.58**
	Holt + Dixon test	44	8	16	1032	0.85	0.73	0.79	0.91	0.65
	Holt + z-score	47	19	13	1021	0.71	0.78	0.75	0.84	0.60
Group 3	LOF	39	15	21	1025	0.72	0.65	0.68	0.84	0.52
	GLOSH	41	6	19	1034	0.87	0.68	0.77	0.93	0.62
	kNNd	42	7	18	1033	0.86	0.70	0.77	0.92	0.63
	Holt + Grubbs test	**104**	**39**	**19**	**2038**	**0.73**	**0.85**	**0.78**	**0.85**	**0.64**
	Holt + Dixon test	90	30	33	2047	0.75	0.73	0.74	0.86	0.59
Group 20	Holt + z-score	106	49	17	2028	0.68	0.86	0.76	0.82	0.62
	LOF	91	36	32	2041	0.72	0.74	0.73	0.84	0.57
	GLOSH	82	20	41	2057	0.80	0.67	0.73	0.89	0.57
	kNNd	97	46	26	2031	0.68	0.79	0.73	0.82	0.57

Remark, source [36]: Precision (PPV)—proportion of positive samples that were correctly classified to the total number of positive predicted samples: $PPV = \frac{TP}{FP + TP}$. Recall (TPR)—the ratio of the correctly classified negative samples to the total number of negative samples: $TPR = \frac{TN}{FP + TN}$. F1—harmonic mean of precision and recall: $F1 = \frac{2PPV \times TPR}{PPV + TPR}$. GM—measure for balanced and imbalanced data: $GM = \sqrt{TPR \times TNR}$, where $TNR = \frac{TN}{FP + TN}$. Jaccard—measure ignores the correct classification of negative samples: $Jaccard = \frac{TP}{TP + FP + Fn}$.

The values in Table 1 indicate the number of points assigned to the TP, FP, FN or TN groups. The situation is summarized by multiple performance measures suitable for imbalanced classification tasks [36].

A positive finding is that there is a low probability of the occurrence of FP and FN. Such cases occurred in less than 3% of points. None of the methods was found to be significantly more successful than the others. The choice of a particular method may be influenced by the data being analyzed, as each method has different advantages. A combination of Holt's analysis and the Grubbs test was selected for subsequent steps because outlier detection is a necessary precondition for changepoint detection. The reliability of outlier detection can be improved by an ensemble of individual methods via a classification model. Nevertheless, the impact of the choice of outlier detection method should be negligible for changepoint detection performance based on method testing summarized in Table 1.

2.3. Changepoint Detection—Method Description

During a preliminary phase, some of the existing (and implemented) methods were tested (see Section 1.3), particularly the bcp and bps methods. The comparison with the new method presented in this contribution is summarized in Table 3, Section 2.5. The existing methods do not seem to be suitable for a short time series. Thus, a new method has been compiled based on known statistical approaches and modified to suit the length of the currently available time series. Detection is based on linear regression, specifically on interleaving data with a straight line and subsequent pattern recognition. In this section, a general approach to the method is described.

Before starting the changepoint detection calculation, the outlier is replaced by the previous point. In case the first point is detected as an outlier, it is replaced by a subsequent point. Equidistant time series of the same length are assumed. Thus, missing values need to be imputed, or only the complete part of the time series can be analyzed. In case a time series of different lengths is analyzed, time rescaling is needed. In the next step, the data are rescaled by dividing by the maximum value. Rescaling is incorporated to allow general parameter setting since various waste types are commonly of different magnitudes. Otherwise, a parameter setting will be needed for each waste type. The algorithm consists of four steps (A–D), which are further described. This is a description of the calculation approach; the evaluation of the presence of a changepoint according to the results of points A–D is described in Section 2.4.

A.   Coefficient of determination (R^2)

In the first step, the whole time series is fitted using linear regression. The quality of this fit is judged by the coefficient of determination (R^2). Based on R^2, the following rules are applied:

- The case data fit very well with linear regression; the presence of an anomaly is not expected.
- In case the fit is poor, the data are likely noisy. In such a case, changepoint detection is tricky, so it is not performed. Some naïve forecasting approaches will be applied.
- In case the fit is neither almost perfect nor very poor, changepoint detection is performed.

B.   Angles

Subsequently, for each pair of consecutive points, a straight line of the form $y = kx + q$ is fitted, and parameters k, q are stored. Moreover, angles between these lines are computed in every inner point of the time series. Based on the computed angles, a point is flagged as suspicious if the angle in this point lies within the specified critical range. If the angle is small, a rapid change followed by another one in the opposite direction occurs. When the angle is large, there is little to no change. None of these situations is considered to be a changepoint.

C.   Slopes of lines

The first round of validation is done via the following steps:

- Let x_P be a "suspicious" point;
- Fit a straight line connecting points x_{P-1} and x_{P+1};
- Fit a straight line connecting points x_{P-2} and x_{P+1};
- Fit a straight line connecting points x_{P-1} and x_{P+2};
- Fit a straight line connecting points x_{P-2} and x_{P+2};
- Slopes of all these lines are recorded, and the following rules are checked:

Sings of slopes of all 4 lines are the same (false detection of changepoints in "teeth" of time series is avoided by this rule);

Angles between these lines and the *x*-axis must be bigger than the specified critical value.

D.   SMAPE (symmetric mean absolute percentage error)

After the first round of validation, which flags suspicious points, the second round of validation is done. In the second round, three scenarios are examined: (1) trend changepoint in x_P; (2) step changepoint between x_{P-1} and x_P; and (3) step changepoint between x_P and x_{P+1}. Validation proceeds as follows (let $\{x_1, \cdots, x_n\}$ be points of the time series):

- Trend changepoint
  - ○ The time series is split into two parts—(1) part $\{x_1, \cdots, x_P\}$ and (2) part $\{x_p, \cdots, x_n\}$;
  - ○ The linear model is fitted, and *KRIT_smape* is computed for both separate parts. The criterion *KRIT_smape* (Equation (2)) is based on the statistics SMAPE. A_i is *i*-th actual value of the time series, F_i is value fitted by linear model for *i*-th point of time series, $mean_{c1}$ is the mean of the first part of the time series and $mean_{c2}$ is the mean of the second part of the time series.

$$KRIT_{smape} = \frac{\frac{1}{p}\sum_{i=1}^{p}\frac{|x_i - F_i|}{(|x_i| + |F_i|)/2}}{\frac{\max\{mean_{c1}; mean_{c2}\}}{\min\{mean_{c1}; mean_{c2}\}}} \tag{2}$$

- Step changepoint (backward)
  - ○ The time series is split into two parts—(1) part $\{x_1, \cdots, x_{P-1}\}$; and (2) part $\{x_p, \cdots, x_n\}$;
  - ○ The linear model is fitted and *KRITsmape* is computed for both separate parts (see Equation (2)).
- Step changepoint (forward)
  - ○ The time series is split into two parts—(1) part $\{x_1, \cdots, x_P\}$; and (2) part $\{x_{p+1}, \cdots, x_n\}$,
  - ○ The linear model is fitted and *KRITsmape* is computed for both separate parts (see Equation (2));
  - ○ Finally, the best scenario is chosen, while *KRITsmape* has to be lower than the selected critical value (critical value is 1) for both parts of the time series. In case this condition is satisfied by multiple scenarios, the scenario with lower *KRITsmape_avg_both* (average of *KRITsmape* both parts) is chosen.

After the calculation of points A–D, the final check is done to avoid false detection of changepoint. The following cases are not allowed with respect to the definitions of anomalies under consideration:

- Multiple changepoints in one time series (due to length of time series);
- The changepoint between the first and second point of the time series;
- The so-called "forbidden area" (the last three points of the time series) is imposed, where changepoint detection is not done. If the "forbidden area" contains outliers, it is extended by the number of outliers.

The reason for the last two rules is increased uncertainty at the beginning and end of the time series. The rule for the end of the time series is stricter to avoid situations

when only one or two points are left after the changepoint. This rule was imposed due to subsequent steps of time series analysis.

2.4. Changepoint Detection—Machine Learning for Parameter Setting

A machine learning approach was applied to the detection of changepoints, which will enable the processing of a large number of time series [37]. Based on the approach described above in Section 2.3 are values A–D given for each tested point in the time series calculated. The aim is the automated detection of changepoints to be as close as possible to the visual assessment of the time series by an expert. Simultaneously, the system is able to learn from new assessments.

The initial information is therefore the determination of the presence of a changepoint based on a visual assessment by experts in the field of WM. The information of visual assessment is the type of changepoint (trend changepoint, step changepoint) and the year of changepoint occurrence. The objective is to establish a procedure such that the holds true:

$$\max(\text{TP} + \text{TN}), \tag{3}$$

where TP denotes the number of correctly identified changepoints, and TN is the number of correctly evaluating the absence of a changepoint. When comparing automated calculation (points A–D) and visual assessment. The correct determination of the occurrence of changepoints is thus maximized. This also minimizes erroneous estimates.

Regression analysis is commonly used in machine learning approaches to assess calculated values with visual assessment outputs [38]. From the point of view of regression analysis Equation (4), the dependent variables in the information about the presence of a changepoint (marked by y). The values of steps A–D represent an independent variable (x). The goal is to find the values of the regression parameters (β), which would be used to describe this dependence with sufficient accuracy.

$$y = f(x, \beta) \tag{4}$$

Function $f(x, \beta)$ in Equation (4) can represent a linear regression model, but also more complex structures.

The whole process of the methodology is shown schematically in Figure 4. It works with training data from the current dataset. These are time series that were randomly selected. The training data enters the learning process, where experts comment on the occurrence of changepoints. At the same time, the values for points A–D are calculated. The procedure is shown in Figure 4. A regression model is built from the outputs of the learning process and the calculated values A–D according to Equation (3). The results of the regression model are assessed according to successful evaluation (TP + TN) and unsuccessful evaluation (FP + FN). In the case of high values (FP + FN), it is possible to consider including additional criteria x in the model. Every year, the data on WM for the past year are supplemented, and the time series is extended, which makes it possible to update the calculations. The model can be further trained if a visual assessment of the time series is supplemented.

Based on the current data, the training data includes 300 time series (100 for group 3 and 200 for group 20), see Section 3. The procedure of visual assessment was made for, so at total of 300 time series by five experts independently. This is a high-quality data set for evaluating changepoints, but it is not possible to create statistically significant regression models. For this, it would be necessary to significantly expand the learning process with new expert assessments. The learning process is constantly being supplemented with new assessments and when it is comprehensive, regression models will be built. At present, however, it is necessary to proceed to a simplified step-by-step form of evaluation in the form of an assessment with critical values. Therefore, it is necessary to determine the critical values with which the calculations will be compared and determine whether it is a changepoint. It is, therefore, a gradual fulfillment of the criteria. Critical values are

determined for a total of quantities (steps A–D in Section 2.4): coefficient of determination R² (A), angles (B), slopes of lines (C), SMAPE (D). The procedure of changepoint evaluation is shown in Figure 5; the other parts remain the same as in Figure 4.
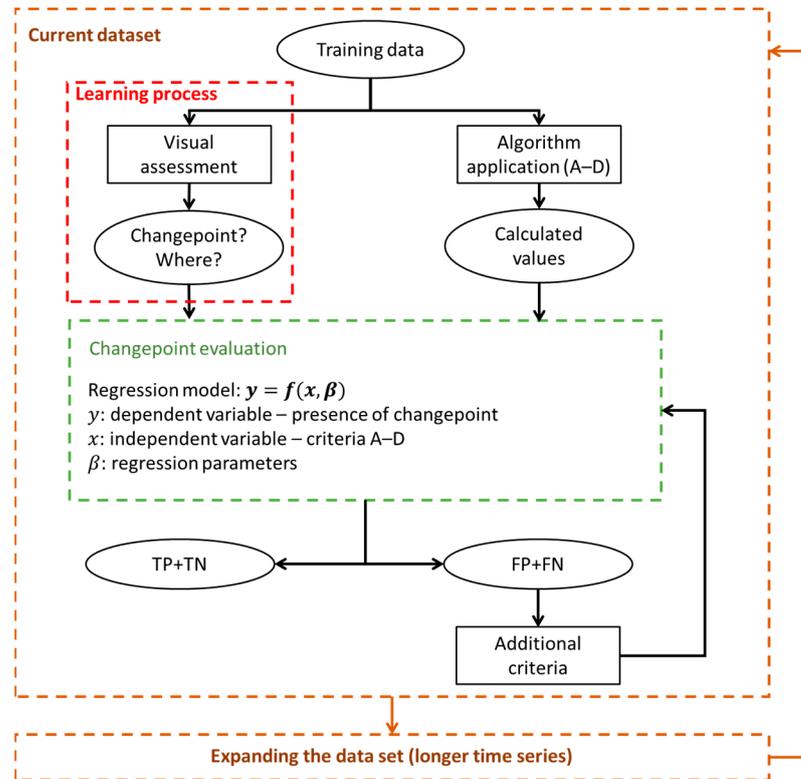

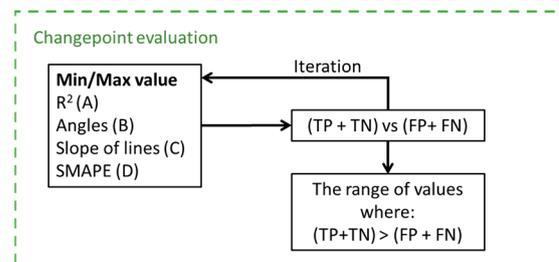**Figure 4.** Schematic representation of the method.

Figure 5. Schematic representation of the critical values setting.

First, a minimum and maximum value was marked for each quantity of steps A–D. In this range, all changepoints would be found, i.e., the maximum value of TP. It means that FN would be zero. However, the consequence is also a large amount of FP. The goal is to shift the minimum and maximum value boundaries for each step A–D so that Equation (3) holds. Thus, an iterative assessment of combinations of values was carried out. All values are already known at this stage—only different combinations of values for A–D are compared. The range of initial minimum and maximum value, the step size for comparing combinations, is shown in Table 2. The critical values are then set such that (TP + TN) > (FP + FN) for the range of values (from minimum to maximum) and their combination with all steps (A–D). The selected critical values based on available data are shown in the right part of Table 2. It should be noted that this is the current setting. With new information, the model will gradually learn and refine.

Table 2. Parameters for setting critical values.

	Initial Values			Selected Critical Values	
	Minimum Value	**Maximum Value**	**Step Size**	**Minimum Value**	**Maximum Value**
R^2	0.05–0.25	0.70–0.95	0.01	0.15	0.85
Angles	60–90°	125–155°	1°	75°	140°
Slopes of lines	35–65°	115–145°	1°	45° * 50° **	135° * 130° **
SMAPE	0	1–2	0.1	0	1.5

Remark: * Value for at least three slopes of lines out of four. ** Value for at least two slopes of lines out of four.

A special case is critical values for slopes of lines. Four values for each tested time series point are compared to the critical value. A suspected changepoint must have all four values with the same sign. When the critical value was required to be met for all four slopes of lines. It was too strict a rule that prevented the changepoint from being found. Therefore, a critical value is set, which must be met for at least three slopes of lines and another value for at least two slopes of lines. In this form, positive results were achieved on the training data.

In transition sections near critical values, it may be useful to use a combination of step-by-step approaches with critical values and a regression model. However, a large number of expert assessments are required to build the regression model.

Comments on the selected critical values based on current data and a visual assessment are provided below (points A–D). It is important to note that the current critical values will gradually adjust with new information, and the model will learn.

A. Critical values for the coefficient of determination (R^2)

R^2 is bounded for the reason that a very low value of R^2 indicates a noisy time series. Such a time series will not be analyzed in terms of changepoints. In Figure 6a,b, two time series from the expert assessment are selected. The coefficients of determination of these time series are close to the current critical value. The coefficient of determination of the time series in Figure 6a ($R^2 = 0.123$) is smaller than the specified critical value and it is clear that this time series is considerably noisy. On the other hand, the time series in Figure 6b have a clearly visible trend, and a potential changepoint can be seen. The coefficient of determination of this time series ($R^2 = 0.158$) is larger than the specified critical value. It is possible that for some time series, especially with the coefficient of determination around the critical value, the opposite situation will occur. Thus, time series with a potential changepoint may also fall below the critical value and, therefore, they will be excluded from the detection.

At the other end of the coefficient of determination scale, the upper critical value was set to 0.85 based on current data. The assumption for this critical value was that a time series with an almost perfect fit does not require changepoint detection. In Figure 6c,d, there are two time series with the R^2 close to the determined critical value. In Figure 6c, the coefficient of determination ($R^2 = 0.845$) is smaller than the critical value, and the point from the year 2014 may be considered suspicious. In Figure 6d, the coefficient of determination ($R^2 = 0.878$) is larger than the critical value, and no changepoint is expected. It is a time series with an almost perfect fit. This critical value was set in the same way as the lower one and thus is a compromise based on the visual expert assessment. If the condition on R^2 is met, the procedure continues to step B. Otherwise, the changepoint is not tested in the selected time series (the time series is either too noisy or the trend is too clear). As the data set expands, the critical values will be updated.
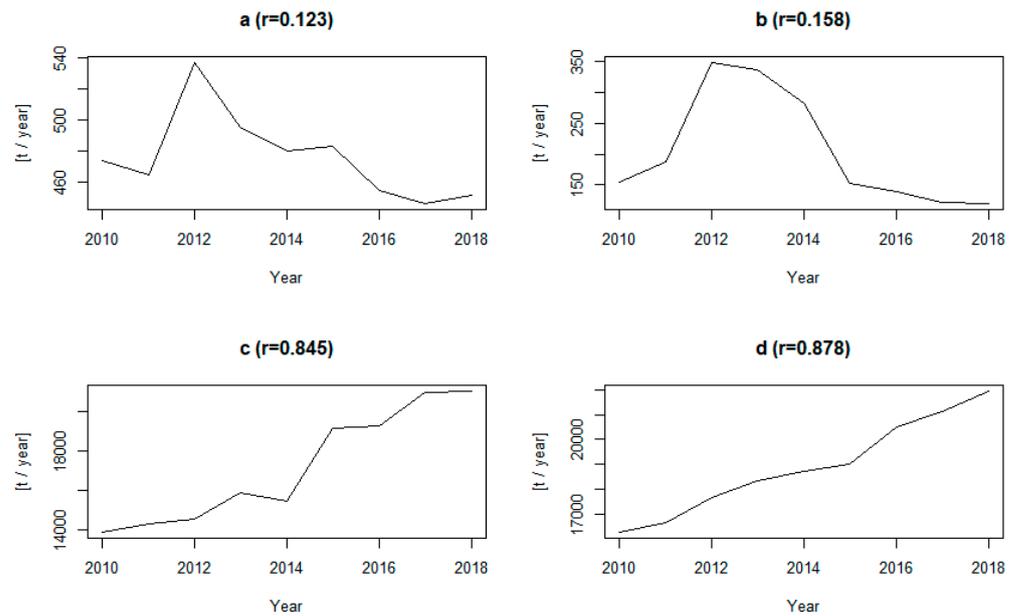
Figure 6. (**a**) Wood processing waste; (**b**) paper and cardboard processing waste; (**c**) hazardous waste from glass manufacture; and (**d**) discarded equipment containing chlorofluorocarbons waste.

B. Critical values for angles

There was a computed angle at each point of the examined time series. This means that each pair of consecutive points was fitted by a straight line and the angle between each pair of consecutive lines was computed. The limitation of the size of these angles was based on the assumption of the shape of the changepoints. The changepoints should be L-shaped or Z-shaped (with right or obtuse angles) and not V-shaped or A-shaped (meaning acute angles). This was taken into account by the experts during the visual assessment. Critical values for angles were set between $75°$ and $140°$. The point in the time series with an angle in this range is flagged as suspicious in terms of changepoint occurrence.

C. Critical values for slopes of lines

For each fitted line from the procedure from paragraph B, the slope of the line was stored, and the rules described in the Materials and Methods section were outlined. The setting of the limitation for the slopes is demonstrated in Figure 7. When this time series was analyzed, it went through the requirement for the coefficient of determination. Simultaneously, four points (years 2012–2016) went through the requirement for angle size. The points in the time series from the years 2013 and 2014 were eliminated based on the requirement for the same signs of slopes of all four lines. At these points, the signs of the slopes were not the same—two were positive, and two were negative. However, a point from 2016 passed the requirement for the same signs of slopes. The requirement for at least three angles between the lines and the *x*-axis larger than $45°$ was not met. As can be seen from Figure 7, the angles between the lines and the *x*-axis at this point are very low. The only point that passed both requirements is the year 2012. This point looks suspicious in terms of changepoint occurrence even after a visual review.
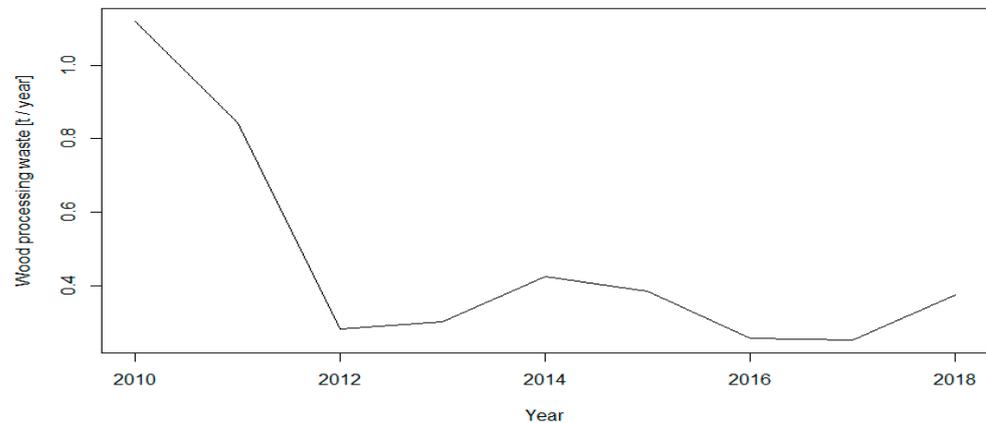
Figure 7. Demonstration of slope limitation.

D.   Critical values for SMAPE

After the detection of suspicious points, it was necessary to find a suitable accuracy measurement for the final validation of the detected changepoint. The assumption of the approach is that the changepoint should split the time series into two parts, which separately have an almost perfect fit. For this purpose, the metric SMAPE proved to be appropriate. The formula of the SMAPE metric is the numerator of the complex fraction from Equation (2). The formula provides a result between 0% and 200%. The denominator in Equation (2) was added to include information on both parts in the calculation and specifically to include the amount of change between these parts (ratio of means). According to the testing data, the critical value for KRITsmape was set at 1.5. Scenarios with a KRITsmape value below this critical value are considered admissible, and the scenario with the lowest average of KRITsmape in both parts is selected.

The model further learns by adding additional expert opinions on the occurrence of changepoints. Each expert estimate adds information, and then it is possible to adjust the critical values. With the annual expansion of the data set, there should be a re-evaluation by experts. Once the data set is large enough, it will be possible to train a regression model with data divided into training, testing and validation. At the moment, there is not enough data available, so the step-by-step procedure has been designed (see Figure 4). A change in the trend is identified in such a time series that passes all criteria. The main goals of the approach are automated calculation updates and model learning using new expert evaluations.

2.5. Changepoint Detection—Performance Evaluation

The presented method for changepoint detection was tested and compared to existing methods (see Section 1.3). The procedure for testing was analogous to the outlier's assessment. So, a total of 300 time series were evaluated (100 for group 3 and 200 for group 20) and this is the test data. In the case of changepoint detection, only 6 points were assessed for each time series because the changepoint is not assumed in 2 first points and 3 last points of the time series. The occurrence of changepoints was first stated by visual assessments and then compared to computed results. As for outliers, the assessed time series were divided into 4 groups (TP, FP, FN, TN); the results are summarized in Table 3.

Table 3. Accuracy of methods for changepoint detection.

		TP	FP	FN	TN	Precision	Recall	F1	GM	Jaccard	Suspicious Time Series
Group 3	**New method**	**8**	**6**	**12**	**574**	**0.57**	**0.40**	**0.47**	**0.75**	**0.31**	**14%**
	bps	1	75	19	505	0.01	0.05	0.02	0.11	0.01	76%
	bcp	1	36	19	544	0.03	0.05	0.04	0.16	0.02	37%
Group 20	**New method**	**3**	**8**	**14**	**1175**	**0.27**	**0.18**	**0.21**	**0.52**	**0.12**	**6%**
	bps	1	156	16	1027	0.01	0.06	0.01	0.08	0.01	79%
	bcp	3	75	14	1108	0.04	0.18	0.06	0.20	0.03	39%

Remark, source [36]: Precision (PPV)—proportion of positive samples that were correctly classified to the total number of positive predicted samples: $PPV = \frac{TP}{FP + TP}$. Recall (TPR)—the ratio of the correctly classified negative samples to the total number of negative samples: $TPR = \frac{TN}{FP + TN}$. F1—harmonic mean of precision and recall: $F1 = \frac{2PPV \times TPR}{PPV + TPR}$. GM—measure for balanced and imbalanced data: $GM = \sqrt{TPR \times TNR}$, where $TNR = \frac{TN}{FP + TN}$. Jaccard—measure ignores the correct classification of negative samples: $Jaccard = \frac{TP}{TP + FP + Fn}$.

From the testing of the four investigated groups (TP, FP, FN, TN), the new method can be considered quite successful. As can be seen from the results, the new method has a higher success in terms of TP. The main advantage of the new method is a significant reduction in FP compared to existing methods. This effect arises because the new method avoids identifying time series with "A" or "V" shaped development as changepoints. In most cases, this is caused by oscillation around the trend, not a real changepoint. In any case, it is recommended to approach the results of this analysis carefully. Ideally, visually verify the true presence of the changepoint. This is possible if the number of time series suspected of the presence of changepoint (TP and FP) is significantly reduced by performing the detection. It should be noted that the stated values assess the individual points of the time series. In fact, there are 100 time series for group 3 and 200 time series for group 20. The last column of Table 3, "Suspicious time series", summarizes the percentage of time series that are suspicious. In this respect, there is a significant benefit in using the new method, which can serve as an indicator for the detection of time series recommended for visual assessment and a final decision on the presence of changepoint. Based on the final control of the results, the time series classified as FN by a new method are, in most cases, ambiguous, with the assessed values close to the critical values (see Section 2.4). Therefore, these are not significant changepoints, and their neglect does not have a negative impact on further work with data.

Higher accuracy of the new method can be achieved by setting parameters individually for a specific dataset. The different waste fractions can vary significantly in their character, and it can be beneficial to adapt the method. The new method allows more degrees of freedom for individual settings compared to existing methods. For this reason, even more successful results are expected for the variable data. This is evident by the TP value compared to the other methods in group 3, which represents the variable data in this testing.

3. Results and Discussion

The case study is realized for the Czech Republic. The dataset for the case study consists of annual data from the period 2010–2018. The dataset contains a broader range (more than 750) of waste types, which can be grouped into subgroups and groups (20; see Appendix A for details). The micro-regional level of data has been used for analysis. The Czech Republic consists of 206 micro-regions. The total has been processed about 9600 time series because some types of waste are produced only in some micro-regions. The following results are presented for aggregated waste types (into 20 waste groups). Furthermore, note that these results are displayed only for the regional level (territories 'CZ0XY') and national level (territory 'cr') data for better clarity (see Appendix B for details).
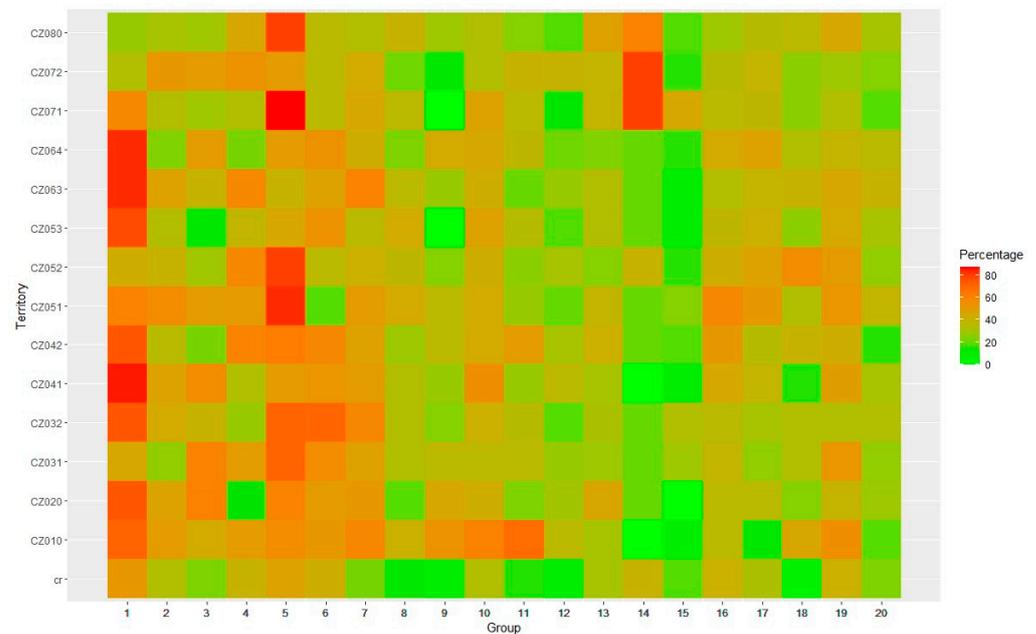
3.1. Outlier Detection

Figures 8–10 provide a graphical representation of the results using the proposed solution (Holt + Grubbs test).
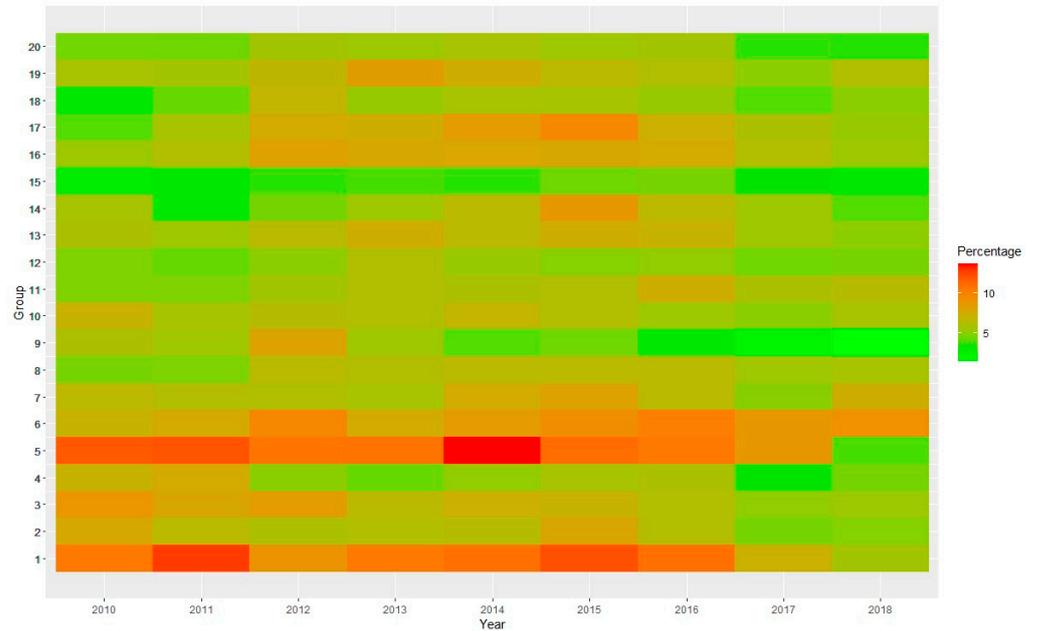
Figure 8. Percentage of time series containing outlier by waste group (see Appendix A and B for details) and year.
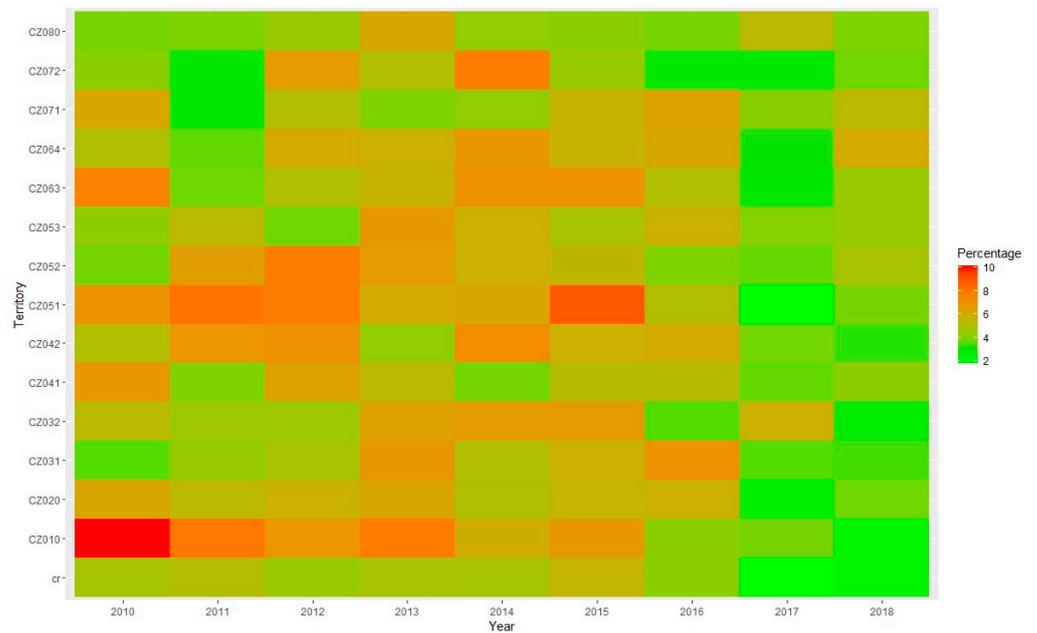
Figure 9. Percentage of time series containing outliers by region and year.

Figure 10. Percentage of time series containing outlier by region and waste group (see Appendix A and B for details).

Figure 8 shows the percentage of time series in a given waste group (group description is included in Appendix A) containing outliers in a given year. This kind of graph allows us to quickly identify problematic waste (group) both overall (high percentage of outliers across the years) and individual years. Individual groups of waste are listed vertically (Group 1–20); the years are shown horizontally (2010–2018). The red color in the graph indicates that a large part of the time series of the given group and year shows an outlier. Groups 1 and 5 generally contain a higher percentage of outliers, and also, the single worst case is for group 5 in 2014. It should be mentioned that the data in groups 1 and 5 are significantly variable, and it is appropriate to deal with the reason for the outlier in these groups. On the other hand, the production of packaging waste (group 15) and municipal solid waste (group 20) are stable. It can therefore be observed that some groups of waste have a higher occurrence of outliers in the long term. In the last year (2018), the outlier was identified with less frequency. The reason is the less reliable identification of outliers at the last point of the time series. The results show that 83% of the time series from group 15 and 76% time series from group 20 are without outliers.

Figure 9 demonstrates the impact of time and territory on the presence of outliers. The chart system is the same as Figure 8, with the difference that micro-regions are aggregated into regions on the vertical axis. The designation of the regions is given in Appendix B. From this kind of graph, a territory containing an unusual number of outliers (also with respect to year) can be identified. This allows for quick identification of potentially problematic combinations of territory and year. A non-systematic distribution of colors in the chart is fine, e.g., for the territory of CZ051 in 2015, more outliers than usual were identified. This may have been caused by the current conditions in the given locality, a registration error, etc., which did not affect other regions. In the case of such a problematic pair (territory and year), it should be questioned whether there is a systematic error in the records or this year was actually specific in a given territory (in this case, 'outliers' could be only 'extremes' and should not be removed or corrected). In this case, CZ010 (see Appendix B) shows a high number of time series containing outliers in 2010. Due to the high incidence of outliers in 2010, it may be useful to omit this year's data for further analysis if there is no problem with the amount of data for a particular analysis. However, there seems to be no big difference between regions, with regional averages (over the whole period) ranging from 4.5% to 6.5%.

The last of the presented graphs (Figure 10) shows the impact of waste group and territory on outliers regardless of the year. It can be seen that, e.g., a high number of time series containing outliers is present in groups 1 and 5 across the territories. Group 14 (waste organic solvents, refrigerants and propulsion media (excluding wastes listed in groups 7 and 8) seem to be problematic for territories CZ071 and CZ072. This is waste that can be closely linked to industry in a given area. On the other hand, the time series from group 15 does not contain a high number of 'problematic' time series. It should also be noted that pairs presented in Figure 10 contain up to 80% of time series with outliers. Such a high number is suspicious and should be examined closely.

In summary, the presence of an outlier was identified in approximately 16% of time series. In the waste production of municipalities, considerable stability can be observed in the percentage of identified series for individual waste fractions. For fractions where at least 100 series were available, all ranged between 10% and 20%. The production of companies generally contained a smaller number of series and thus also a greater variability of results. Of the main, and therefore more numerous fractions, paper, plastic (both 28%), and metals (36%, but only 55 series are available) have the most outliers.

3.2. Changepoint Detection

Similar plots, as in the case of outlier detection, can be created. Years 2010 and 2016–2018 are excluded due to problematic identification at the beginning and end of the time series. Figure 11 shows the occurrence of changepoints for waste groups (vertical axis) and year (horizontal axis). Each waste group includes multiple time series; therefore, the occurrence of changepoints is expressed as a percentage of all time series. Figure 11 shows a high number of changepoints present in 2012 for multiple waste groups. The reason, among other things, may be changes in the method of recording data, as the data used are recorded annually by all waste producers. Figure 12 (for territories) agrees with the previous result and shows that 2012 is problematic for most of the regions; however, CZ010 contains a particularly high number of changepoints. However, it is not clear why 2012 contains such a high number of time series. Such cases need to be investigated closely.
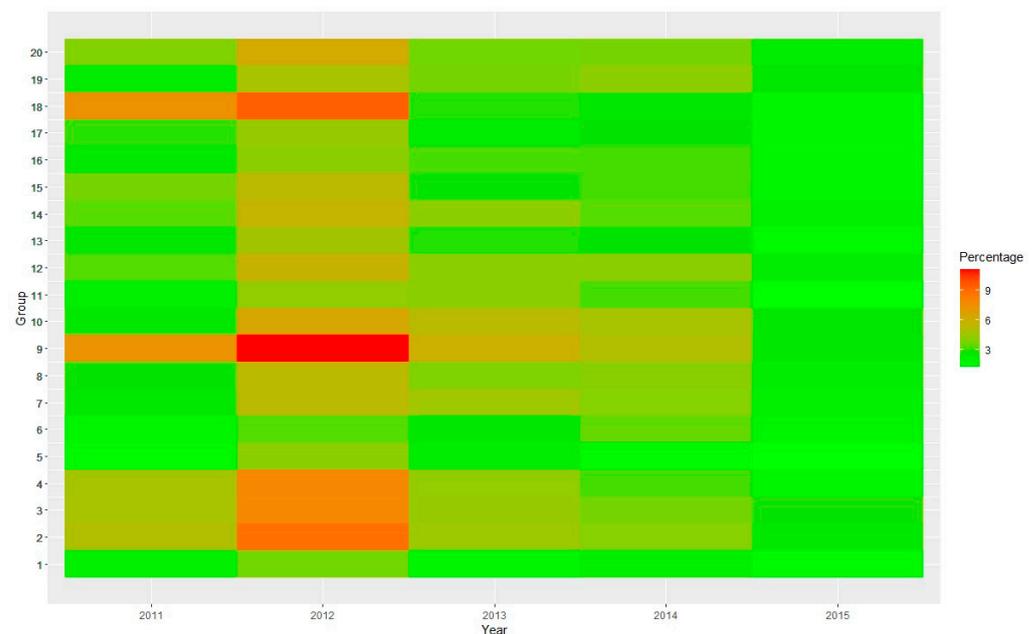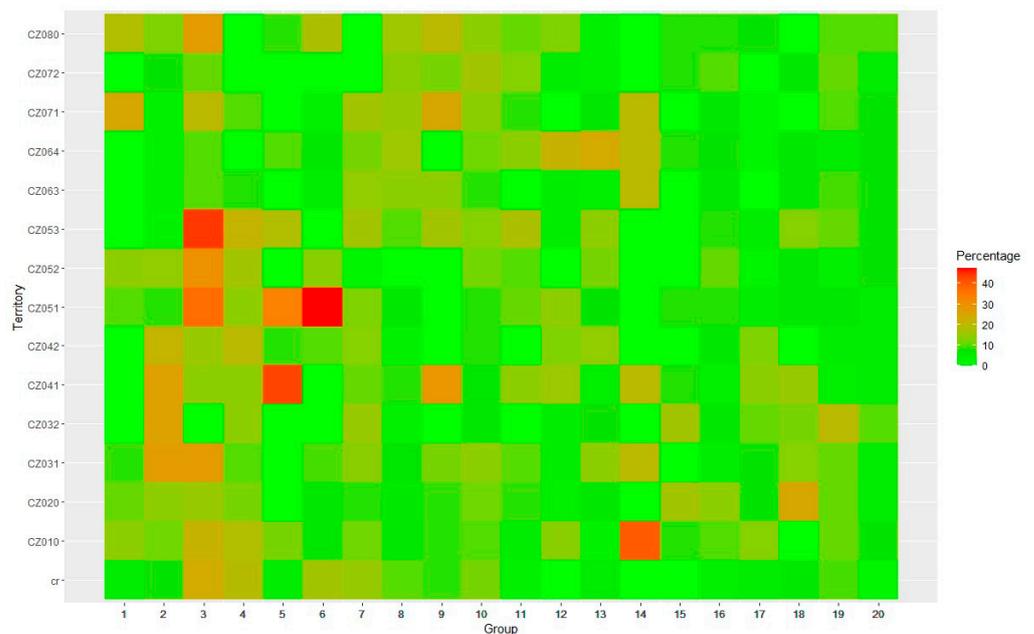

Figure 11. Percentage of time series containing changepoint by waste group (see Appendix A and B for details) and year.
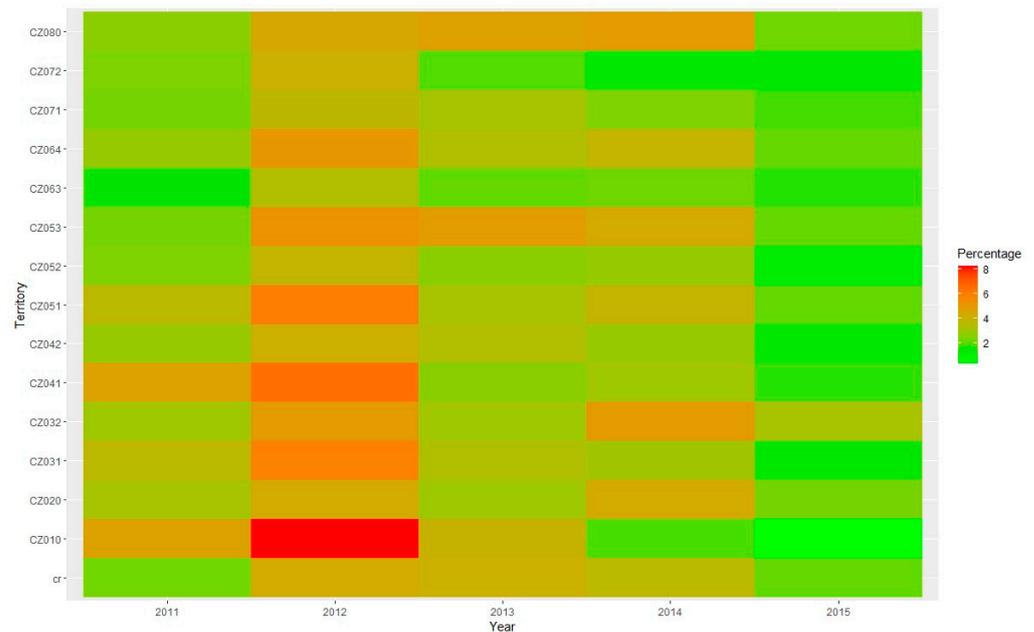
Figure 12. Percentage of time series containing changepoint by region and year.

Finally, Figure 13 demonstrates that in most cases, only one to three waste groups per territory contain most of the changepoints. Once again, the number of changepoints present in this graph is much higher when compared to Figures 11 and 12 (same as in the case of outliers). As Figure 13 shows, up to 40% time series of some waste groups has identified changepoints. These results show that changepoint detection is essential for further work with data, and pre-processing is crucial in time series. The quality studies should include this part of data analysis; however, usually the data pre-processing is not performed at all or at most outlier detection [9]. Nevertheless, the changepoint detection can have a fundamental impact on the quality of the output from the analyses, as shown by the results of the case study on WM in the Czech Republic.

Figure 13. Percentage of time series containing changepoint by region and waste group (see Appendix A and B for details).

Summary

The presented method for changepoint detection achieved the best results compared to the previous approaches (bsp and bcp) (see Table 3 in Section 2.5). Different indicators were evaluated based on the TP, FP, FN and TN values for two different waste groups: Group 3 (represents highly variable data) and Group 20 (represents relatively stable data). It is particularly important to reduce the number of FP using the new method. This is a case where a changepoint is detected but does not appear in the data. This problem was significantly eliminated with the new method. This is a major benefit because, in the case of high values for FP, some information is neglected for further work with the data. By using the presented method, better results can be achieved during data pre-processing and thereby improve the quality of planning in the area of WM.

The first step for detecting changepoints is to detect outliers. From an overall perspective, it seems that the problem with outliers is rather diminishing in time for most groups (especially in the last 2 years). The percentage of outliers in the first seven years (2010–2016) is between 5 and 6%, while in the last 2 years, this number drops under 4%. This behavior is expected since data collection quality and control mechanisms are improving. Waste time series from mining (group 1) and petroleum industry (group 5) contain the biggest percentage of outliers (about 10%). On the other hand, 4 or 5% of MSW time series (group 15 or 20 respectively) contain outliers. As expected, these numbers are lower for changepoints, e.g., the average percentages of time series containing changepoint range from 2 to 6.5% per waste group (group 9 being the worst).

In total, 31.1% of time series contained (at least one) outlier and 16.4% changepoint. Anomalies were detected, especially for waste fractions with an indistinct trend (e.g., group 3). Group 3 includes waste from wood processing. In this time series, a changepoint can be often seen in the period of the bark beetle calamity in the Czech Republic. This change can be responded to thanks to anomaly detection. Otherwise, the forecast would be formed without reacting to this period. On the contrary, group 20 (waste from citizens and similar) has a relatively constant trend. However, changepoints can often appear as a result of legislative changes, technological progress, infrastructure changes, etc. Although anomalies are more common at lower levels of territory, they should be detected for all data. Thus, more accurate forecasts can be achieved.

The limitation of the method is primarily that it is necessary to have a sufficient number of expert evaluations. Thanks to this, the system has the opportunity to learn and achieve quality results. Furthermore, the evaluation must be repeated with each extension of the time series. That is, in the case of annual detail, the update should take place once a year. Considering the need to repeat calculations, the process is completely automated. The system is designed to place minimum requirements on the available data. It is thus adapted to a very short time series that may show an unstable trend.

4. Conclusions and Future Work

The presented paper dealt with the issue of anomaly detection in a short time series focused on the WM field. Although it is a well-known fact that WM data contain anomalies, little to no attention was given to their detection previously. Only naïve methods for outlier detection in WM were found in the literature, while no paper dealing with changepoint detection was found. The Holt method combined with the Grubbs test is recommended for outlier detection based on the results provided in Section 2.2.

The changepoint detection methods used in the literature are often not applicable for such a short time series. Therefore, the possibility of creating its own method based on basic statistical procedures such as linear regression, coefficient of determination and SMAPE, and setting simple rules was explored. The crucial aspect of the method is to define appropriate rules and set critical values correctly to make detection as accurate as possible and convenient for the length of time series in the field of WM. The critical values are set using principles of supervised machine learning. As the time series extends, the critical values will be updated. The learning of the model is possible by additional visual

assessment. According to the current data, the method proved to be successful and reliable. The probability of type I error is a maximum of 2% in all tested waste groups, and the probability of type II error is less than 8%. In summary, for about 90% of the time series, the changepoints were correctly identified.

However, it should be noted that no objective information exists about the presence of anomalies in the dataset under examination. There is no absolutely correct method with which to compare results. Thus, expert judgment is the only way to assess the results, and it needs to incorporate knowledge of legislative changes and other external factors influencing waste generation. In every case, anomalies identified by proposed algorithms should be judged by an expert. The new changepoint detection approach was compared with existing approaches that are usable for short time series (Section 2.5). The new approach is considered to be the most successful for this type of data from the available methods, as it comes closest to expert evaluation. The main benefit of the proposed approach is the possibility of automating the process of deciding on the presence of an anomaly and the possibility of learning. Thanks to machine learning, the system will learn based on the opinions of several experts. When a comprehensive data set is achieved, the approach will be less dependent on the subjective opinion of the individual.

In the field of waste management, this contribution represents a fundamental contribution leading to the improvement of production forecasts and/or waste management. Without high-quality forecasts, it is impossible to create adequate plans for the waste economy and to fulfill legislative goals. The benefit of the contribution can, therefore also be perceived from the point of view of supporting the sustainability of the use of natural resources. From the perspective of data processing theory, the proposed method represents a unique approach to time series processing. The method, therefore, has the potential to be used in other areas of data processing that show a similar character, especially short time series.

The need for a sufficient amount of expert evaluation is one of the main limitations of the approach. With a higher number of ratings, a higher quality of results can be achieved. This is the main point of further research. The method should move from step-by-step criteria for identifying change to a regression model. However, a large number of expert evaluations are needed. In transition sections, it may be useful to use a combination of these approaches. Another possible direction for future research is the usage of some transformation of the original data. E.g., logarithmic transformation reduces the number of outliers as well as improves other properties of data. Logarithmic transformation is widely used for processing data from various fields [39]. However, in such a case, a zero-value replacement strategy needs to be developed and examined thoroughly. A larger number of suspicious cases investigated will be beneficial, as well as a wider range of experts. Another way to improve the whole framework is by an enhancement of the algorithmic detection by expert judgment to adjust the parameters of the proposed method.

Author Contributions: Conceptualization, R.Š.; methodology, M.R. and K.Š.; investigation, M.R., and V.S.; data curation, K.Š.; writing—original draft preparation, M.R. and V.S.; writing—review and editing, V.S. and R.Š.; visualization M.R. and K.Š.; supervision, R.Š.; project administration, R.Š.; funding acquisition, R.Š. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The output of the project is a methodology. The data on which the development was carried out is not public.

Appendix A

Table A1. Waste group numbers.

Waste Group Number	Description
1	Wastes from geological exploration, mining, treatment and further processing of minerals and stone
2	Wastes from agriculture, horticulture, fisheries, forestry, hunting and food production and processing
3	Wastes from wood processing and manufacture of boards, furniture, cellulose, paper and paperboard
4	Wastes from the leather, fur and textile industries
5	Wastes from oil refining, natural gas refining and pyrolytic coal processing
6	Wastes from inorganic chemical processes
7	Wastes from organic chemical processes
8	Wastes from the manufacture, processing, distribution and use of paints (paints, varnishes and enamels), adhesives, sealants and printing inks
9	Wastes from the photographic industry
10	Wastes from thermal processes
11	Wastes from chemical surfaces, surfaces of metal and other materials and hydrometallurgy of non-ferrous metals
12	Wastes from shaping and physical and mechanical surface treatment of metals and plastics
13	Oil wastes and wastes of liquid fuels (excluding edible oils and wastes of groups 5, 12 and 19)
14	Waste organic solvents, refrigerants and propulsion media (excluding wastes listed in groups 7 and 8)
15	Waste packaging; absorbents, cleaning cloths, filter materials and protective clothing, not elsewhere specified or included
16	Wastes not otherwise specified in the catalogue
17	Construction and demolition wastes (including excavated soil from contaminated sites)
18	Wastes from health care and veterinary care and/or research related to them (excluding kitchen waste and waste from catering facilities, which are non-central health)
19	Wastes from waste treatment plants (use and disposal), wastewater treatment plants for the treatment of these waters and the production of water for human consumption and industrial water
20	Municipal wastes (household wastes and similar trade, industrial and official wastes), including separately collected components

Appendix B

Territory Code	Territory Level	Territory Name
CZ	Country	Czech Republic
CZ010	Region	Prague the Capital city
CZ020	Region	Central Bohemian Region
CZ031	Region	South Bohemian Region
CZ032	Region	Pilsen Region
CZ041	Region	Karlovy Vary Region
CZ042	Region	Ústí nad Labem Region
CZ051	Region	Liberec Region
CZ052	Region	Hradec Králové Region
CZ053	Region	Pardubice Region
CZ063	Region	Vysočina Region
CZ064	Region	South Moravian Region
CZ071	Region	Olomouc Region
CZ072	Region	Zlín Region
CZ080	Region	Moravian-Silesian Region

References

1. Petropoulos, F.; Apiletti, D.; Assimakopoulos, V.; Babai, M.Z.; Barrow, D.K.; Taieb, S.B.; Bergmeir, C.; Bessa, R.J.; Bijak, J.; Boylan, J.E.; et al. Forecasting: Theory and practice. *Int. J. Forecast.* **2022**, *38*, 705–871. [CrossRef]
2. Zgurovsky, M.; Sineglazov, V.; Chumachenko, E. Intelligence Methods of Forecasting. *Stud. Comput. Intell.* **2021**, *904*, 313–361. [CrossRef]
3. Atkinson, A.C.; Riani, M.; Corbellini, A. The Box–Cox Transformation: Review and Extensions. *Stat. Sci.* **2021**, *36*, 239–255. [CrossRef]
4. Šomplák, R.; Smejkalová, V.; Rosecký, M.; Szásziová, L.; Nevrlý, V.; Hrabec, D.; Pavlas, M. Comprehensive Review on Waste Generation Modeling. *Sustainability* **2023**, *15*, 3278. [CrossRef]
5. Kuznetsova, E.; Cardin, M.-A.; Diao, M.; Zhang, S. Integrated decision-support methodology for combined centralized-decentralized waste-to-energy management systems design. *Renew. Sustain. Energy Rev.* **2019**, *103*, 477–500. [CrossRef]
6. Ribic, B.; Pezo, L.; Sincic, D.; Loncar, B.; Voca, N. Predictive model for municipal waste generation using artificial neural networks—Case study City of Zagreb, Croatia. *Int. J. Energy Res.* **2019**, *43*, 5701–5713. [CrossRef]
7. Niska, H.; Serkkola, A. Data analytics approach to create waste generation profiles for waste management and collection. *Waste Manag.* **2018**, *77*, 477–485. [CrossRef]
8. Cubillos, M.; Wulff, J.N.; Wøhlk, S. A multilevel Bayesian framework for predicting municipal waste generation rates. *Waste Manag.* **2021**, *127*, 90–100. [CrossRef]
9. Alcay, A.; Montañés, A.; Simón-Fernández, M.-B. Waste generation and the economic cycle in European countries. Has the Great Recession decoupled waste and economic development? *Sci. Total Environ.* **2021**, *793*, 148585. [CrossRef]
10. Kannangara, M.; Dua, R.; Ahmadi, L.; Bensebaa, F. Modeling and prediction of regional municipal solid waste generation and diversion in Canada using machine learning approaches. *Waste Manag.* **2018**, *74*, 3–15. [CrossRef]
11. Tozlu, A.; Abusoglu, A.; Ozahi, E.; Anvari-Moghaddam, A. Municipal solid waste-based district heating and electricity production: A case study. *J. Clean. Prod.* **2021**, *297*, 126495. [CrossRef]
12. Rashid, M.I.; Shahzad, K. Food waste recycling for compost production and its economic and environmental assessment as circular economy indicators of solid waste management. *J. Clean. Prod.* **2021**, *317*, 128467. [CrossRef]
13. Mohammadi, E.; Singh, S.J.; Habib, K. How big is circular economy potential on Caribbean islands considering e-waste? *J. Clean. Prod.* **2021**, *317*, 128457. [CrossRef]
14. Singh, S.P.; Jawaid, M.; Chandrasekar, M.; Senthilkumar, K.; Yadav, B.; Saba, N.; Siengchin, S. Sugarcane wastes into commercial products: Processing methods, production optimization and challenges. *J. Clean. Prod.* **2021**, *328*, 129453. [CrossRef]
15. Capasso, I.; Liguori, B.; Ferone, C.; Caputo, D.; Cioffi, R. Strategies for the valorization of soil waste by geopolymer production: An overview. *J. Clean. Prod.* **2021**, *288*, 125646. [CrossRef]
16. Smejkalová, V.; Šomplák, R.; Nevrlý, V.; Burcin, B.; Kučera, T. Trend forecasting for waste generation with structural break. *J. Clean. Prod.* **2020**, *266*, 121814. [CrossRef]
17. Aminikhanghahi, S.; Cook, D.J. A survey of methods for time series change point detection. *Knowl. Inf. Syst.* **2017**, *51*, 339–367. [CrossRef]
18. Aggarwal, C.C. *Outlier Analysis*; Springer: New York, NY, USA, 2013; ISBN 978-1461463955.
19. Braei, M.; Wagner, S. Anomaly Detection in Univariate Time-series: A Survey on the State-of-the-Art. *arXiv* **2020**, arXiv:2004.00433.
20. Choi, K.; Yi, J.; Park, C.; Yoon, S. Deep Learning for Anomaly Detection in Time-Series Data: Review, Analysis, and Guidelines. *IEEE Access* **2021**, *9*, 120043–120065. [CrossRef]
21. Blázquez-García, A.; Conde, A.; Mori, U.; Lozano, J.A. A review on outlier/anomaly detection in time series data. *arXiv* **2020**, arXiv:2002.04236. [CrossRef]
22. Chalapathy, R.; Chawla, S. Deep Learning for Anomaly Detection: A Survey. *arXiv* **2019**, arXiv:1901.03407.
23. Chandola, V.; Banerjee, A.; Kumar, V. Anomaly detection: A survey. *ACM Comput. Surv.* **2009**, *41*, 1–58. [CrossRef]
24. Rybová, K.; Burcin, B.; Slavík, J. Spatial and non-spatial analysis of socio-demographic aspects influencing municipal solid waste generation in the Czech Republic. *Detritus* **2018**, *1*, 3–7. [CrossRef]
25. Truong, C.; Oudre, L.; Vayatis, N. Selective review of offline change point detection methods. *Signal Process.* **2020**, *167*, 107299. [CrossRef]
26. Li, Y.; Lin, G.; Lau, T.; Zeng, R. A Review of Changepoint Detection Models. *arXiv* **2019**, arXiv:1908.07136.
27. Kawahara, Y.; Sugiyama, M. Sequential Change-Point Detection Based on Direct Density-Ratio Estimation. *Stat. Anal. Data Min.* **2012**, *5*, 114–127. [CrossRef]
28. Kawahara, Y.; Yairi, T.; Machida, K. Change-Point Detection in Time-Series Data Based on Subspace Identification. In Proceedings of the Seventh IEEE International Conference on Data Mining, Omaha, NE, USA, 28–31 October 2007; pp. 559–564. [CrossRef]
29. Adams, R.P.; Mackay, D. Bayesian Online Changepoint Detection. *arXiv* **2007**, arXiv:0710.3742.
30. Chandola, V.; Vatsavai, R.R. Scalable Time Series Change Detection for Biomass Monitoring Using Gaussian Process. In Proceedings of the 2010 Conference on Intelligent Data Undestanding, Mountain View, CA, USA, 5–6 October 2010.
31. R Core Team. *R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria*; European Environment Agency: Copenhagen, Denmark, 2021. Available online: https://www.R-project.org/ (accessed on 4 October 2023).

32. Gamallo, P. Using the Outlier Detection Task to Evaluate Distributional Semantic Models. *Mach. Learn. Knowl. Extr.* **2019**, *1*, 211–223. [CrossRef]

33. Dean, R.B.; Dixon, W.J. Simplified statistics for small numbers of observations. *Anal. Chem.* **1951**, *23*, 636–638. [CrossRef]

34. Thomson, M.; Jowthian, P. *Notes on Statistics and Data Quality for Analytical Chemists*; Birkbeck University of London: London, UK, 2011. [CrossRef]

35. Breunig, M.M.; Kriegel, H.-P.; Ng, R.T.; Sander, J. LOF: Identifying density-based local outliers. *ACM SIGMOD Rec.* **2000**, *29*, 93–104. [CrossRef]

36. Tharwat, A. Classification assessment methods. *Appl. Comput. Inform.* **2021**, *17*, 168–192. [CrossRef]

37. Nakano, K.; Chakraborty, B. Effect of Data Representation for Time Series Classification—A Comparative Study and a New Proposal. *Mach. Learn. Knowl. Extr.* **2019**, *1*, 1100–1120. [CrossRef]

38. Gupta, V.; Mishra, V.K.; Singhal, P.; Kumar, A. An Overview of Supervised Machine Learning Algorithm. In Proceedings of the 2022 11th International Conference on System Modeling and Advancement in Research Trends, (SMART), Moradabad, India, 16–17 December 2022; pp. 87–92. [CrossRef]

39. Verma, M.; Gharpure, D.C.; Wagh, V.G. Pre-processing of data using logarithmic transformation to improve the spatial resolution of an EIT system for biomedical applications. *J. Phys. Conf. Ser.* **2019**, *1272*, 012021. [CrossRef]