

**Anhang zu:**

**Verfahren zur Bildung  
und Beschreibung  
zustandshomogener  
Abschnitte und  
repräsentativer  
Kennzahlen für das  
Erhaltungsmanagement**

von

Dietmar Berthold  
Andreas Großmann

LEHMANN + PARTNER GmbH  
Erfurt

**Berichte der  
Bundesanstalt für Straßenwesen**

**Straßenbau Heft S 195**

**bast**

## **ANHANG 1**

**Übersicht zu den Verfahren zur Bildung homogener Abschnitte**



In der Übersicht werden zusammenfassend die Eigenschaften und besonderen Merkmale der einzelnen Verfahren aufgelistet, die eine Grundlage für die Entscheidung über die weitere Untersuchung bildeten. Die Nummerierung in der linken Spalte entspricht der Abschnittsnummer in Kapitel 3. Bei Verfahren, die nicht weiter untersucht werden, ist die letzte Spalte grau eingefärbt.

<b>Abschn.-nr.</b>	<b>Name des Verfahrens und dessen Merkmale</b>	<b>Folgerung</b>
<b>3.2.1</b>	<b>Verfahren nach Rübensam / Schulze</b>	Vergleichs- verfahren, wird untersucht
	- Stand der Technik in Deutschland	
	- Verfahren wurde für die Anwendung auf BAB konzipiert (ohne Ortsdurchfahrten)	
	- geringe Ausprägung von Trends, Randproblem	
	- geringe Schärfe der Homogenisierung	
	- Egalisierung der Zustandsdaten	
<b>3.2.2</b>	<b>Verfahren der kumulierten Summen</b>	keine weitere Untersuchung
	- einfache mathematische Handhabung	
	- Plausibilitätsüberprüfung eines Abschnittswechsels über statistischen Signifikanztest	
	- für größere Datenmengen ungeeignet, da grundlegend eine Visualisierung der Daten notwendig ist	
	- keine Berücksichtigung von Mindestabschnittslängen möglich	
	- Unterteilung in mehrere Unterabschnitte nicht möglich, wenn die jeweiligen Mittelwerte alle unter oder über dem Mittelwert des betreffenden Abschnittes liegen	
<b>3.2.3</b>	<b>Verfahren der kumulativen Differenzen (CDA)</b>	keine weitere Untersuchung
	- einfache mathematische Handhabung	
	- zahlreiche Unterabschnitte mit oft zu geringen Längen (keine Berücksichtigung von Mindestabschnittslängen)	
	- sehr zeitaufwändig für größere Datenmengen	
	- keine Unterabschnittsbildung möglich, wenn die jeweiligen Mittelwerte alle unter oder über dem Mittelwert des betreffenden Abschnittes liegen	
	- nicht geeignet für in zu kleinen Abständen aufgenommene Erfassungsdaten	
<b>3.2.4</b>	<b>Verfahren der absoluten Differenzen</b>	keine weitere Untersuchung!
	- einfache mathematische Handhabung	
	- Bildung zu vieler Unterabschnitte mit zu kurzen Abschnittslängen	

Abschn.-nr.	Name des Verfahrens und dessen Merkmale	Folgerung
<b>3.2.5</b>	<b>Verfahren LCPC</b>	keine weitere Untersuchung
	- kann in einem automatisch funktionierenden iterativen Prozess angewendet werden	
	- Eignung nur für normalverteilte Daten, einige Zustandsmerkmale entsprechen jedoch nicht dieser Verteilung	
	- keine systematischen Trends oder Wechselbeziehungen zwischen den erfassten Zustandsmerkmalen	
	- Erfassungsdaten müssen in gleichmäßigen Abständen erfasst worden sein	
	- Mindestdatenmenge erforderlich	
	- statistischer Parameter $\alpha$ frei wählbar, keine Empfehlung aus der Literatur; bei ungeeigneter Wahl ergibt sich ein nicht zielführendes Ergebnis	
<b>3.2.6</b>	<b>Klassifikations- und Regressionsbäume</b>	keine weitere Untersuchung
	- einfache mathematische Handhabung	
	- Festlegung von Mindestabschnittslängen möglich	
	- Bildung vieler Unterabschnitte, die Zusammenfassung ist zeitaufwändig und komplex	
	- Optimierungsproblem, wenn mehr als zwei Unterabschnitte gefunden werden	
	- ohne Vorglättung ist das Verfahren nicht zielführend	
	- nicht massendatentauglich	
<b>3.2.7</b>	<b>Circular Binary Segmentation (CBS)</b>	weitere Untersuchung
	- Erfolgreiche Anwendung in der Biologie zur Teilung eines Genoms, um Abschnitte mit gleicher Anzahl von DNA-Kopien zu erhalten	
	- relativ komplexer mathematischer Hintergrund	
<b>3.2.8</b>	<b>Verfahren mit Bayes'schem Ansatz</b>	weitere Untersuchung
	- Anwendung im österreichischen PMS	
	- Festlegung von Mindestabschnittslängen möglich	
	- Daten müssen flächendeckend vorhanden sein und extreme Ausreißer sollten vorab eliminiert werden	
	- Ansatz des Bayes'schem Theorem zur Berechnung von bedingten Wahrscheinlichkeiten	

Abschn.-nr.	Name des Verfahrens und dessen Merkmale	Folgerung
<b>3.2.9</b>	<b>Qualitätskontrollkarten</b>	keine weitere Untersuchung
	- Segmentierungsverfahren, bei dem anhand der geschätzten Varianz bestimmt werden kann, ob eine Abschnittsgrenze gezogen werden muss oder nicht	
	- Ergebnis maßgeblich abhängig von zu wählender Spannweite	
	- viele frei wählbare Parameter	
	- keine einheitliche Betrachtungsweise der unterschiedlichen Merkmale möglich	
<b>3.2.10</b>	<b>Clusterverfahren</b>	kein alleinstehendes Verfahren
	- Berücksichtigung mehrerer Kriterien (hier Zustandsmerkmale) für die Abschnittsbildung	
	- in der Regel kein eindeutiges Segmentierungsergebnis	
	- Anwendung von Ansätzen aus diesem Verfahren sinnvoll für die Ermittlung repräsentativer Kennzahlen der Zustandshomogenität	
<b>3.2.11</b>	<b>Wavelet – Entrauschen</b>	kein alleinstehendes Verfahren
	- gut geeignet für Glättung der Daten	
	- bessere Ausgangsgrundlage für Segmentierung	
	- Verringerung der Anzahl möglicher Segmente	
	- nicht hinreichend als eigenständiges Verfahren, Vorstufe für andere Verfahren	
<b>3.2.12</b>	<b>Verfahren nach Zhang und Flintsch</b>	keine weitere Untersuchung
	- zweistufiges Verfahren: Glätten der Daten durch Eliminieren von Ausreißern, anschließend Verfahren der Klassifikations- und Regressionsbäume	
	- Festlegung von Mindestabschnittslängen möglich	
	- Komplexität steigt enorm bei großen Datenmengen (Optimierung der Zusammenfassung der zu kurzen Teilabschnitte, rechnerische Komplexität), siehe auch 3.2.6	
<b>3.2.13</b>	<b>Verfahren nach D'Apuzzo und Nicolosi</b>	keine weitere Untersuchung
	- sehr schnelle Berechnungsvorgänge möglich bei „gewöhnlichem“ Datenumfang	
	- Bestimmung der optimalen Lage der Change-Points wenn Anzahl vorab bekannt	
	- Voruntersuchung des Datensatzes notwendig	
	- nicht geeignet bei großen Datenmengen	

Abschn.-nr.	Name des Verfahrens und dessen Merkmale	Folgerung
3.2.14	<b>Verfahren nach Lindner, Stephan, Rübensam, Schulze und Staroste</b>	keine weitere Untersuchung
	- für Aufbaudaten entwickelt, vorliegend nicht zielführend	
3.2.15	<b>Verfahren nach Kunze, Rübensam, Schulze und Staroste</b>	keine weitere Untersuchung
	- Erweiterung des Verfahrens nach Rübensam / Schulze (siehe 3.2.1) - benötigt Aufbaudaten, diese sind derzeit netzweit nicht qualitätsgesichert verfügbar	
3.2.16	<b>Verfahren nach Hajdin, Botzen, Lindenmann und Schiffmann</b>	vorerst keine weitere Untersuchung
	- heuristische Optimierung der Eingabedaten auf Basis der Auswerteabschnitte	
	- sowohl zustands- als auch maßnahmenbezogene Kriterien fließen in die Abschnittsbildung mit ein	
	- wirtschaftliches Optimum auf 40 Jahre  - Ansatz verfolgt die Idee, dass es eine Menge von Maßnahmen gibt, die so verändert wird, dass eine Reduzierung der Netzgesamtkosten erreicht wird	

## **ANHANG 2**

### **Beschreibung des favorisierten ZHA-Verfahrens**



## Allgemeines

Der Programmablauf für die Segmentierung in zustandshomogene Abschnitte untergliedert sich in drei Teilprozesse.

### Vorbereitung

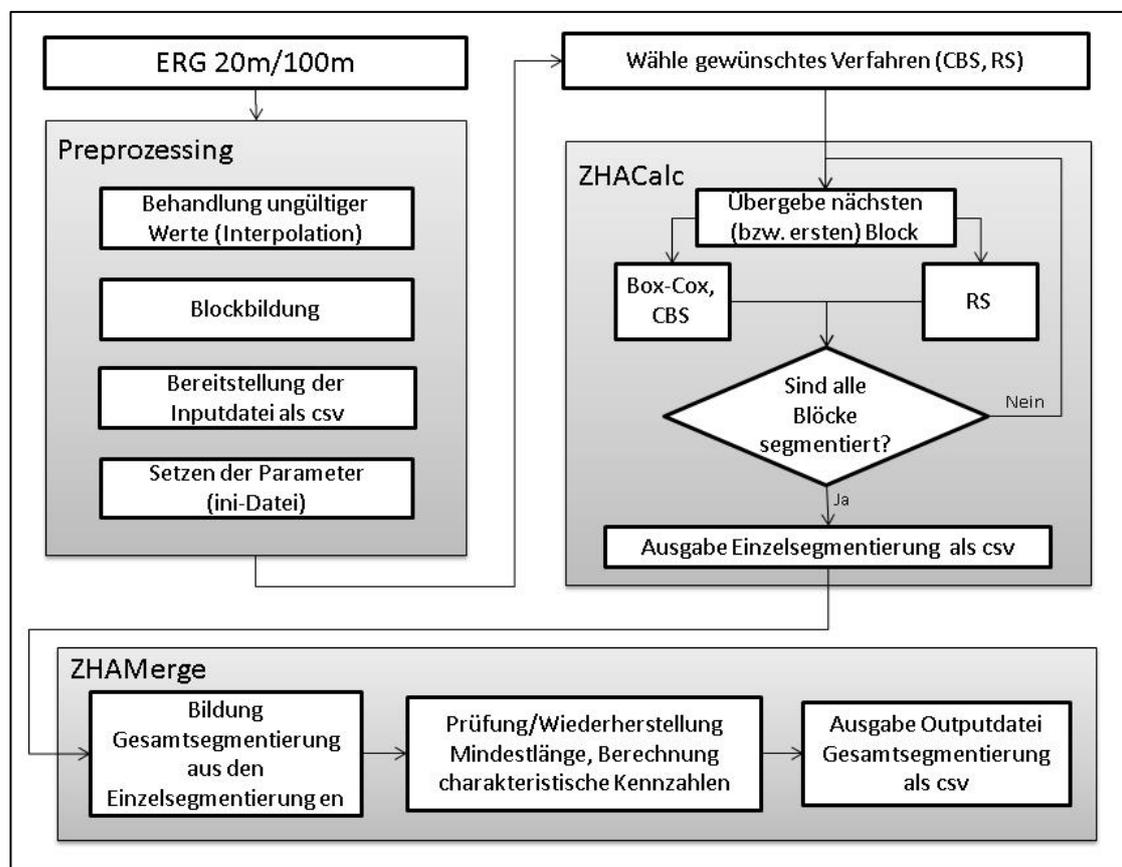
Der erste Prozess besteht im Preprocessing der Inputdatei für jede Zustandsgröße. Diese werden entsprechend den folgenden Kapiteln „Inputdaten“, „Ungültige Werte“ und „Blockbildung“ für die Einzelsegmentierung vorbereitet. Die Struktur der Inputdatei im Format csv (comma separated values) ist in „Beschreibung der Inputdatei“ zu finden.

### Einzelsegmentierung

Im zweiten Schritt werden diese Inputdateien mit der **C++-Software „ZHACalc“** verarbeitet. Diese Software führt für jede Zustandsgröße einen eigenen Berechnungslauf durch, dementsprechend ist die Wahl des Verfahrens (CBS oder RS) flexibel. Für den Ablauf der Verfahren siehe die entsprechenden Pseudocodes und Flowcharts weiter unten. Die Begründungen für die empfohlenen Steuerparameter für jedes Verfahren sind dem Bericht zu entnehmen. Im Ergebnis der Berechnungen wird für jede Zustandsgröße eine eigene Outputdatei ebenfalls im Format csv bereitgestellt.

### Gesamtsegmentierung

Im letzten Prozess werden die Einzelergebnisse pro Zustandsgröße mit Hilfe des **R-Skriptes<sup>1</sup> „ZHAMerge“** zu einer Gesamtsegmentierung zusammengefasst. Für alle auf diese Weise entstehenden Segmente wird die charakteristische Kennzahl berechnet. Falls die Mindestsegmentlänge unterschritten wird, wird diese mittels des Gesamtsegmentierungsverfahrens nach Rübensam/Schulze wiederhergestellt. Das Ergebnis wird anschließend wiederum als eine Outputdatei (siehe „Beschreibung der Outputdatei, Gesamtsegmentierung“) bereitgestellt.



<sup>1</sup> R ist eine freie Programmiersprache für statistische Berechnungen und Grafiken.

## **Vorbereitung**

### **Inputdaten**

Die Datengrundlage bildet eine Datei in der Struktur der ZEB-Ergebnisdatei ERG.dbf (bei Verwendung des einheitlichen 20m-Rasters muss diese separat aus den Rasterrohdaten erzeugt werden). Jede Zeile der Datei repräsentiert einen Teilabschnitt eines Fahrstreifens einer Fahrbahn im gewählten Raster (20m / 100m). Die zustandshomogenen Abschnitte können Netzknoten übergreifend berechnet werden.

Die ZEB-Ergebnisdatei liegt normalerweise bereits in einer geeigneten Sortierung vor (nach Straße, Fahrstreifenlage, Fahrstreifennummer, Abschnittsfolgennummer, Von-Station, Bis-Station); falls nicht, ist genau diese Sortierung herzustellen. Die Sortierung ist durch Hinzufügen einer ID zu fixieren.

### **Ungültige Werte**

Die Inputdatei (Basis ERG.dbf) kann ungültige/undefinierte Werte für einzelne Zustandsgrößen enthalten (Codierung mit -90 bis -99, Gültigkeitsflag der ZEB). Falls es sich um nur ein oder zwei aufeinander folgende ungültige/undefinierte Werte handelt, werden diese durch lineare Interpolation der benachbarten gültigen Werte ersetzt. Diese Maßnahme kann verhindern, dass im Endergebnis an dieser Stelle mehrere kurze Segmente entstehen.

### **Blockbildung**

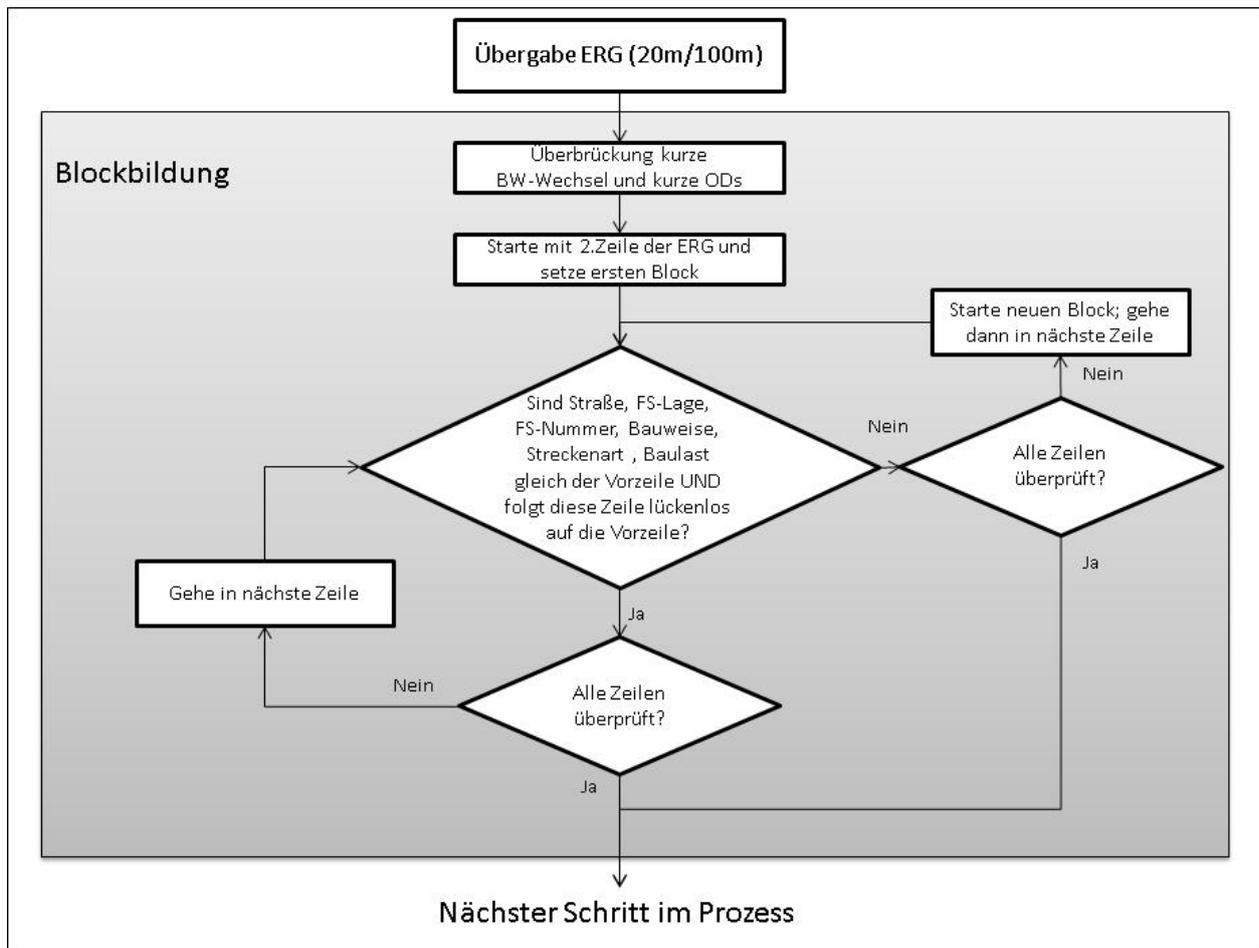
Die Datei wird in einem weiteren vorgeschalteten Schritt in sogenannte Blöcke unterteilt. Ein Block ist ein linear zusammenhängender Bereich desselben Fahrstreifens in gleicher Bauweise, Streckenart und Baulast, d.h. Blockgrenzen werden gesetzt bei:

- tatsächlichen Unterbrechungen im Straßenverlauf (insbesondere bilden auch Seitenarme und Bereiche mit getrennt verlaufenden Fahrbahnen jeweils eigene Blöcke),
- Wechsel des Fahrstreifens (gemäß obiger Sortierung zuerst alle L1, dann alle L2, ..., alle R1),
- Wechsel der Bauweise (Asphalt, Beton, Pflaster, unbekannt),
- Wechsel der Streckenart (OD, FS),
- Wechsel der Baulast (HBL, GBL).

Optional sind auch weitere Blockgrenzen möglich (z.B. am Wechsel der Dienststelle). Technisch wird der "Block" als eine zusätzliche Spalte der Inputdatei mit einer BlockID „block“ angelegt.

Die Bildung von Segmenten durch das ZHA-Verfahren erfolgt stets innerhalb eines Blocks, d.h. ein Segment ist immer ein Teil eines Blocks. Blockübergreifende Segmente werden nicht gebildet (Blockgrenzen sind Vorab-Grenzen im Sinne des Verfahrens nach Rübensam / Schulze).

Kurze Bauweisenwechsel (z.B. ein oder zwei Teilabschnitte mit Bauweise Asphalt inmitten einer Betonstrecke = Fall bituminöser Teilersatz) oder kurze Ortsdurchfahrten (bis zu einer bestimmten OD-Länge, Inputvariable) werden in die benachbarten Blöcke integriert. Dadurch sollen zu viele kurze Blöcke und damit zu viele kurze zustandshomogene Abschnitte von vornherein vermieden werden.



### Herstellen der Symmetrie bei normalverteilten Zustandsgrößen

Für die Verfahren CBS und BAYES wird eine symmetrische Normalverteilung der Inputdaten vorausgesetzt. Falls die Zustandsgröße zwar normalverteilt ist, aber keine gute Symmetrie aufweist, dann kann die Anwendung einer bestimmten Box-Cox-Transformation Abhilfe schaffen. Mit geeigneter Wahl deren Parameter  $\lambda_1$  und  $\lambda_2$  ist es fast immer möglich eine annähernde Symmetrie herzustellen und somit die Voraussetzungen des Verfahrens CBS besser zu erfüllen.

$$x_{trans} = \begin{cases} \frac{(x + \lambda_2)^{\lambda_1} - 1}{\lambda_1} & \text{für } \lambda_1 \neq 0 \\ \log(x + \lambda_2) & \text{für } \lambda_1 = 0 \end{cases}$$

Empfohlene Parameter  $\lambda_1$  und  $\lambda_2$  pro Zustandsgröße (für die nicht-normalverteilten Zustandsgrößen des TP 3 ist diese Frage gegenstandslos) sind:

ZG	$\lambda_1$	$\lambda_2$
ZGAUN	0,0 (log. Transformation)	0,08
ZGSPT	0,0 (log. Transformation)	0,05
ZGSPH	0,0 (log. Transformation)	0,05
ZGGRI	2,0 (quadr. Transformation)	0,2

### Beschreibung der Inputdatei (\*.csv)

block	= Blocknummer, ganzzahlig, positiv (siehe oben bei Blockbildung)
idx	= laufender Index ID, ganzzahlig, positiv (Primärschlüssel/Sortierschlüssel)
klasse	= Straßenklasse (*)
strasse	= Straßenbezeichnung (*)
sort	= Abschnittsfolgennummer, numerisch (für NK-übergreifende Blockbildung)
aoa	= Abschnitts-Kennung = VNK & NNK (*)
aoalen	= Abschnittslänge in Meter, ganzzahlig
vst	= Von-Station in Meter, ganzzahlig (*)
bst	= Bis-Station in Meter, ganzzahlig (*)
snr	= alternative Nummerierung des Fahrstreifens, redundant zu lage + fstrnr
lage	= Lage zur Fahrbahnachse ('L' links, 'R' rechts) (*)
fstrnr	= Fahrstreifennummer, numerisch, 1, 2, 3 oder 4 (*)
bauw	= Bauweise 'A', 'B', 'P' oder 'X' (*)
odfs	= Streckenart 'F' oder 'O' (*)
baulast	= Baulast 'H' oder 'G' (*)
zgaun	= Zustandsgröße AUN in cm <sup>3</sup> (*)
zgspt	= Zustandsgröße MSPT in mm (*) – alternativ SPT_sig
zgsph	= Zustandsgröße MSPH in mm (*) – alternativ SPH_sig
zggrl	= Zustandsgröße GRI_80, GRI_60 oder GRI_40 (*) – alternativ GRI_sig
zgriss	= Zustandsgröße RISS in % (*)
zgrsfa	= Zustandsgröße RSFA in % (*)
zglqrp	= Zustandsgröße LQRP in % (*)
zglqrl	= Zustandsgröße LQRL in m (*)
zgrsfb	= Zustandsgröße RSFB in % (*)

Spalten mit (\*) stammen direkt aus der ZEB-Ergebnisdatei.

### Beschreibung der Outputdatei (\*.csv) - Einzelsegmentierung

Die Outputdatei übernimmt die Struktur der Inputdatei. Zusätzlich werden pro Zustandsgröße drei neue Spalten angefügt.

*Beispiel:*

*Für die Zustandsgröße zgaun werden die Spalten zgaun\_hanr, zgaun\_avg und zgaun\_std hinzugefügt, analog für alle anderen Zustandsgrößen.*

zg*_hanr	laufende Nummer des zustandshomogenen Abschnitts
zg*_avg	Mittelwert der Zustandsgröße im zustandshomogenen Abschnitt
zg*_std	Standardabweichung der Zustandsgröße im zustandshomogenen Abschnitt

### Beschreibung der ini-Datei (Textformat)

In der ini-Datei werden die Eingabe- und Ausgabedateien des ZHA-Verfahrens sowie alle Steuerparameter festgelegt.

- das für die Zustandsgröße zu verwendende Segmentierungsverfahren (CBS, RS)
- Steuerparameter für CBS:
  - kritischer Wert  $\alpha$  aus dem Intervall [0,5 ; 0,9999]
  - Parameter  $\lambda_1$  und  $\lambda_2$  der Box-Cox-Transformation

- Steuerparameter für RS:
  - kritische Werte  $d_{\text{krit}}$ ,  $v_{\text{krit}}$  (abhängig von der Zustandsgröße)
  - Parameter  $q$  und  $d$  (zur Bildung des gleitenden Mittelwerts und der dividierten Differenzen)
- Steuerparameter für BAYES:
  - Schwellenwert  $p_{\text{krit}}$ , Teilblocklänge  $\text{seqpart}$ , Überlappungslänge  $\text{overlap}$
  - Parameter  $\lambda_1$  und  $\lambda_2$  der Box-Cox-Transformation
- Mindestlänge  $n_{\text{Min}}$  in Anzahl Raster (Mindestsegmentlänge  $L_{\text{Min}} = n_{\text{Min}} \cdot \text{Rasterlänge}$ )

Empfohlene Steuerparameter pro Zustandsgröße:

ZG	Verfahren	$\alpha$	$d_{\text{krit}}$	$v_{\text{krit}}$
ZGAUN	CBS	0,99	-	-
ZGSPT	CBS	0,99	-	-
ZGSPH	CBS	0,99	-	-
ZGGRI	CBS	0,99	-	-
ZGRISS	RS	-	5,0	10,0
ZGRSFA	RS	-	5,0	10,0
ZGLQRP	RS	-	10,0	20,0
ZGLQRL	RS	-	2,0	4,0
ZGRSFB	RS	-	10,0	20,0

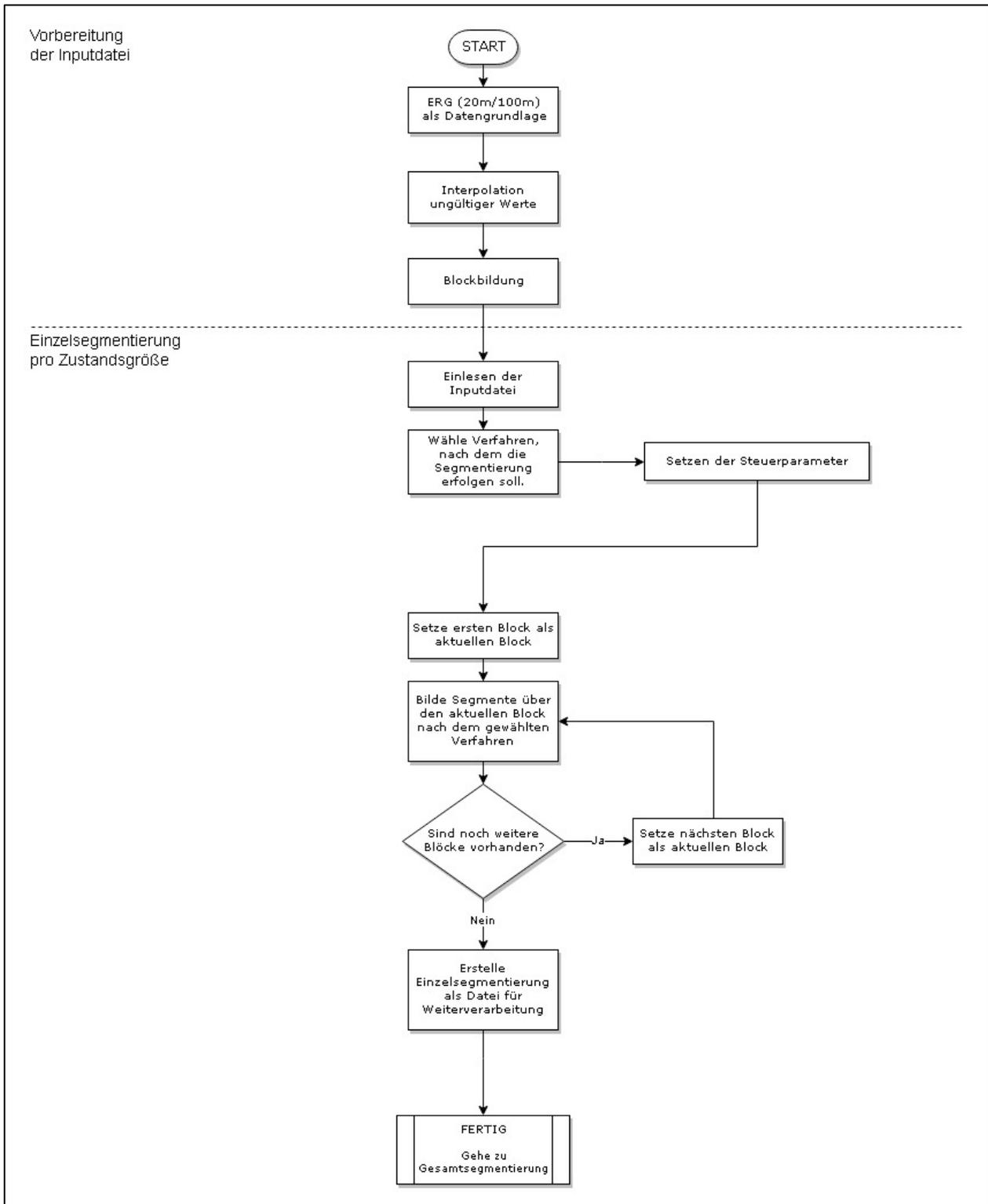
#### Beispiel einer ini-Datei

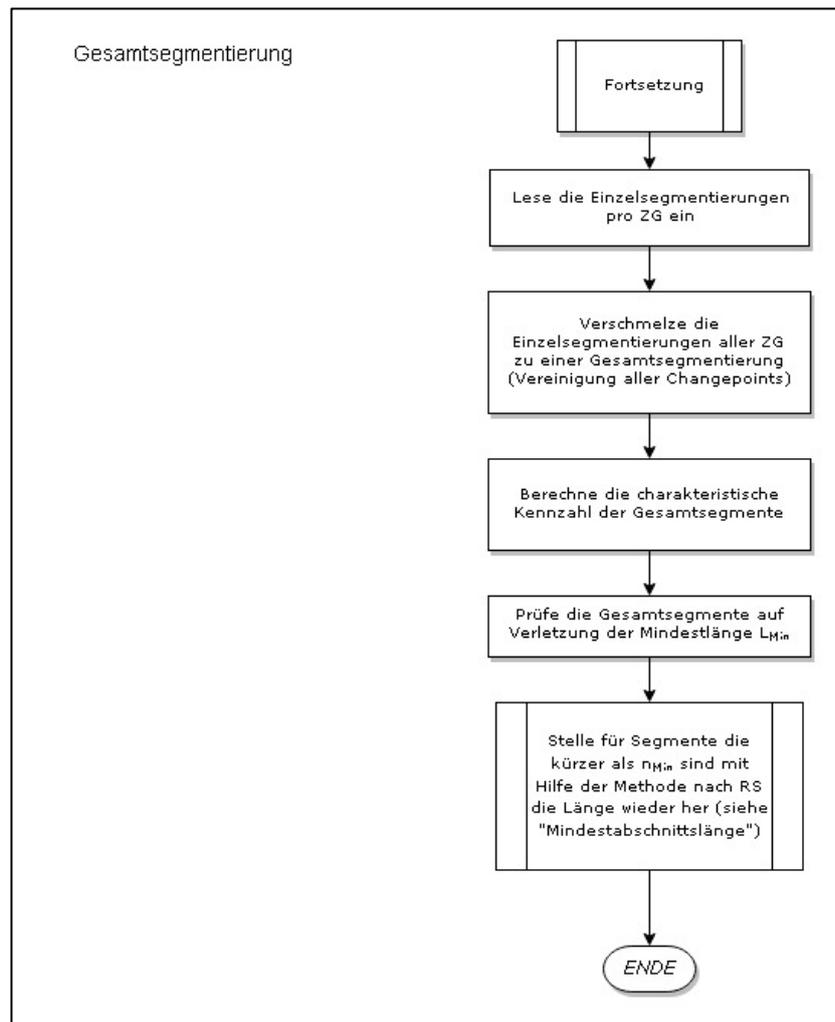
```

[Inp-Output]
input=zhacalc_inp_zgriss_100m_bab_by.csv
output=zhacalc_rs_out_zgriss_100m_bab_by_lmin020.csv
log=zhacalc_rs_out_zgriss_100m_bab_by_lmin020.log
[Methode (rs, cbs, bayes)]
methode=rs
[Variables-RS]
q=3
d=3
dgmkrit=5.0
vgmkrit=10.0
[Variables-Bayes]
seqpart=30
overlap=10
pkrit=0.8
[Variables-CBS]
alpha=0.95
[Variables-Global]
lMin=020
[Box-Cox-Transformation]
lambda1=1
lambda2=1

```

## Programmablauf – Flowchart





## Einzelsegmentierung

### Verfahren CBS (Pseudocode)

Inputdaten: Block  $\{x_1, x_2, \dots, x_n\}$  mit  $n$  Elementen (nach Box-Cox-Transformation),  $n_{min}$ ,  $\alpha \in [0,5 ; 0,9999]$

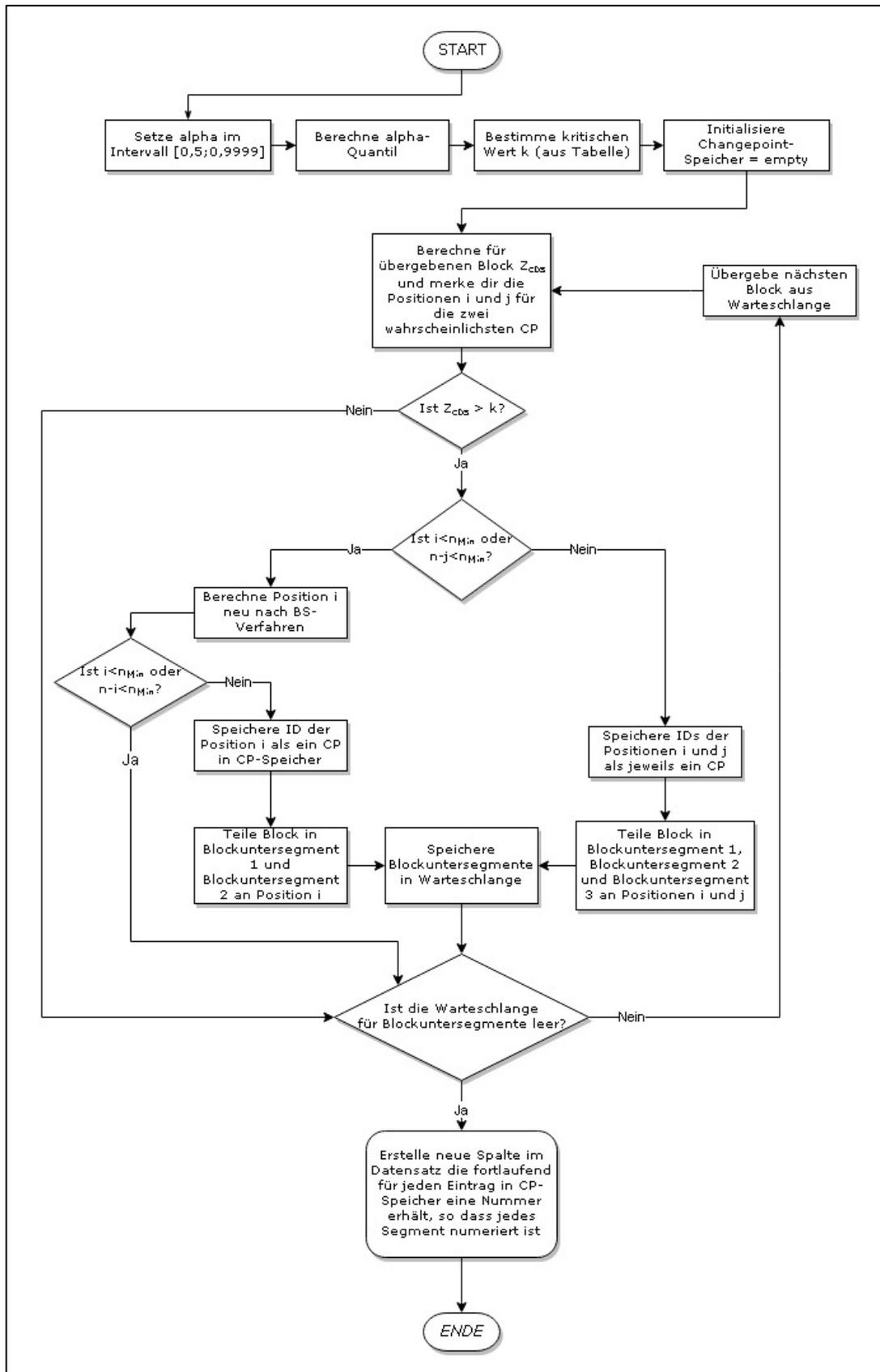
Formel 1:  $\bar{x}_r = \sum_{i=1}^r x_i, 1 \leq r \leq n$

Formel 2:  $Z_{i,j} = (\sqrt{1/(j-i) + 1/(n-j+i)})^{-1} * (\frac{\bar{x}_j - \bar{x}_i}{j-i} - \frac{\bar{x}_n - \bar{x}_j + \bar{x}_i}{n-j+i}), 1 \leq i < j \leq n$

Formel 3:  $Z_i = (\sqrt{1/i + 1/(n-i)})^{-1} * (\frac{\bar{x}_i}{i} - \frac{\bar{x}_n - \bar{x}_i}{n-i}), 1 \leq i < n$

1	<b>Berechne</b> Mittelwert $\mu$ und Standardabweichung $\sigma$ des Blocks
2	<b>Setze</b> $k := \mu + \sigma * (\alpha\text{-Quantil der Standardnormalverteilung } N(0,1) )$
3	<b>Setze</b> CP-Speicher := (leere) Liste, blockposition := ID vom ersten Wert im Block
4	<b>Definiere Funktion</b> CBS-Schritt(block){ Bestimme $i$ und $j$ für $Z_{CBS} = \max_{1 \leq i < j \leq n}  Z_{ij} $ (siehe Formel 2) <b>Falls</b> $Z_{CBS} > k$ <b>Falls</b> $i < n_{min}$ oder $n - j < n_{min}$ Bestimme $i$ nach BS-Verfahren $Z_{BS} = \max_{1 \leq i < n}  Z_i $ (siehe Formel 3) Setze $j = 0$ <b>Falls</b> $i < n_{min}$ oder $n - i < n_{min}$ Tue nichts. <b>Sonst</b> block.u1 := Liste mit $block[1], block[2], \dots, block[i-1]$ als Elemente block.u2 := Liste mit $block[j], block[j+1], \dots, block[n]$ als Elemente CBS-Schritt(block.u1) CP-Speicher.addElement(i+ blockposition) CBS-Schritt(block.u2) <b>Sonst</b> block.u1 := Liste mit $block[1], block[2], \dots, block[i-1]$ als Elemente block.u2 := Liste mit $block[j], block[j+1], \dots, block[j-1]$ als Elemente block.u2 := Liste mit $block[j], block[j+1], \dots, block[n]$ als Elemente CBS-Schritt(block.u1) CP-Speicher.addElement(i+ blockposition) CBS-Schritt(block.u2) CP-Speicher.addElement(j+ blockposition) CBS-Schritt(block.u3) <b>Sonst</b> Tue nichts. }
5	<b>Rufe</b> CBS-Schritt(Block) auf //CP-Speicher enthält nach dem Befehl alle gefundenen CP im Block
6	<b>Setze</b> Output-Tabelle := Input-Tabelle
7	<b>Erstelle</b> neue Spalte an der Output-Tabelle (haNr_zg***)
8	HANummer := 1 <b>Solange</b> Länge der Output-Tabelle(Zeilenzahl) > Laufindex := 1 <b>Falls</b> Output-Tabelle(ID-Spalte).hasElement(Laufindex) in ID-Speicher HANummer := HANummer + 1 Output-Tabelle(haNr_zg***) .writeElement(HANummer) <b>Sonst</b> Output-Tabelle(haNr_zg***) .writeElement(HANummer) Laufindex := Laufindex + 1
9	<b>Return</b> Output-Tabelle

## Verfahren CBS (Flowchart)



### Verfahren Bayes (Pseudocode)

Inputdaten: Block  $\{x_1, x_2, \dots, x_n\}$  mit  $n$  Elementen (nach Box-Cox-Transformation),  $n_{min}$ , Teilblocklänge  $sgp$ , Überlappungslänge  $ovl$ , Schwellenwert  $p_{krit}$

Formeln nach Thomas [35]:

Berechnung von Hilfsgrößen:

$$\begin{aligned}
 \bar{x}_{\langle t,r \rangle} &= (r-1)^{-1} \sum_{t=2}^r x_t & R_{\langle t,r \rangle} &= \sum_{t=2}^r (x_t - \bar{x}_{\langle t,r \rangle})^2 \\
 \bar{x}_{\langle t,n-r \rangle} &= (n-r-1)^{-1} \sum_{t=r+1}^{n-1} x_t & R_{\langle t,n-r \rangle} &= \sum_{t=r+1}^{n-1} (x_t - \bar{x}_{\langle t,n-r \rangle})^2 \\
 \bar{x}_{\langle t,n \rangle} &= (n-2)^{-1} \sum_{t=2}^{n-1} x_t & R_{\langle t,n \rangle} &= \sum_{t=2}^{n-1} (x_t - \bar{x}_{\langle t,n \rangle})^2 \\
 \bar{x}_{\langle t-1,r \rangle} &= (r-1)^{-1} \sum_{t=2}^r x_{t-1} & R_{\langle (t-1),r \rangle} &= \sum_{t=2}^r (x_t - \bar{x}_{\langle t,r \rangle})(x_{t-1} - \bar{x}_{\langle t-1,r \rangle}) \\
 \bar{x}_{\langle t+1,n-r \rangle} &= (n-r-1)^{-1} \sum_{t=r+1}^{n-1} x_{t+1} & R_{\langle (t+1),n-r \rangle} &= \sum_{t=r+1}^{n-1} (x_t - \bar{x}_{\langle t,n-r \rangle})(x_{t+1} - \bar{x}_{\langle t+1,n-r \rangle}) \\
 \bar{x}_{\langle t-1,n \rangle} &= (n-2)^{-1} \sum_{t=2}^{n-1} x_{t-1} & R_{\langle (t-1),n \rangle} &= \sum_{t=2}^{n-1} (x_t - \bar{x}_{\langle t,n \rangle})(x_{t-1} - \bar{x}_{\langle t-1,n \rangle}) \\
 \bar{x}_{\langle t+1,n \rangle} &= (n-2)^{-1} \sum_{t=2}^{n-1} x_{t+1} & R_{\langle t-1,r \rangle} &= \sum_{t=2}^r (x_{t-1} - \bar{x}_{\langle t-1,r \rangle})^2 \\
 & & R_{\langle t+1,n-r \rangle} &= \sum_{t=r+1}^{n-1} (x_{t+1} - \bar{x}_{\langle t+1,n-r \rangle})^2 \\
 & & R_{\langle t-1,n \rangle} &= \sum_{t=2}^{n-1} (x_{t-1} - \bar{x}_{\langle t-1,n \rangle})^2 \\
 & & R_{\langle (t,t+1),n \rangle} &= \sum_{t=2}^{n-1} (x_t - \bar{x}_{\langle t,n \rangle})(x_{t+1} - \bar{x}_{\langle t+1,n \rangle}) \\
 & & R_{\langle t+1,n \rangle} &= \sum_{t=2}^{n-1} (x_{t+1} - \bar{x}_{\langle t+1,n \rangle})^2
 \end{aligned}$$

Berechnung der a-posteriori Wahrscheinlichkeiten  $p(Mr|CP)$  und  $p(\text{change})$ :

$$\begin{aligned}
 p(Mr|CP) &= p(M_r) \times \frac{\Gamma\left(\frac{r-3}{2}\right) \Gamma\left(\frac{n-r-3}{2}\right)}{\sqrt{(r-1)(n-r-1)} \sqrt{R_{\langle t-1,r \rangle} \cdot R_{\langle t+1,n-r \rangle}}} \\
 &\times \left\{ R_{\langle t,r \rangle} - \frac{[R_{\langle (t,t-1),r \rangle}]^2}{R_{\langle t-1,r \rangle}} \right\}^{-(r-3)/2} \\
 &\times \left\{ R_{\langle t,n-r \rangle} - \frac{[R_{\langle (t,t+1),n-r \rangle}]^2}{R_{\langle t+1,n-r \rangle}} \right\}^{-(n-r-3)/2} \\
 B_{r0,orig} &= \frac{1}{4\pi} \sqrt{\frac{3}{2}} \left[ \frac{n-2}{(r-1)(n-r-1)} \right]^{1/2} \frac{\Gamma\left(\frac{r-3}{2}\right) \Gamma\left(\frac{n-r-3}{2}\right)}{\Gamma\left(\frac{n-4}{2}\right)} \frac{R_{\langle t-1,n \rangle}}{\sqrt{R_{\langle t-1,r \rangle} \cdot R_{\langle t+1,n-r \rangle}}} \\
 &\times \frac{\left\{ R_{\langle t,n \rangle} - \frac{[R_{\langle (t,t-1),n \rangle}]^2}{R_{\langle t-1,n \rangle}} \right\}^{(n-6)/2}}{\left\{ R_{\langle t,r \rangle} - \frac{[R_{\langle (t,t-1),r \rangle}]^2}{R_{\langle t-1,r \rangle}} \right\}^{(r-3)/2} \left\{ R_{\langle t,n-r \rangle} - \frac{[R_{\langle (t,t+1),n-r \rangle}]^2}{R_{\langle t+1,n-r \rangle}} \right\}^{(n-r-3)/2}}
 \end{aligned}$$

$$B_{r0} = \frac{1}{2} (B_{r0,orig} + B_{r0,rev})$$

$$p(\text{change}) = \frac{(n-7)^{-1} \sum_{\text{all } r} B_{r0}}{[(n-7)^{-1} \sum_{\text{all } r} B_{r0}] + 1} = \frac{\sum_{\text{all } r} B_{r0}}{(n-7) + \sum_{\text{all } r} B_{r0}}$$

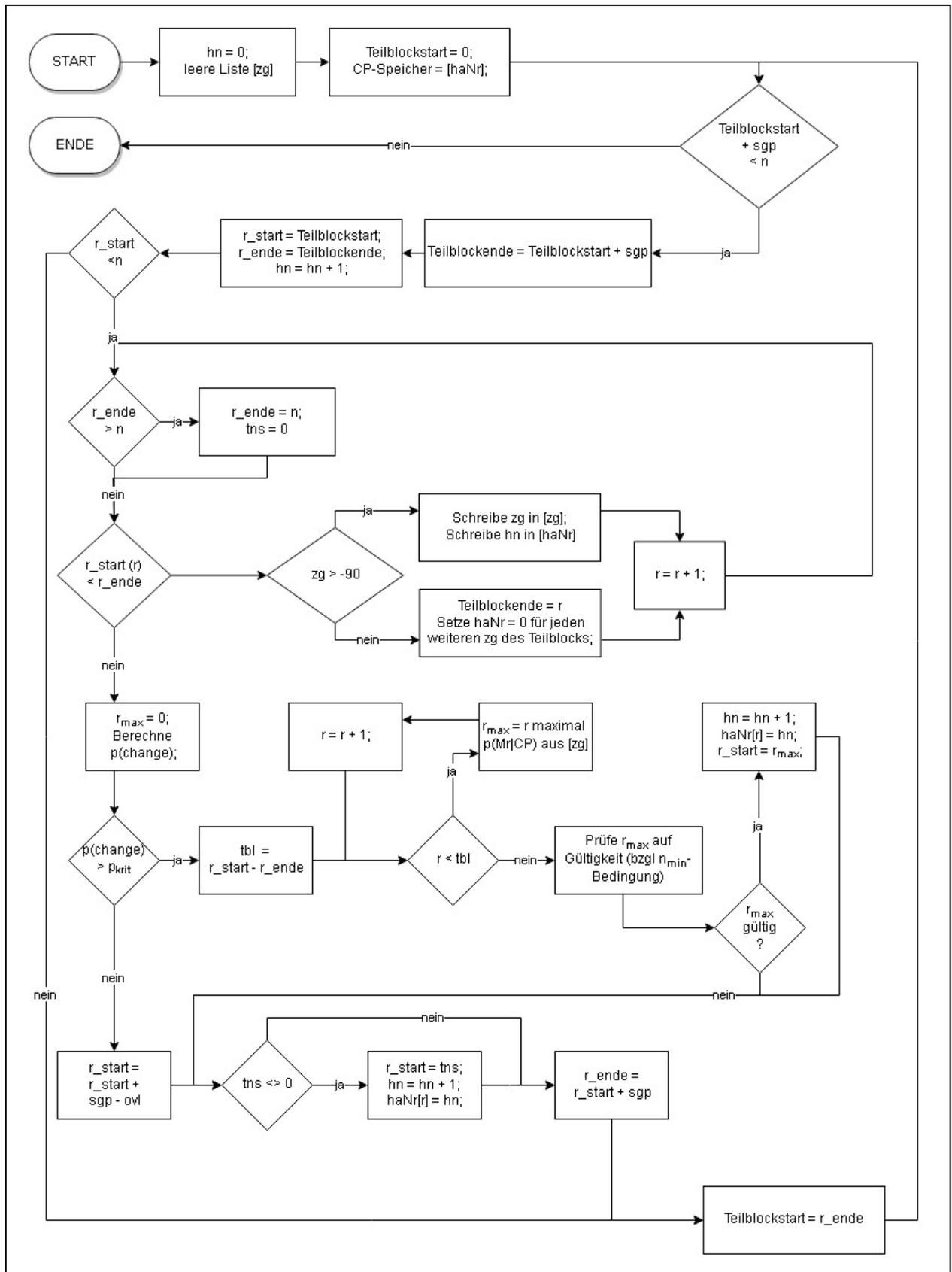
### Verfahren Bayes (Pseudocode – First Run)

1	<b>Setze</b> $hn = 0$ , CP-Speicher = (leere) Liste [haNr], Liste [zg], Teilblockstart = 0
2	<b>Solange</b> Teilblockstart + sgp < n <b>Setze</b> Teilblockende = Teilblockstart + sgp <b>Setze</b> r_start = Teilblockstart, r_ende = Teilblockende $hn = hn + 1$  <b>Solange</b> r_start < n <b>Falls</b> r_ende > n $r\_ende = n$ $tns = 0$  <b>Solange</b> r_start (r) < r_ende <b>Falls</b> zg > -90 Schreibe zg in Liste [zg] Schreibe hn in Liste [haNr]  <b>Sonst</b> Teilblockende = r <b>Solange</b> Teilblockende < n und zg > -90 $haNr = 0$ $r = r + 1$  $tns = r$ $r = r + 1$ $r_{max} = 0$  <b>Berechne</b> p(change) nach o.g. Formel  <b>Falls</b> p(change) > $p_{krit}$ <b>Solange</b> $r < (r\_start - r\_ende)$ <b>Ermittle</b> r aus Liste [zg] wo $p(M_r CP)$ maximal nach o.g. Formel $r_{max} = r$ (= Position des CP) $r = r + 1$  <b>Falls</b> $r_{max}$ gültig (bzgl. $n_{min}$ -Bedingung) $hn = hn + 1$ $HaNr[r] = hn$ $r\_start = r_{max}$  <b>Sonst</b> $r\_start = r\_start + sgp - ovl$  <b>Falls</b> $tns \neq 0$ $r\_start = tns$ $hn = hn + 1$ $haNr[r] = hn$ $r\_ende = r\_start + sgp$  Teilblockstart = r_ende

## Verfahren Bayes (Pseudocode – Subsequent Run)

1	<b>Setze</b> nochange = 0, durchlauf = 0, change = true
2	<b>Solange</b> durchlauf < 1000 und nochange < 50 <b>Setze</b> change = false, durchlauf = durchlauf + 1, Teilblockstart = 0  ===== First Part
2.1	<b>Setze</b> r_start = Teilblockstart  <b>Solange</b> r_start < n r_ende = Teilblockende <b>Schreibe</b> zg > -90 in Liste [zg] <b>Berechne</b> p(change) nach o.g. Formel  <b>Falls</b> p(change) > p <sub>krit</sub> <b>Ermittle</b> r aus Liste [zg] wo p(M <sub>r</sub>  CP) maximal nach o.g. Formel <b>Falls</b> r <sub>max</sub> gültig (bzgl. n <sub>min</sub> -Bedingung) change = true Passe [haNr] an  <b>Sonst</b> r_ende = r_start r = r + 1
2.2	===== Second Part  <b>Setze</b> r_start = Teilblockstart, r_ende = r_start  <b>Solange</b> r_start < n CP1 = r_ende, CP2 = Teilblockende des Nachbarteilblocks <b>Falls</b> CP1 <> CP2  <b>Falls</b> zg > -90 <b>Schreibe</b> zg in [zg] <b>Berechne</b> p(change) nach o.g. Formeln <b>Falls</b> p(change) > p <sub>krit</sub> <b>Ermittle</b> r aus Liste [zg] wo p(M <sub>r</sub>  CP) maximal nach o.g. Formel <b>Falls</b> r <sub>max</sub> + r_start < CP1 und r <sub>max</sub> gültig Zähle folgende haNr hoch change = true CP1 = r <sub>max</sub> + r_start <b>Falls</b> r <sub>max</sub> + r_start > CP1 und r <sub>max</sub> gültig Zähle folgende haNr herunter change = true CP1 = r <sub>max</sub> + r_start  <b>Sonst</b> change = true Zähle folgende haNr herunter CP1 = r_ende  r_start = CP2
3	<b>Falls</b> change = false nochange = nochange + 1

## Verfahren RS (Flowchart – First Run)



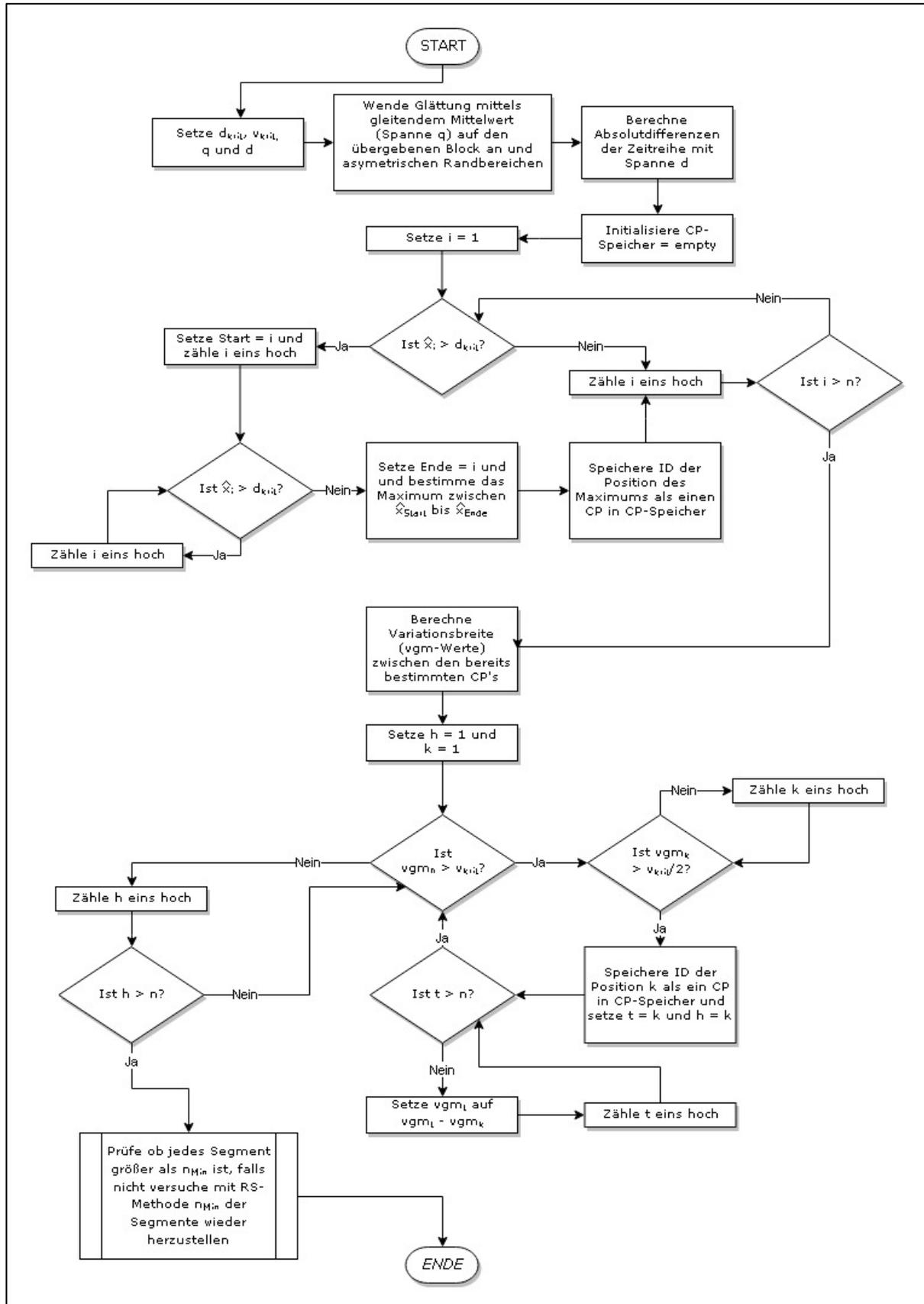


## Verfahren RS (Pseudocode)

Inputdaten: Block  $\{x_1, x_2, \dots, x_n\}$  mit n Elementen,  $n_{min}$ ,  $d$ ,  $q$ ,  $d_{krit}$ ,  $v_{krit}$

1	<p><b>Für</b> i von 1 bis n // Berechnung der geglätteten Zeitreihe <math>\tilde{x}</math></p> <p><b>Falls</b> <math>i \leq q</math></p> $\tilde{x}_i = \frac{1}{2q+1} \sum_{r=1-i}^q x_{i+r}$ <p><b>Sonst</b></p> <p><b>Falls</b> <math>n - i &lt; q</math></p> $\tilde{x}_i = \frac{1}{2q+1} \sum_{r=-q}^{n-i} x_{i+r}$ <p><b>Sonst</b></p> $\tilde{x}_i = \frac{1}{2q+1} \sum_{r=-q}^q x_{i+r}$
2	<p><b>Für</b> i von 1 bis n // Berechnung der Reihe der Absolutdifferenzen <math>\hat{x}</math></p> <p><b>Falls</b> <math>i \leq d</math></p> $\hat{x}_i =  \tilde{x}_{i+d} - \tilde{x}_1 $ <p><b>Sonst</b></p> <p><b>Falls</b> <math>n - i &lt; d</math></p> $\hat{x}_i =  \tilde{x}_n - \tilde{x}_{i-d} $ <p><b>Sonst</b></p> $\hat{x}_i =  \tilde{x}_{i+d} - \tilde{x}_{i-d} $
3	CP-Speicher := (leere) Liste; Start := 0; Ende := 0
4	<p><b>Für</b> i von 1 bis n</p> <p><b>Falls</b> <math>\hat{x}_i &gt; d_{krit}</math> und Start = 0 Start := i</p> <p><b>Falls</b> <math>\hat{x}_i &lt; d_{krit}</math> und Start <math>\neq</math> 0 Ende := i</p> <p><b>Falls</b> Start <math>\neq</math> 0 und Ende <math>\neq</math> 0 Bestimme <math>\max_{Start \leq c &lt; Ende} \hat{x}_c</math> und merke c CP-Speicher.addElement(c + blockposition) Start := 0 Ende := 0</p>
5	Start := 1; Ende := 0
6	<p><b>Für</b> j von 1 bis m := CP-Speicher.length()+1</p> <p><b>Falls</b> Start = 1 Ende := CP-Speicher.getElementOnPosition(j)</p> <p><b>Sonst</b></p> <p><b>Falls</b> j = m Ende := n + blockposition</p> <p><b>Sonst</b> Start := CP-Speicher.getElementOnPosition(j-1) Ende := CP-Speicher.getElementOnPosition(j)</p> <p><math>v_{gm_j} = \max(\tilde{x}_{Start}, \dots, \tilde{x}_{Ende}) - \min(\tilde{x}_{Start}, \dots, \tilde{x}_{Ende})</math> //Berechnung Variationsbreite <math>v_{gm}</math></p>
7	<p><b>Für</b> h von 1 bis n</p> <p><b>Falls</b> <math>v_{gm_h} &gt; v_{krit}</math></p> <p><b>Für</b> k von 1 bis j</p> <p><b>Falls</b> <math>v_{gm_k} &gt; \frac{v_{krit}}{2}</math> CP-Speicher.addElement(k + blockposition)</p> <p><b>Für</b> t in k+1 bis n <math>v_{gm_t} - v_{gm_k}</math> Setze k auf h+1</p>
8	Prüfe ob jeder Abschnitt größer als $n_{min}$ ist, falls nicht versuche mit RS-Methode die Mindestsegmentlängen wieder herzustellen (siehe Pseudocode Mindestsegmentlänge)
9	Fertig

Verfahren RS (Flowchart)



## Gesamtsegmentierung

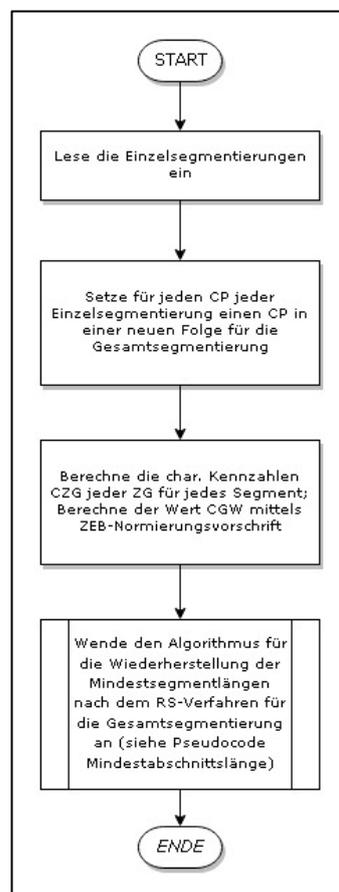
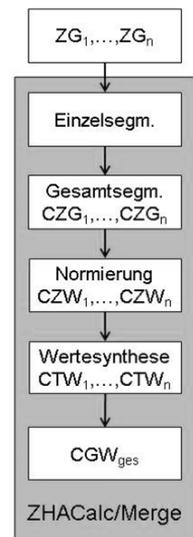
Die Gesamtsegmentierung erfolgt nach Vorliegen aller Einzelsegmentierungen. Dabei werden zunächst alle Changepoints der Einzelsegmentierungen vereinigt. Da dies in der Regel die Mindestlänge vieler Segmente zerstört, muss diese im nachfolgenden Schritt durch Verschiebung einzelner Changepoints wieder hergestellt werden (die Verschiebung erfolgt im jeweiligen Raster der ERG). Um zu entscheiden, in welche Richtung die Verschiebung erfolgen muss, wird die charakteristische Kennzahl der Segmente berechnet.

Eine Segmentgrenze (Changepoint) wird nur innerhalb desselben Blocks verschoben (keine blockübergreifenden Änderungen). Infolge einer CP-Verschiebung sollen auch keine Segmente kleiner als die Mindestlänge werden. Sind beide benachbarten Abschnitte eines CPs größer als die Mindestlänge, dann wird mittels der Summe der Absolutbeträge der Differenzen der charakteristischen Kennzahlen entschieden, in welche Richtung der CP verschoben wird.

Die Berechnung der charakteristischen Kennzahlen erfolgt nach nebenstehendem Schema (siehe dazu auch Abschnitt 4.4 des Berichts). Als Berechnungsvorschriften für die charakteristische Kennzahl wurden das arithmetische Mittel und das p-Quantil implementiert. Der Wert p ist eine Programmvariable. Beim Übergang von den charakteristischen Zustandsgrößen (CZG) zu den charakteristischen Zustandswerten (CZW) wird die Normierungsvorschrift der ZEB zu Grunde gelegt.

Mit dem R-Skript ZHAMerge werden die Einzelsegmentierungen folgendermaßen zusammengefasst:

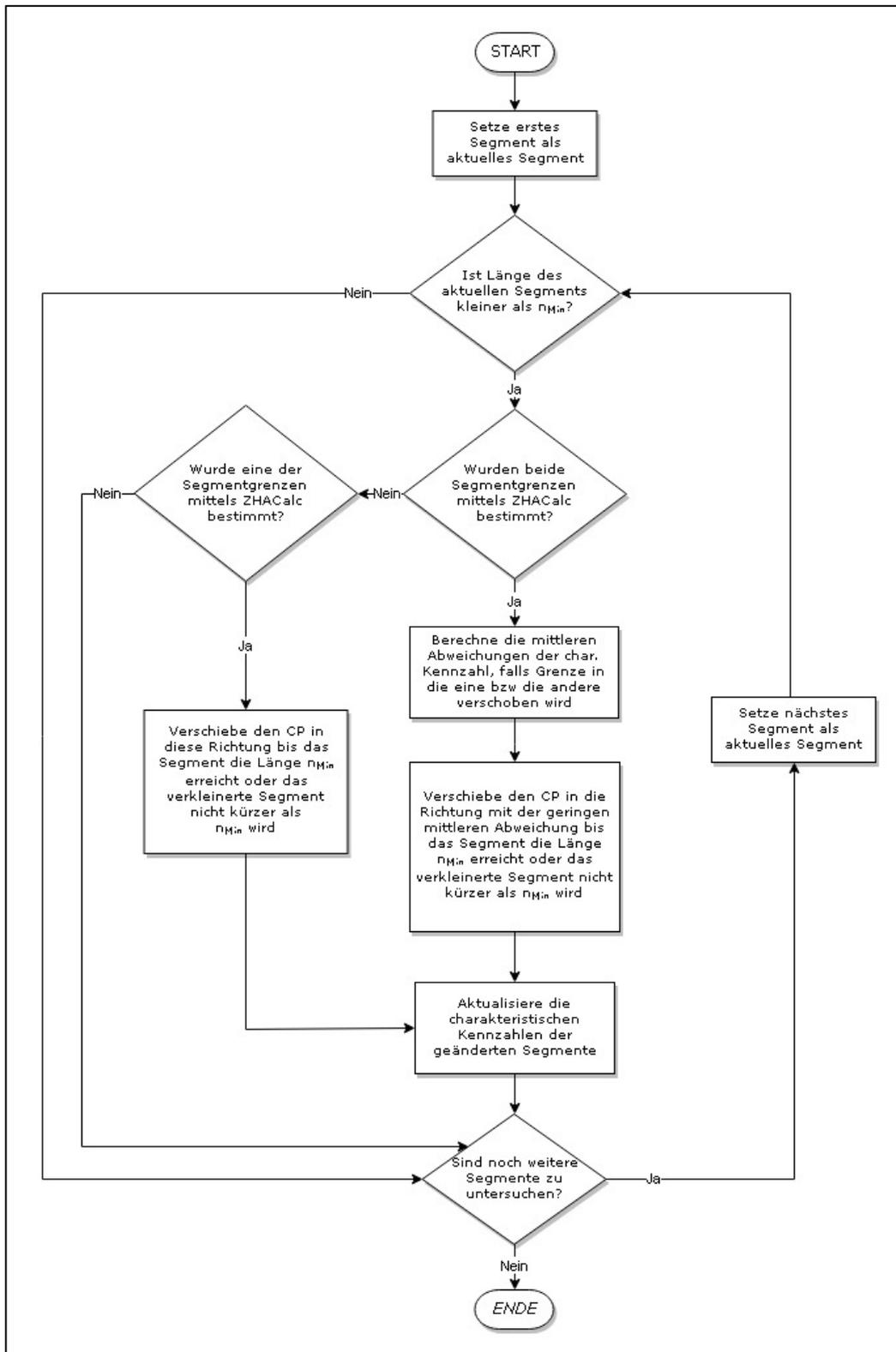
### Verfahren ZHAMerge (Flowchart)



**Verfahren Mindestsegmentlänge (Pseudocode)**

1	<p><b>Für</b> jedes Segment</p> <p><b>Falls</b> Länge des Segments <math>&lt; n_{min}</math></p> <p><b>Falls</b> beide Segmentgrenzen als Changepoints mittels ZHACalc bestimmt wurden</p> <p>Berechne mittlere Abweichung nach links (=: mAL)</p> <p>Berechne mittlere Abweichung nach rechts (=: mAR)</p> <p><b>Falls</b> mAL <math>&lt;</math> mAR</p> <p>Verschiebe linken Changepoint nach links bis das Segment größer als <math>n_{min}</math> oder das links benachbarte Segment gleich <math>n_{min}</math> wird</p> <p><b>Sonst</b></p> <p>Verschiebe rechten Changepoint nach rechts bis das Segment größer als <math>n_{min}</math> oder das rechts benachbarte Segment gleich <math>n_{min}</math> wird</p> <p><b>Sonst</b></p> <p><b>Falls</b> genau eine der beiden Segmentgrenzen als Changepoint mittels ZHACalc bestimmt wurde</p> <p>Verschiebe diesen Changepoint in diese Richtung bis das Segment größer als <math>n_{min}</math> oder das benachbarte Segment gleich <math>n_{min}</math> wird</p> <p><b>Sonst</b></p> <p>Tue nichts</p> <p>Aktualisiere die charakteristische Kennzahlen der veränderten Segmente</p>
2	Fertig

## Verfahren Mindestsegmentlänge (Flowchart)



**Beschreibung der Outputdatei (\*.csv) - Gesamtsegmentierung**

Diese Outputdatei besitzt die Spalten der Outputdatei der Einzelsegmentierung pro Zustandsgröße. Zusätzlich werden weitere Spalten angelegt: Pro Zustandsgröße eine Spalte für den charakteristischen Zustandswert (CZW) dieser Zustandsgröße in Bezug auf die Gesamtsegmentierung und zwei weitere Spalten zur Darstellung der Gesamtsegmentierung (charakteristische Kennzahl CGW und laufende Nummer).

*Beispiel:*

*Für die Zustandsgröße zgaun wird die Spalte ges\_zgaun hinzugefügt (analog für alle anderen Zustandsgrößen). Die Spalten gesamt und hanr\_gesamt enthalten dann die Beschreibung der Gesamtsegmentierung.*

ges_zgaun	charakteristischer Zustandswert CZW der Zustandsgröße im zustandshomogenen Abschnitt bezogen auf die Gesamtsegmentierung
gesamt	charakteristische Kennzahl $CGW_{ges}$ des zustandshomogenen Abschnitts der Gesamtsegmentierung
hanr_gesamt	laufende Nummer des zustandshomogenen Abschnitts der Gesamtsegmentierung