Application of change-point analysis to the selection of representative data in creep experiments

Setareh Zomorodpoosh¹, Nicklas Volz¹, Steffen Neumeier² and Irina Roslyakova¹
¹ Interdisciplinary Center for Advanced Materials Simulations, Ruhr-Universität Bochum, Universitätstr. 150, 44801 Bochum, Germany
² Department of Material Science and Engineering, Institute I, Friedrich-Alexander-Universität Erlangen-Nürnberg, Martenstr. 5, 91058 Erlangen, Germany
E-mail: setareh.zomorodpoosh@rub.de
Keywords: automated data selection, change-point analysis, non-stationary time series, creep

Abstract
The high volume of data resulting from a rapidly increasing number of experiments in materials science necessitates an efficient preparing of the data before any analysis. In addition, due to the large datasets in some experiments, it is essential to reduce the data sample to a small number of representative data points. In this study, three statistical methods for the change-point analysis are tested for the automated selection of representative creep data which provides large possibilities to speed up the data preparation for their further analysis. Moreover, this approach aids the practitioner to produce consistent and unique representative data for each experiment more efficiently.

1. Introduction

Due to the spreading of information technologies and the increase in data volumes and complexity, automated methods of data analysis are becoming a topic of much interest [1]. One of the essential steps in data analysis is data cleaning that can strongly improve the quality of the final result [2]. Thus, before any statistical analysis can be performed the so-called 'raw-data' has to be cleaned. Basic operations of data cleaning include removing corrupt and inaccurate data and imputing missing data [3]. Another important step is data reduction in which the number of data points is reduced while the integrity of the whole dataset is preserved. The size of the representative data should be sufficient enough to capture complete behavior during the experiment while excluding the possible data dependencies. This makes the process of data modeling more efficient as it produces a more robust and better result [4]. Therefore, this step is an important part of the scientific process and there is a wide variety of established approaches to perform it. One of the possible solutions for data reduction is based on change-point analysis which is capable of detecting changes on time ordered data like creep experimental data [5, 6]. Creep is defined as time-dependent plastic deformation and failure of material over long times under the influence of mechanical stress below the yield stress and high temperature. Studying the creep behavior at high temperature is an important aspect of many technologies such as nuclear power plants, turbine blades, jet engines [7]. The rate and the value of deformation are mainly dependent on temperature, applied load, exposure duration, and some other factors like microstructure, heat treatment, and orientation [8–11]. The creep experiment is performed in order to verify the mechanical properties of novel alloys and determine the possible ranges of application for the tested materials. The results can guide further experimental or data-driven alloy design [12, 13]. During the experiment, a txt-file with saved measurements is generated by an experimental machine and the next challenging task is the selection of key experimental points which should represent the creep behavior of the studied alloy. These issues can be tackled by the introduction of universal automated procedure for the handling of experimental data. Such an approach should provide better possibilities to speed up the data preparation for their further analysis and reduce the human factor. The process includes creating a dataset, data cleaning, and data reduction. While the data cleaning procedure can be programmed easily [14], the selection of a representative sample is not a trivial task and requires the knowledge of practitioners. As mentioned before, one of the possible solutions for data reduction is based on change-point analysis. Change-
point analysis is one of the statistical approaches that have been widely adopted by researchers to detect so-called change-points which should indicate statistically significant changes in the behavior of studied phenomena in a quasi/non-stationary time series data in several fields such as climate change, image processing, etc [15–19]. A change-point is a point at which the parameters (for instance mean and variance) of an underlying distribution or the parameters of a model used to describe time series changes abruptly [20–23].

In this study, we establish an automated approach for cleaning and reduction of data acquired in creep experiments on Co-base superalloys [24]. In this direction, a method has been developed to treat mini-creep data [6] and then successfully extended to the double shear experiments with constant and variable loading conditions [25]; both for Ni-base superalloys.

2. Material and experiments

The comparably new class of $\gamma'$-strengthened Co-base superalloys is of high research interest since they are considered as a competitor to the conventional Ni-base superalloys. These alloys are utilized as turbine components such as vanes, blades, and disks, in the hottest regions of aero engines [7]. Their creep properties are essential to be known and as important design criterion, because these parts are exposed to very high stresses and temperatures. In this study, the experimentally investigated alloy VF80 is a single crystalline (SX) $\gamma'$-strengthened Co-base superalloy (Co-9Al-10.6W-2.9Ta in at.%). The SX bar was cast using a lab scale Bridgman casting unit and a withdrawal rate of 3mm/min. Subsequently, the bar was solution heat treated at 1350 °C for 24 h and aged at 900 °C for 100 h in vacuum followed by furnace cooling. Afterwards, cylindrical (001)-oriented samples were wire eroded from the bar and finally cut and ground to about 7.5 mm in height and about 5 mm in diameter. The creep tests were performed in a commercially available load cell with a maximum capacity of 100 kN. A linear variable differential transducer (LVDT) was used for the measurement of the change in length. Data points of force and displacement were recorded in different time intervals. The time intervals were given by the experimentalist according to the changing speed of the creep test. In the primary creep regime, where the creep behavior is changing quickly, more data points need to be recorded in a certain time interval in contrast to the secondary creep regime, where less data points per interval are sufficient and useful to reduce the noise. The resulting txt-file including data on force, displacement, and time was the starting point for the evaluation described in the following sections.

3. Methodology

As explained in the introduction, automatic data handling consists of the two main steps: data cleaning and the selection of representative data sample, which is the core part proposed in the automated data treatment procedure. During this step, the initial data are reduced from hundreds/thousands to a small set of approximately 20–60 entries (depending on the size of the whole dataset) without losing the valuable information about the progress of the experiment. In this section, we discuss the procedure of data handling in details.

All algorithms are implemented in R, which is an open-source ‘language and environment for statistical computing and graphics’ [26]. R has been selected for our analysis due to its broad applicability for data mining and a variety of built-in statistical and mathematical methods.

3.1. Data cleaning

Before the data can be used as an input for any statistical algorithm, data cleaning is required to obtain the correct values from the raw-file written by a creep machine. The raw file includes the information about conditions of the experiment and values of applied force $F(N)$, displacement $\delta$(mm), and initial area/length ($A_0$(mm$^2$))/$L_0$(mm)). The displacement $\delta$ reported from creep test has to be corrected due to the influence of machine stiffness and settling effects. The parameters $a$ and $b$ correct the displacement by equation (1).

$$\delta_{corr} = \delta + \frac{a}{b},$$

where $F(\delta) = a + b\delta$ in which $a$ and $b$ are intercept and slope respectively. Here, segmented regression provided by the R-package segmented [27, 28] is used to calculate $a$ and $b$. Moreover, true plastic strain ($\varepsilon_{pl}^{true}$) at each time step needs to be calculated in the following steps. First, we calculate technical strain (2)
Then, the true strain is calculated by

\[ \varepsilon_{\text{true}} = \ln(1 + \varepsilon_{\text{tech}}). \]  

Afterwards, technical stress is calculated by

\[ \sigma_{\text{tech}} = \frac{F}{A_0}. \]  

The dependency between technical stress \( \sigma_{\text{tech}} \) versus technical strain \( \varepsilon_{\text{tech}} \) curve can be derived from the linear regression model of the form

\[ \sigma_{\text{tech}}(\varepsilon_{\text{tech}}) = c + d\varepsilon_{\text{tech}}, \]

where \( c \) and \( d \) are the intercept and slope, respectively, and have to be estimated from creep experiments. Finally, we calculate the true plastic strain using equation

\[ \varepsilon_{\text{true}}^{\text{pl}} = \varepsilon_{\text{true}} - \frac{\sigma_{\text{tech}}}{d}. \]

Thereafter, the incubation period, and the data entries after rupture (if occurring) have to be removed. The incubation interval is characterized by the increase in the applied force measured on the specimen till a constant value is obtained as shown in figure 1. The rupture time is associated with the rapid decrease to zero of the applied force. In this case, machine errors describe the mistakenly measured values due to high sensitivity of the device. In other words, during this step the stationary force interval and the outliers need to be determined.

To prepare the data for change-points selection, assume \( Z = \{ Z_1, \ldots, Z_q \} \) is an initial dataset from the raw-file, where

\[ Z_i = \{ t_i, \delta_i, \ldots, F_i \}, \quad i = 1, 2, \ldots, q \]

and \( q \) is a number of observations. First, remove the negative forces to fulfill the experimental setup requirement:

\[ \bar{Z} = Z_i \subset Z: F_i \geq 0, \quad i = 1, 2, \ldots, q. \]

Then, it is essential to remove outliers and noisy data. Hereby, we assume the middle of the experiment represents its steady-state regime. Therefore, we find the median, which is the most robust statistics to the outliers \[\text{[3]}, \] of applied force in the dataset from the previous step

\[ F_\text{med} = \text{med}(F_i), \quad F_i \subset Z. \]

Then, the outlier and noisy data are removed by selecting the data entries where the applied force differs for less than 5% from the found median \( F_\text{med} \). The range of accepted error can be different depending on the case study

\[ Z_{\text{CUT}} = \bar{Z}_i \subset \bar{Z}: |F_i - F_\text{med}| < 0.05F_\text{med}. \]

Dataset \( Z_{\text{CUT}} \) obtained during this routine is written in the file called CUT. The results of the application of above described procedure are shown in figure 2. Here, the two red dashed lines determine the stationary force range and the blue points are the values stored in CUT file.

---

**Figure 1.** The loading period in the creep experiment for VF80, 950 °C, 400 MPa.
After the removal of noisy data and calculating the true plastic strain $\varepsilon_{\text{true} \text{pl}}$, it is necessary to recalculate time $t$ and strain $\varepsilon$ corresponding to the beginning of actual creep $(t_0, \varepsilon_0)$ using formulas

$$ t_i = t_i - t_0, \quad i \in \mathbb{N}, $$

where $t_i$ denotes the $i^{th}$ time point.

$$ \varepsilon_i^{\text{norm}} = \varepsilon_i^{\text{pl} \text{true}} - \varepsilon_i^{\text{pl} \text{true}}, \quad i \in \mathbb{N}. $$

Additionally, strain rate $\dot{\varepsilon}$ is defined by equation (9)

$$ \dot{\varepsilon} = \frac{1}{2} \frac{\varepsilon_{i+1}^{\text{norm}} - \varepsilon_i^{\text{norm}}}{t_{i+1} - t_i} + \frac{1}{2} \frac{\varepsilon_i^{\text{norm}} - \varepsilon_{i-1}^{\text{norm}}}{t_i - t_{i-1}}, \quad i \in \mathbb{N}. $$

The newly obtained dataset is recorded in the file NORM. The creep strain/time curve recorded in the NORM file has already the correct shape. However, it is necessary to select the representative sample in order to obtain a smooth curve for the strain rate/strain dependency.

### 3.2. Selection of representative data

As can be seen in figure 3(a), the creep strain/time has a non-trivial shape. Therefore, some additional data manipulations are required in order to capture complete creep behavior in time. An interval of the fast increase in creep strain during the first few minutes of the experiment is of concern for the research. However, the experimental data that belong to this interval can be considered as outliers for several methods. This occurs also
due to the small strain change during this stage compared to the whole experiment. Nevertheless, this part of the creep test is very important and has to be captured in the refined data. The idea on how to do it has only mathematical meaning. It is suggested to perform the data split into two subsets defined by the minimum of strain rate in the strain rate/strain curve [6]. Afterward, the algorithm of point selection is called separately on both subsets providing that a certain amount of data from each interval is selected. The obtained results of the data splitting procedure described above are shown in figure 3. The red dashed lines on both graphs display the point of the data split where the strain rate has the minimum value. After this procedure, the representative data points can be detected by change-point analysis.

3.2.1. Basic theory in change-point analysis
This problem is associated with time series analysis [22, 29]. A time series is an ordered sequence of values at equally spaced time intervals [30]. It is called stationary when all statistical properties like mean, variance, functional form, etc are constant over time. When at least one of the parameters of the series is changing with time, it is called non-stationary [30]. Thus the strain-time dependency obtained after data cleaning described earlier spans a non-stationary time series \( Y_{1:n} \), where \( n \) is a number of observation in the cleaned set, \( y_i = \epsilon_i \) is the value of creep strain \( \epsilon_i \) at measured time point \( t_i \). The mathematical interpretation of this problem can be formulated as detection of multiple change-points within non-stationary time series. By the term change-points, we consider those data points, which divide a dataset into distinct stationary segments [20, 22]. In this work, we assume that the number of change points \( m \) is known and fixed, while their positions \( \eta_{1:n} = (\eta_1, ..., \eta_m) \), \( m < n \) are to be estimated. All change-point positions are integer values between 1 and \( n - 1 \). The start and the end indices of the change-points are set as \( \eta_0 = 0 \) and \( \eta_{m+1} = n \). Change-points are ordered such that \( \eta_i < \eta_j \) if and only if \( i < j \). In the theory, multiple change-points detection can be obtained by extension of single change-point detection algorithms. Nevertheless, in practice, with the increasing size of a dataset, the number of possibilities to choose \( m \) change-points is increased rapidly. This makes the problem of multiple change-point estimation very computationally heavy [31, 32]. Taking into account the heterogeneous character of experimental set-ups and their different period of duration, we are interested in the application of such statistical methods which can be employed to both small and large datasets.

3.2.2. Implementation in R
A wide range of R-packages is available on CRAN repository [26] for the problem of multiple change-points analysis [33–36]. We have selected two applicable packages for creep data: wbs [34] and changepoint [33, 37]. Suitable methods from these packages are based on the binary segmentation method, which is considered to be the most widespread algorithm for multiple change-points estimation [32]. Binary segmentation can be used to extend the likelihood-ratio or any other statistic approach for single data point selection to multiple change-points case. It begins with applying the single change-point detection to the whole algorithm for the entire dataset. If no change-point is detected, the algorithm stops, otherwise, it splits the dataset into two segments and applies single point detection algorithm for both segments. The procedure is repeated until no more change-points are found [32]. The schematic of the binary segmentation method is shown in figure 4. In order to detect a single change-point within the method, the hypothesis testing needs to be performed. The null \( H_0 \) and the alternative \( H_1 \) hypotheses are formulated as follows:

\[
H_0 : \text{no change-point is determined,} \\
H_1 : \text{single change-point is determined.}
\]

To test the null hypothesis against the alternative one, a general test statistic \( w(\cdot) \) dependent on data, an estimator of change-point position \( \hat{\eta}(\cdot) \), and a rejection parameter \( \xi \) are required. If \( w( \cdot ) > \xi \), the statistical significance is declared, and the null hypothesis \( H_0 \) is rejected. This means that the change-point is detected and its position is determined by the estimator \( \hat{\eta}(y_{\cdot:}) \). The position of change-point \( \eta_i \) for the original dataset is calculated using \( \hat{\eta}(y_{\cdot:}) \). The binary segmentation method is computationally fast (typically order of \( O(n \log n) \)) and easy to implement. However, it is an approximate method and the results of change-point detection are significantly dependent on the choice of the rejection parameter \( \xi \). The procedure of single change-point detection is performed recursively. Therefore, the method can not be applied to the configuration of the dataset, where the spacing between change points is greater than \( n^{3/4} \). Additionally, each stage of detection is dependent on the previous one which can also be a disadvantage for non-sequential datasets [38]. The difference in results among the selected methods of both packages is explained by the difference in the general test statistic, the estimator for the change-points positions and the threshold parameter are summarized in table 1 [6].

Standard binary segmentation (sbs) method from the R-package wbs and binseg.mean.cusum (mean) method from the R-package changepoint identify changes in mean for the data where there is no assumption about the data distribution is made. Both approaches use the CUSUM (cumulative sum control chart) statistic which is defined by
\[ y_{s,b} = \frac{e - \eta}{\sqrt{n(e - s + 1)}} \sum_{t=s}^{n} y_t = \frac{\eta - s + 1}{\sqrt{n(e - \eta)}} \sum_{t=\eta+1}^{e} y_t, \]

where \( s \leq b < e \) with \( s \) is the first element of set, \( e \) is the last element of set, \( n = e - s + 1 \) and \( \eta \) is a variable of the unknown change-point position to define a test statistic \([20]\). On each step of the binary segmentation method the value of test statistic

\[ w(y_{s,e}) = \max_{\eta:s \leq \eta < e} \hat{y}_{s,e}^\eta, \]

is calculated. When it exceeds the value of threshold parameter \( \xi \), i.e. \( w(y_{s,e}) > \xi \), then the change-point is detected and its position can be determined by

\[ \hat{\eta} = \arg \max_{\eta:s \leq \eta < e} |\hat{y}_{s,e}^\eta|. \]

The \textit{mean} and \textit{sbs} methods are similar to each other, however, due to the difference in the choice of threshold parameter \( \xi \) they provide slightly different results. In method \textit{sbs} the threshold is calculated automatically from the given number of change-points, while the \textit{mean} requires the manually given value. Additionally, these methods could not provide suitable results applied for the whole dataset due to the special interest of researchers in the specific intervals of the data. Therefore, we split the data into two subsets by the local minimum of the derivative of the creep strain and then apply the methods to each subset separately. The selected points for each subset indicate the representative points for the whole dataset. \textbf{Binseg.var.css(var)} method from the package \textbf{changepoint} is used to find multiple changes in the variance for the data where the distribution of the input data is unknown and no assumption about it is made. In this method the test statistic for the binary segmentation algorithm is determined as

\[ w(y_{\hat{\eta},e}) = \max_{\eta:s \leq \eta < e} D_{\eta}, \]

where \( D_{\eta} = \frac{C_{\eta}}{C_{e-s+1}} - \frac{\eta}{e - s + 1}, \)

and \( C_{\eta} = \sum_{t=\eta+1}^{e} y_t^2 \) is a cumulative sum of squares. When it exceeds the value of the threshold parameter, i.e. \( w(y_{\hat{\eta},e}) > \xi \), the change-point location is estimated as

![Figure 4. Schematic demonstration of the binary segmentation method \([39]\).](image-url)
\[ \hat{\eta} = \arg \max_{\eta < \epsilon} |D_\eta|, \]

It requires a manual selection of the threshold like method \textit{mean}. Method \textit{var} is able to provide satisfying results of point selection without splitting the data into physical meaningful subsets. However, the difference in the results between described methods is not significant enough to count any of them unsuitable. Moreover, there is no unified criterion on how to determine the best approach among the suggested algorithms, except the subjective visual inspection. In this study, the criterion for selection of the most appropriate change-point algorithm is based on the evaluation of statistical similarity of two datasets [41].

3.2.3. Comparison of considered methods

For evaluation of selected methods to form the representative datasets, manually treated experimental data have been compared with the results applied by the algorithms described in section 3. To compare manually and automatically selected data, several similarity measures have been calculated. The main idea behind it is based on cluster analysis, which is one of the most common technique in data analysis. It identifies the subgroups (or cluster) in the dataset such that data points in the same subgroup are very similar according to a similarity measure while data points in different clusters are very different [42]. Some of clustering comparison metrics are Adjusted Rand Index (ARI), Fowlkes-Mallows Index (FMI), Jaccard Index (JI), and Variation of Information Index (VII) [43]. The higher the value of ARI, FMI, and JI the more similar are the two datasets, and the reverse is true for VII, where two perfectly matching partitions produce 0 score. For random clustering, ARI returns a value around zero (negative values are possible) and for perfectly matching clustering ARI is 1. FMI and JI values are strictly in [0, 1]. All these indexes show how two datasets are similar to each other. The function \texttt{dsClustCompare} in \texttt{R}-package \texttt{semiArtificial} [44] delivers the values of ARI, FMI, JI, and VII by entering two datasets as an input. In this study, the creep data points selected by practitioners are considered as the correct change points and are compared to the data obtained from each change-point algorithms. The method with the higher similarity index of ARI, FMI, and JI while lower index of VII are considered as the most appropriate method for selection of creep data.

4. Results and discussion

The algorithms \textit{sbs}, \textit{mean} and \textit{var} for change-points detection described in the previous section 3 were sequentially applied to the experiments. The automatically proceeded data were compared to the results of the manual points selection by the practitioner and are shown in figures 5–7. By the visual inspection (figures 5(a), 6(a), and 7(a)), it can be concluded that all three methods have been successful in the reproduction of the correct strain/time behavior. It can be observed that the selected representative data by method \textit{var} are distributed through the interval more uniformly, compared to the other methods, with more points at the beginning of the experiment and fewer points at the end (figure 7). However, for all three methods, the strain rate/strain dependency is represented by a smooth distribution without visible disturbance according to the figures 5(b), 6(b), and 7(b). The selected points correspond well to the results of the manual procedure. All three approaches provide good results in capturing the minimum strain rate \( \dot{\varepsilon}_{\min} \) as well as describing the early times of the creep.
Due to the similarity of the provided results, the most appropriate algorithm for point selection by visual inspection can only be determined subjectively. Therefore, we compared the data using similarity measurement discussed in previous section \textsuperscript{3}. Table \textsuperscript{2} reports the different metrics for all three algorithms. In addition, the corresponding R-package and computational time for each method are summarized. The results imply that overall \textit{var} performs better than \textit{mean} and \textit{sbs}. As can be seen, VII has the lowest value for \textit{var} while FMI and JI are slightly greater than the values for \textit{sbs}. Moreover, ARI for both \textit{var} and \textit{sbs} are equal. Therefore, \textit{sbs} is also a good candidate to be selected as the most appropriate method especially because of its lowest computational time among the other methods. On the other hand, the obtained result from \textit{mean} shows the least similarity with the manual data. In general, all three proposed algorithms for the points selection provide

\begin{table}
\centering
\caption{Calculated statistics for identifying similarity of the manually selected data with each statistical method for VF80, 950 °C, 400 MPa.}
\begin{tabular}{lcccccc}
\hline
Algorithm & VII & FMI & JI & ARI & \text{R-package} & \text{Computational time, s} \\
\hline
\textit{mean} & 0.709 & 0.829 & 0.707 & 0.662 & changepoint & 0.964 \\
\textit{sbs} & 0.445 & 0.910 & 0.835 & 0.822 & wbs & 0.936 \\
\textit{var} & 0.443 & 0.912 & 0.839 & 0.822 & changepoint & 0.983 \\
\hline
\end{tabular}
\end{table}

Figure 6. Comparison of \textit{mean} (blue) and manual (black) change-points selection methods for VF80, 950 °C, 400 MPa: (a) plastic strain versus time and (b) strain rate versus plastic strain.

Figure 7. Comparison of \textit{var} (blue) and manual (black) change-points selection methods for VF80, 950 °C, 400 MPa: (a) plastic strain versus time and (b) strain rate versus plastic strain.
appropriate results and any of these procedures can be used for the data treatment of creep experiments, nevertheless var is selected as the most appropriate model regarding the results in table 2.

5. Conclusion

In this study, the procedure for automated data treatment of Co-base creep experiments is proposed. For selection of representative data, change-point analysis technique based on three different criteria are applied to the experiments and compared to the data that was evaluated manually by the practitioners. This approach has been successfully applied to 40 more experiments performed at the Friedrich-Alexander-Universität Erlangen-Nürnberg within the SFB/Transregio 103 B3 project. It was also applied to Ni-base creep experiments (ca. 30 experiments) performed at Ruhr-Universität Bochum within the SFB/Transregio 103 A1 project and some results were published by Turchenko [6]. So, it means that the proposed procedure has been tested and verified on raw creep data from two different laboratories. This approach helps the practitioner to follow a standard rule and produce consistent and unique representative data for each experiment in a more efficient way.

Acknowledgments

The work was supported by the Deutsche Forschungsgemeinschaft through the projects B3 and T2 of the collaborative research center SFB Transregio 103, ‘From Atoms to Turbine Blades—a Scientific Approach for Developing the Next Generation of Single Crystal Superalloys’. We also acknowledge support by the DFG Open Access Publication Funds of the Ruhr-Universität Bochum.

Data availability

The manually and automatically selected representative data are available to download from https://data.mendeley.com/datasets/r53wxcvt5g/1.

ORCID iDs

Setareh Zomorodpoosh © https://orcid.org/0000-0002-9198-6156

References

[1] Shyr J and Spisic D 2014 Automated data analysis Comput. Stat. 6 359–66
[2] Müller H and Freytag I-C 2005 Problems, methods, and challenges in comprehensive data cleansing Professoren des Institute Für Informatik
[3] Nishet R, Elder J and Miner G 2009 Handbook of Statistical Analysis and Data Mining Applications (New York: Academic)
[4] Han J, Pei J and Kamber M 2011 Data Mining: Concepts and Techniques (Amsterdam: Elsevier) (https://doi.org/10.1016/C2009-0-61819-5)
[5] Kallrath J ed. 2012 Algebraic Modeling Systems: Modeling and Solving Real World Optimization Problems vol 104 (Berlin, Heidelberg: Springer) (https://doi.org/10.1007/978-3-642-23392-4)
[6] Turchenko E 2016 Automated data proceeding and its application to creep experiments Master’s Thesis Ruhr-Universität Bochum, Bochum, Germany
[7] Reed R C 2008 The Superalloys: Fundamentals and Applications (New York, USA: Cambridge university press) (https://doi.org/10.1017/CBO9780511541285)
[8] Wollgramm P, Bürger D, Parsa A, Neuking K and Eggeler G 2016 The effect of stress, temperature and loading direction on the creep behaviour of Ni-base single crystal superalloys miniature tensile specimens Materials at High Temperature 33 346–60
[9] Wollgramm P, Wu X and Eggeler G 2016 On the temperature dependence of creep behavior of Ni-base single crystal superalloys Superalloys 2016: Proc. 13th International Symposium on Superalloys ed M Hardy et al (Tennessee, USA: TMS) pp 711–8
[10] Latief F H and Kakehi K 2012 Effects of heat treatment and crystallographic orientation on creep behavior of aluminized nickel-base single crystal superalloy CM186LC Int. J. Electrochem. Sci. 7 9290–302
[11] Latief F H and Kakehi K 2013 Effects of Re content and crystallographic orientation on creep behavior of aluminized Ni-base single crystal superalloys Mater. Des. 49 485–92
[12] Rettig R, Matuszewski K, Müller A, Helmer H E, Ritter N C and Singer R F 2016 Development of a low-density rhenium-free single crystal nickel-based superalloy by application of numerical multi-criteria optimization using thermodynamic calculations Proc. of the 13th International Symposium of Superalloys pp 35–44
[13] Müller A, Roslyakova I, Sprenger M, Git P, Rettig R, Markl M, Körner C and Singer R F 2019 MultiOpt ++ : a fast regression-based model for the development of compositions with high robustness against scatter of element concentrations Modell. Simul. Mater. Sci. Eng. 27 2 (024001)
[14] Toulemonde E-L 2020 Data preparation: automated data preparation, R package version 0.4.3
[15] Amimikhangahi S, Wang T and Cook D J 2018 Real-time change point detection with application to smart home time series data IEEE Trans. Knowl. Data Eng. 31 1010–23
[16] Reeves J, Chen I, Wang X L, Lund R and Lu Q Q 2007 A review and comparison of changepoint detection techniques for climate data. *Journal of Applied Meteorology and Climatology* **46** 900–15

[17] Itoh N and Kurths J 2010 Change-point detection of climate time series by nonparametric method *Proc. of the World Congress on Engineering and Computer Science* vol 1, pp 455–8

[18] Arif S N A M, Mohsin M F M, Bakar A A, Hamdan A R and Abdullah S M S 2017 Change point analysis: a statistical approach to detect potential abrupt change. *Jurnal Teknologi* **79** 147–159

[19] Coulson D and Joyce I 2006 Indexing variability: a case study with climate change impacts on ecosystems. *Ecol. Indic.* **6** 749–69

[20] Page E S 1954 Continuous inspection schemes. *Biometrika* **41** 100–15

[21] Shumway R H and Stoffer D S 2006 Spectral analysis and filtering. *Time Series Analysis and Its Applications: With R Examples* (New York, USA: Springer) pp 174–270

[22] Aminikhanghahi S and Cook D J 2017 A survey of methods for time series change point detection. *Knowledge and Information Systems* **51** 339–67

[23] Truong C, Oudre I and Vayatis N 2019 Selective review of offline change point detection methods. *Signal Process.* **167** 107299

[24] Bezold A, Zoln N, Xue F, Zenk C, Neumeier S and Göken M 2020 On the precipitation-strengthening contribution of the Ta-containing Co3 (Al, W)-phase to the creep properties of γ/γ′ cobalt-base superalloys. *Metallurgical and Materials Transactions A* **51** 1567–74

[25] Almodallaleh M 2018 Anwendung von automatischer Datenverarbeitung auf Doppelscherversuche von Nickelbasissuperlegierung. *Master’s Thesis* Ruhr-Universität Bochum, Bochum, Germany

[26] R Core Team 2013 R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria

[27] Muggeo V M 2003 Estimating regression models with unknown break-points. *Statistics in Medicine* **22** 3055–71

[28] Muggeo V M 2008 segmented: an R package to fit regression models with broken-line relationships. *R News* **8** 20–5

[29] Jandhyala V, Fotopoulos S, MacNeill I and Liu P 2013 Inference for single and multiple change-points in time series. *J. Time Ser. Anal.* **34** 423–46

[30] Nason G P 2006 Stationary and Non-Stationary time Series Statistics in Volcanology ed H Mader and S C Coles vol 1 (United Kingdom: Geological Society of London) pp 129–421862392080

[31] Eckley I A, Fearnhead P and Killick R 2011 Analysis of changepoint models. *Bayesian Time Series Models* ed D Barber, A T Cemgil and S Chiappa (Cambridge: Cambridge University Press) 205–24

[32] Killick R, Eckley I A, Jonathan P and Chester U 2011 Effizient detection of multiple changepoints within an oceano-graphic time series. *Proc. of the 58th world Science Congress of ISI*

[33] Killick R, Haynes K and Eckley I A 2016 changepoint: an R package for changepoint analysis, R package version 2.2.2

[34] Baranowski R and Fryzlewicz P 2015 wbs: wild binary segmentation for multiple change-point detection, R package version 1.3

[35] Cleynen A, Rigail G and Koskas M 2016 Segmentor: a fast segmentation algorithm, R package version 2.0

[36] Chen H, Zhang N R and Chu L 2019 gSeq: graph-based change-point detection (g-Segmentation), R package version 0.6

[37] Killick R and Eckley I A 2014 Changepoint: an R package for changepoint analysis *Journal of Statistical Software* **58** 1–19

[38] Fryzlewicz P et al 2014 Wild binary segmentation for multiple change-point detection *The Annals of Statistics* **42** 2243–81

[39] Roslyakova I 2009 Analyse und Optimierung von Teilaspekten der Ultraform-Produktion *Master’s Thesis* Hochschule Fulda, Fulda, Germany

[40] Inclan C and Tiao G C 1994 Use of cumulative sums of squares for retrospective detection of changes of variance. *J. Am. Stat. Assoc.* **89** 913–23

[41] Ashby F G and Ennis D M 2007 Similarity measures *Scholarpedia* **2** 4116

[42] Xu R and Wunsch D 2008 *Clustering* vol 10 (New York: Wiley)

[43] Wagner S and Wagner D 2007 Comparing clusterings: an overview. *Universität Karlsruhe, Fakultät für Informatik Karlsruhe*

[44] Robnik-Sikonja M 2019 semiArtificial: generator of semi-artificial data, R package version 2.3.1