

# Analyse und Bewertung von Tragwerksprozessen – Detektion von Verzweigungspunkten

Martin Liebscher<sup>1</sup>, Michael Beer<sup>2</sup>, Wolfgang Graf<sup>3</sup>

<sup>1</sup>Dynamore GmbH, Stuttgart

<sup>2</sup>Department of Civil Engineering, National University of Singapore

<sup>3</sup>Institut für Statik und Dynamik der Tragwerke, TU Dresden

## **Abstract:**

In diesem Beitrag wird ein Verfahren zur Detektion von Verzweigungspunkten in unscharfen Tragwerksprozessen vorgestellt. Alle ermittelten reellwertigen Realisierungen eines unscharfen Tragwerksprozesses werden in eine zutreffende Anzahl Zeitschritte diskretisiert. Die Detektion von Verzweigungspunkten erfolgt zu jedem Zeitschritt mit Hilfe von Clusterverfahren. Es wird das deterministische k-Medoid Clusterverfahren eingesetzt. Die unüberwachte Klassifikation gelingt durch Einsatz eines Gütemaßes. Die Kenntnis von Verzweigungspunkten unterstützt die Analyse eines unscharfen Tragwerksprozesses und ermöglicht u.a. die Identifikation mechanischer Ursachen.

## **Keywords:**

Tragwerksprozesse, Verzweigungspunkte, Clusteranalyse, Unschärfe

## 1 Detektion von Verzweigungspunkten in Tragwerksprozessen

Tragwerksprozesse sind das Resultat der Abbildung von zeitabhängigen Tragwerksparametern  $\underline{x}(\tau)$  auf zeitabhängige Tragwerkantworten  $\underline{z}(\tau)$  (Tragwerksprozesse).

$$\mathbf{f}: \underline{x}(\tau) \rightarrow \underline{z}(\tau) \quad (1)$$

Das deterministische Berechnungsmodell stellt dabei den Abbildungsoperator  $\mathbf{f}$  dar. Sind die zeitabhängigen Tragwerksparameter  $\underline{x}(\tau)$  unscharf, können sie mit Hilfe eines geeigneten Unschärfemodells mathematisch beschrieben werden. Weisen Tragwerksparameter z.B. stochastische Eigenschaften auf, können sie als stochastische Prozesse  $\underline{X}(\tau)$  modelliert werden. Mit Hilfe einer stochastischen Tragwerksanalyse unter Verwendung des deterministischen Berechnungsmodells werden stochastische Tragwerksprozesse  $\underline{Z}(\tau)$  erhalten.

$$\underline{X}(\tau) \rightarrow \underline{Z}(\tau) \quad (2)$$

Wird beim Entwurf von Tragwerken neben der Erfüllung notwendiger Entwurfsnebenbedingungen auf ein robustes Tragverhalten [3] im Zeitbereich oder zu ausgewählten Zeitpunkten orientiert, so ist die Analyse des gesamten unscharfen Tragwerksprozesses erforderlich. Zur Charakterisierung und Bewertung unscharfer Tragwerksprozesse sind u. a. Informationen über die Existenz von Verzweigungspunkten und die Zeitpunkte  $\tau_{v_i}$  ihres Auftretens wesentlich. Als einführendes Beispiel für die Identifikation der Verzweigungspunkte  $v_i$  und deren Auswertung dient die Crash-Simulation eines Fahrzeugbauteils. Der Crash

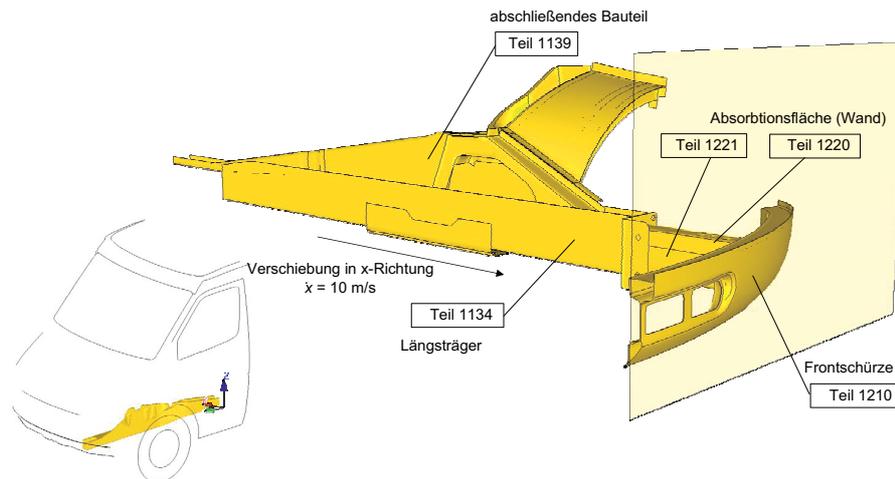


Abbildung 1: Baugruppe eines Fahrzeugs nach [7]; Mit freundlicher Genehmigung der DaimlerChrysler AG

des in Abb. 1 dargestellten Bauteils wird numerisch mit einer konstanten Vorschubgeschwindigkeit  $v_0$  auf eine starre Wand simuliert. Vier Blechdicken als Geometrieparameter des Bauteils und ein Materialparameter (Fließspannung) unterliegen produktionsbedingten Schwankungen. Die Unschärfe dieser Tragwerksparameter wird mit der Charakteristik Zufälligkeit modelliert, die als normalverteilt angenommen wird. Die verwendeten Wahrscheinlichkeitsverteilungsparameter  $\mu_i$  und  $\sigma_i$  der Normalverteilung  $F_i(x_i, \mu_i, \sigma_i)$  sind in Tab. 1 wiedergegeben. Die Krafteinwirkung des Bauteils auf die starre Wand wird als stochastischer Tragwerksprozeß  $Z(\tau)$  erhalten. Die ermittelten Realisierungen  $z(\tau)$  des stochastischen Prozesses sind in Abb. 2 dargestellt. Die Realisierungen  $z(\tau)$  des stochastischen Prozesses sind zu Beginn des Betrachtungszeitraums  $\tau = 10\text{ms}$  weitgehend homogen, sie bilden eine gemeinsame initiale Prozeßgruppe  $P_0(\tau) = M(\tau)$ . Die erste deutliche Diversifikation des Strukturverhaltens tritt zum Zeitpunkt  $\tau_{v_1} = 27,5\text{ms}$  auf. Ab diesem Zeitpunkt (Verzweigungspunkt  $v_1$ ) ist eine Unterscheidung der Realisierungen in zwei in sich homogene Prozeßgruppen  ${}_0P_1(\tau)$  und  ${}_0P_2(\tau)$  mit  ${}_0P_1(\tau) \cap {}_0P_2(\tau) = P_0(\tau)$  möglich. Für jede Prozeßgruppe wird ein repräsentativer Tragwerksprozeß ( $z_1(\tau) \in {}_0P_1(\tau)$  und  $z_2(\tau) \in {}_0P_2(\tau)$ ) bestimmt und das Tragwerksverhalten zum Zeitpunkt  $\tau_{v_1} = 27,5\text{ms}$  untersucht. Die repräsentativen Tragwerksprozesse und das zugehörige Systemverhalten zum Zeitpunkt  $\tau_{v_1}$  sind in Abb. 2 visualisiert. Die Kenntnis von Verzweigungspunkten erlaubt eine detailliertere Interpretation des Strukturverhaltens

Tabelle 1: Stochastische Tragwerksparameter; Werte normiert

Zufallsgröße $X_i$	Beschreibung	Parameter	Mittelwert $\mu_i$	Standardabweichung $\sigma_i$
$X_1$	Blechdicke	T 1139	10,0	0,20
$X_2$	Blechdicke	T 1134	10,0	0,21
$X_3$	Blechdicke	T 1210	10,0	0,62
$X_4$	Blechdicke	T 1221	10,0	0,50
$X_5$	Fließspannung	SF 1134	10,0	0,50

im Zeitbereich. Insbesondere können durch Auswertung der Berechnungsergebnisse zum detektierten Zeitpunkt  $\tau_{v_1}$  Ursachen für die Diversifikation der Tragwerksantworten erkannt werden. Das unterschiedliche Beulverhalten ist hervorgehobenen Bereich (Abb. 2) klar erkennbar und eine mechanische Ursache für die Bildung untereinander heterogener Prozeßgruppen  ${}_0P_1(\tau)$  und  ${}_0P_2(\tau)$ . Aus diesem Beispiel kann geschlußfolgert werden, daß insbesondere bei einer nichtlinearen Analyse kein homogener (ähnlicher) Verlauf der Realisierungen  $\underline{z}(\tau) \in M$  von  $\underline{Z}(\tau)$  über den gesamten betrachteten Zeitbereich erwartet werden kann.

Liegen ein hochdimensionaler unscharfer Prozess  $\underline{X}(\tau)$  und umfangreiche, komplexe Simulationsergebnisse  $\underline{Z}(\tau)$  vor, ist eine visuelle Detektion von Verzweigungspunkten  $v_j$  nicht effektiv bzw. nicht möglich. Die Lösung gelingt numerisch. Für die Existenz eines Verzweigungspunktes werden folgende Kriterien formuliert:

1. Im Zeitbereich  $[\tau_{t-1}, \tau_t]$  nimmt die Unschärfe des Tragwerksprozesses stark zu.
2. Im Zeitbereich  $[\tau_{t-1}, \tau_t]$  spaltet sich der Tragwerksprozeß in homogene Prozeßgruppen auf.

Die Zunahme der Unschärfe (Kriterium 1) ist eine notwendige, die Bildung von Prozeßgruppen (Kriterium 2) eine hinreichende Bedingung für die Existenz eines Verzweigungspunktes  $v_j$ . Alle detektierten Verzweigungspunkte  $v_j$  werden zur Menge der Verzweigungspunkte  $\Upsilon = \{v_1, \dots, v_j, \dots, v_{n_v}\}$  zusammengefaßt.

Kriterium 1 kann durch die Auswertung eines Unschärfemaßes  $\mathcal{U}$  numerisch geprüft werden. Das Unschärfemaß  $\mathcal{U}$  ist in Abhängigkeit von der vorliegenden Unschärfecharakteristik zu wählen. Für stochastische Prozesse kann die empirische Varianz  $\text{VAR}(\underline{Z}(\tau))$  angewendet werden. Kriterium 1 ist erfüllt, wenn im Zeitintervall  $[\tau_{t-1}; \tau_t]$

$$\mathcal{U}(\underline{\tilde{Z}}(\tau_{t-1})) \ll \mathcal{U}(\underline{\tilde{Z}}(\tau_t)) \quad (3)$$

gilt. Die Bildung von Prozeßgruppen im Intervall  $[\tau_{t-1}; \tau_t]$  kann durch geeignete numerische Verfahren auf der Basis von Clusterverfahren erfolgen. Initial verlaufen alle ermittelten Realisierungen  $\underline{z}(\tau)$  des stochastischen Tragwerksprozesses homogen. Sie bilden eine Prozeßgruppe (Cluster)  $P_0(\tau)$ . Die in  $P_0(\tau)$  zusammengefaßten Realisierungen  $\underline{z}(\tau)$  werden in  $n_t$  Zeitschritte diskretisiert. Die Menge  $P_0(\tau)$  kann dadurch in Punktteilungen  $P_{0,t}$  zerlegt werden. Jede Teilmenge  $P_{0,t}$  repräsentiert eine diskrete Punktmenge von Prozeßgrößen  $\underline{z}_t = \underline{z}(\tau = \tau_t)$  zum Zeitpunkt  $\tau = \tau_t$  (s. Abb. 3). Zu jedem Zeitschritt  $\tau_t$  wird die zutreffende Clusterung der Punktmenge  $P_{0,t}$  mit Hilfe einer Klassifizierungsfunktion bestimmt

$$g: P_{0,t} \rightarrow \mathcal{C}_t \mid \mathcal{C}_t = \{C_{1,t}, \dots, C_{n_{C,t},t}\} \quad (4)$$

Jede so erhaltene Clusterkonfiguration  $\mathcal{C}_t$  wird durch ein Gütemaß  $G$  bewertet.

$$\mathcal{C}_t \times \mathcal{C}_t \rightarrow G \in \mathbb{R} \quad (5)$$

Anhand der erhaltenen Maßzahl kann beurteilt werden, ob die gefundene Clusterkonfiguration die Punktmenge  $P_{0,t}$  mit  $n_{C,t}$  Clustern zutreffend beschreibt. Als Gütemaß wird der Silhouettenkoeffizient  $SC$  nach Gl. (17) vorgeschlagen. Für dieses Gütemaß kann ein Grenzwert  $SC_{min}$  festgelegt werden, der eindeutig eine Substrukturierung der Punktmenge in gut separierte Cluster indiziert. Als Grenzwert hat sich

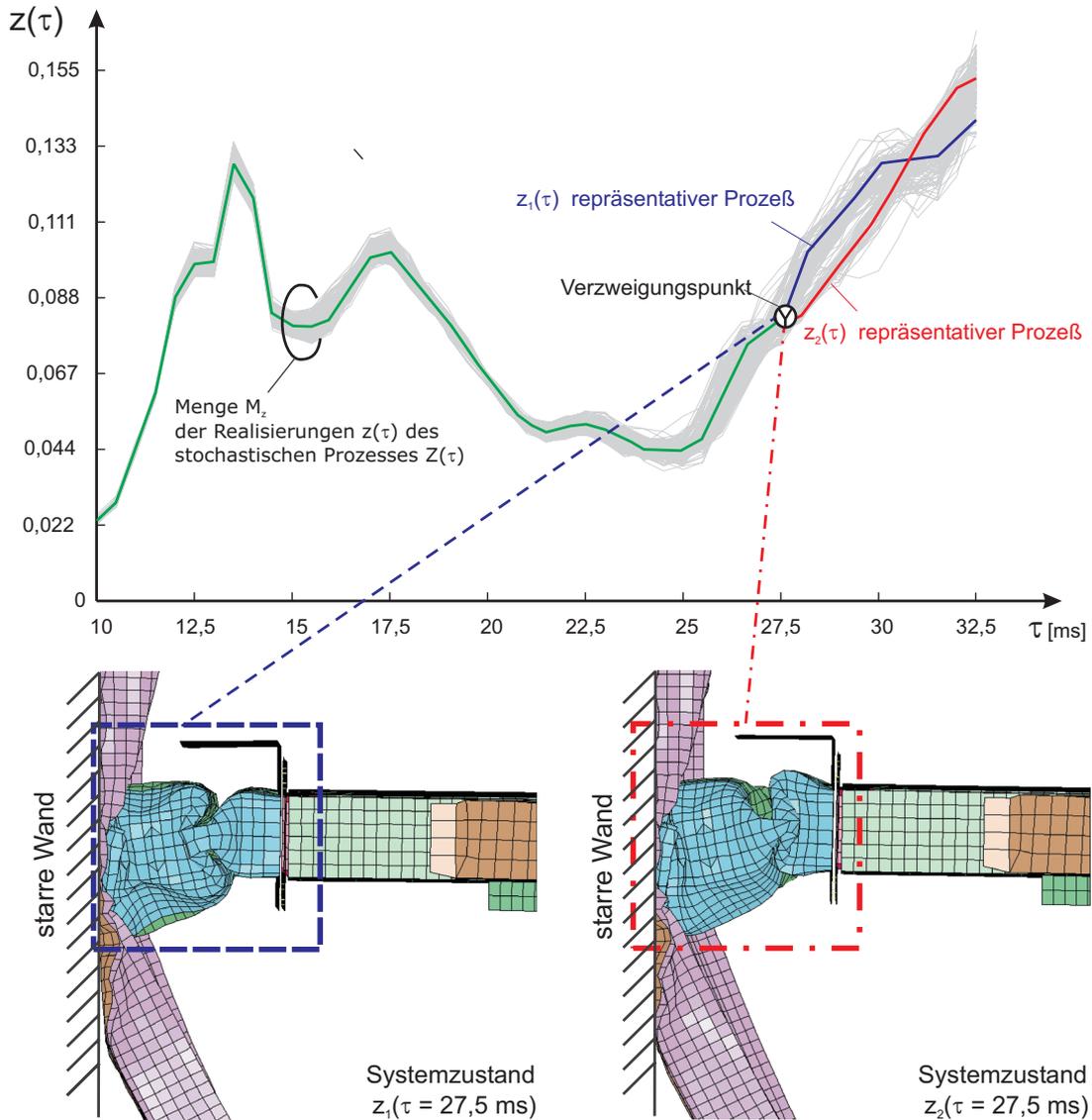


Abbildung 2: Stochastischer Tragwerksprozeß  $Z(\tau)$ ; ausgewählte Prozesse  $z_1(\tau)$ ,  $z_2(\tau)$  und Systemzustände zum Zeitpunkt  $\tau_{v_1} = 27,5 \text{ ms}$

dabei der Wert  $SC_{min} = 0,75$  bewährt (vgl. Tab. 3). Wird dieser Grenzwert  $SC_{min}$  überschritten, so wird ein Verzweigungspunkt  $v_j$  zum Zeitpunkt  $\tau_{i-1}$  detektiert. Entsprechend der ermittelten Clusterkonfiguration  $\mathcal{C}_i$  wird die Prozeßgruppe  $P_0(\tau)$  zum Zeitpunkt  $\tau_i$  in Unterprozeßgruppen  ${}_0P_1(\tau), \dots, {}_0P_k(\tau), \dots, {}_0P_{n_{C,t}}(\tau)$  unterteilt.

$$P_0(\tau) \rightarrow {}_0P_1(\tau), \dots, {}_0P_k(\tau), \dots, {}_0P_{n_{C,t}}(\tau) \quad (6)$$

Es gilt:

$$P_0(\tau) = \bigcup_{k=1}^{n_{C,t}} {}_0P_k(\tau) \quad (7)$$

$$C_k = {}_0P_{k,t} \quad \forall k = 1, \dots, n_{C,t}. \quad (8)$$

Jede so erhaltene Prozeßgruppe  ${}_0P_k(\tau)$  wird separat auf Verzweigungspunkte untersucht. Wird zu einem späteren Zeitpunkt  $\tau_q$  ein Verzweigungspunkt detektiert, erfolgt wiederum eine Substrukturierung

in Unterprozeßgruppen. Für diese gefundenen  $n_{C,q}$  Unterprozeßgruppen wird folgende Schreibweise vereinbart:

$${}_{0-k}P_l(\tau) \mid l = 1 \dots n_{C,q} \quad (9)$$

mit

$${}_0P_k(\tau) = \bigcup_{l=1}^{n_{C,q}} {}_{0-k}P_l(\tau). \quad (10)$$

Dieses rekursive Vorgehen ist in Algorithmus 1 beschrieben. Für einen unscharfen Prozeß  $Z(\tau)$  ist die

---

### Algorithm 1 Algorithmus zur Identifikation der Verzweigungspunkte

---

#### 1. Initialisierung

- Prozeßgruppe:  $\dots P_k(\tau)$
- Startzeitschritt:  $\tau_a$
- Grenzwert des Silhouettenkoeffizienten  $SC_{min}$

#### 2. Iteration über $\tau_t \mid \tau_t \in [\tau_a; \tau_{n_t}]$ , solange gilt: $SC(\mathcal{C}_t) \leq SC_{min}$

- a) Übergang zum nächsten Zeitschritt

$$t = t + 1$$

- b) Bilden der zu klassifizierenden Punktmenge

$$\dots P_{k,t} = \dots P_k(\tau_t)$$

- c) Ermitteln der zutreffenden Clusterung zum Zeitpunkt  $\tau_t$

$$\dots P_{k,t} \rightarrow \mathcal{C}_t \mid \mathcal{C}_t = \{C_{1,t}, \dots, C_{n_{C,t}}\}$$

- d) Bestimmen des Gütemaßes Silhouettenkoeffizient  $SC$

- i.  $SC(\mathcal{C}_t) > SC_{min}$  – Bildung von Teilprozeßgruppen detektiert; Zeitpunkt  $\tau_{t-1}$  ist ein Verzweigungspunkt
- ii. Unterteilen der Prozeßgruppe  $\dots P_k(\tau)$  in  $n_{C,t}$  Teilprozeßgruppen

$$\dots P_k(\tau) \rightarrow \dots {}_{-k}P_1(\tau), \dots, \dots {}_{-k}P_{n_{C,t}}(\tau)$$

auf der Grundlage der ermittelten Clusterkonfiguration  $\mathcal{C}_t$

- iii. Wiederholtes Ausführen des Algorithmus 1 für jede Teilprozeßgruppe  $\dots {}_{-k}P_1(\tau), \dots, \dots {}_{-k}P_{n_{C,t}}(\tau)$

---

Detektion der zwei vorhandenen Verzweigungspunkte  $\Upsilon = \{v_1, v_2\}$  in Abb. 3 dargestellt. Die entstandene hierarchische Strukturierung des unscharfen Prozesses ist erkennbar. Bei stark verrauschten unscharfen Tragwerksprozessen kann es erforderlich sein, die Sensitivität dieses Algorithmus zur Detektion von Verzweigungspunkten zu justieren. Der Zeitpunkt  $\tau_{t-1}$  ist genau dann ein Verzweigungspunkt, wenn gilt

$$SC(C_i) > SC_{min} \mid t = t, t + 1, \dots, s_\tau. \quad (11)$$

Nur wenn zu allen  $s_\tau$  folgenden Zeitpunkten eine zutreffende Einteilung der Mengen  $\dots P_{k,t}$  möglich ist, existiert zum Zeitpunkt  $\tau_{t-1}$  ein Verzweigungspunkt.

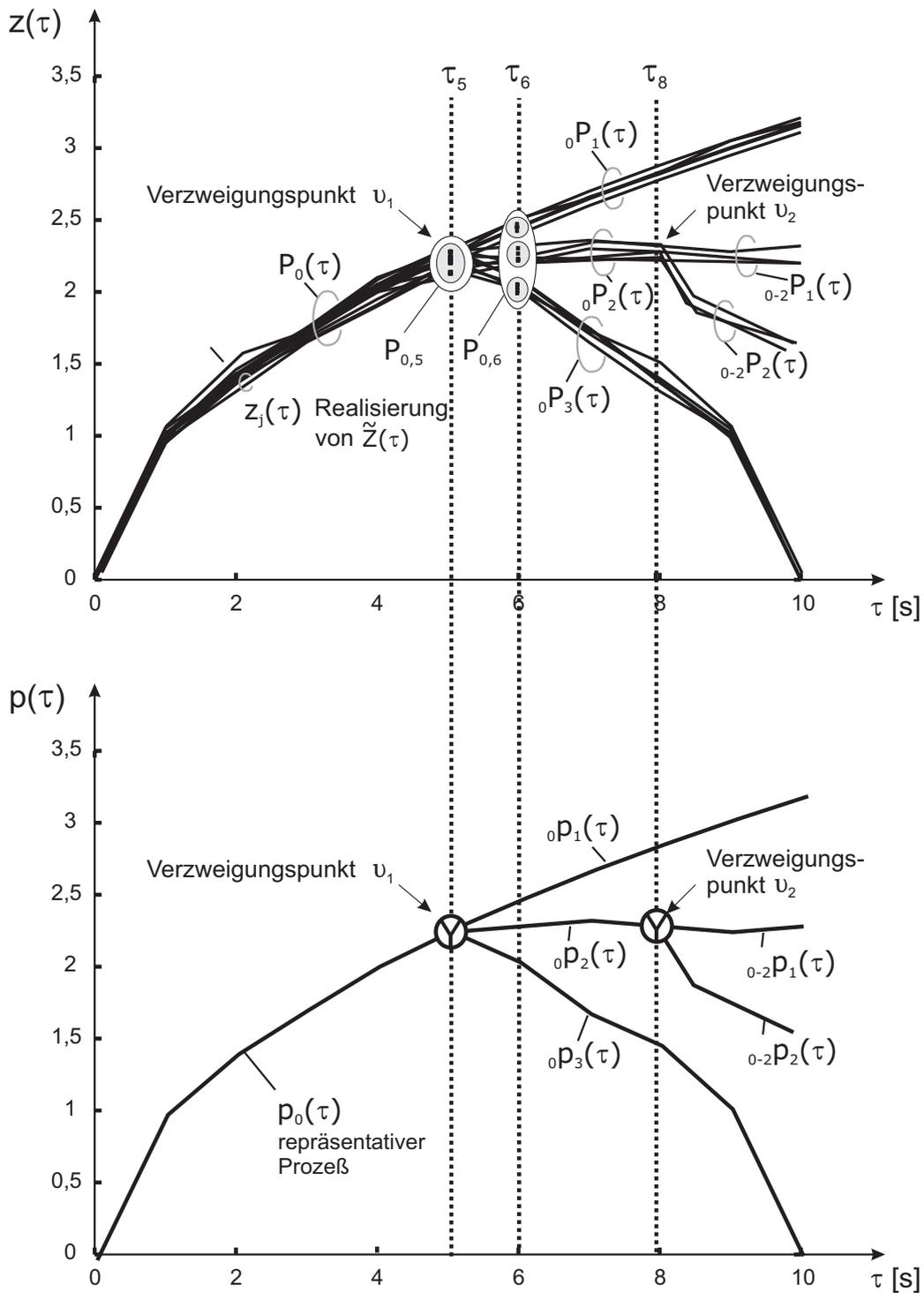


Abbildung 3: Unscharfer Prozeß  $Z(\tau)$ ; detektierte Verzweigungspunkte  $v_1$  und  $v_2$ ; repräsentative Prozesse  $\dots p_{\dots}(\tau)$

## 2 Clusteranalyse

Die Clusteranalyse ist ein Verfahren, mit dem in Datensätzen Strukturen gefunden werden können. Hauptanwendungsgebiete sind die Mustererkennung [4, 5], die Datenanalyse und die Informationsreduktion [1, 2]. Durch eine Clusteranalyse wird eine Menge von Elementen in homogene Teilmengen unterteilt. Üblicherweise werden Elemente auch als Daten und homogene Teilmengen als Cluster oder Klassen bezeichnet. Die Klassifikation der Elemente basiert auf zwei grundlegenden Forderungen:

- Die Elemente eines Clusters sollen untereinander möglichst ähnlich sein, d. h., es wird Homogenität innerhalb eines Clusters gefordert.
- Die Elemente verschiedener Cluster sollen möglichst unterschiedlich sein, d. h., es wird Heterogenität zwischen den Clustern gefordert.

*Homogenität* bezeichnet die Ähnlichkeit der einem Cluster  $C_k$  zugeordneten Punkte  $\underline{z} \in C_k \mid \underline{z} = \underline{z}_i = \underline{z}(\tau = \tau_i)$ . Ein Cluster  $C_k$  ist homogen, wenn alle Punkte  $\underline{z} \in C_k$  einander ähnlich sind. Ähnlichkeit kann mit Hilfe der Punktabstände bewertet werden. Die Ähnlichkeit nimmt zu, wenn die Summe der Abstände zwischen Punkten eines Clusters minimal wird.

$$\sum_{\underline{z}_i, \underline{z}_j \in C_k} d(\underline{z}_i, \underline{z}_j) \Rightarrow \text{Min} \forall C_k, k = 1, \dots, n_C \quad (12)$$

*Heterogenität* bezeichnet die Verschiedenheit der Cluster. Die Cluster  $C_k$  und  $C_l$  sind heterogen, wenn ihre Punkte  $\underline{z}_i \in C_k$  und  $\underline{z}_j \in C_l$  verschieden sind. Die Verschiedenheit nimmt zu, wenn die Summe der Abstände zwischen den Punkten maximal wird.

$$\sum_{\underline{z}_i \in C_k, \underline{z}_j \in C_l} d(\underline{z}_i, \underline{z}_j) \Rightarrow \text{Max} \mid k \neq l \quad (13)$$

In Abb. 4 ist ein Beispieldatensatz dargestellt; es sind drei Cluster erkennbar. Das Ziel der Clusteranalyse ist es, diese Cluster algorithmenbasiert zu detektieren. Die Zuordnung der Elemente zu einem Cluster erfolgt durch eine Klassifizierungsfunktion  $g$ , Gl. (4). Die Klassifizierungsfunktion zerlegt die Grundmenge der Elemente in Teilmengen  $C_k \in \mathcal{C}$ . Die Clusteranalyse liefert ein (lokales) Optimum der Klassifikation. Die verschiedenen Clusterverfahren führen aufgrund unterschiedlicher Konzeptionen i. d. R. auch zu unterschiedlichen Klassifikationsergebnissen. Das k-Medoid Clusterverfahren, ein deterministisches Clusterverfahren, ist in Abschn. 2.1 vorgestellt.

Für die Klassifikation der Elemente muß die Anzahl  $n_C$  der betrachteten Klassen a priori bekannt sein. Mit dem Klassifikator  $g$  wird eine optimale Zuordnung der Elemente  $\underline{z}$  zu einem Cluster oder zu mehreren Clustern  $C_1, C_2, \dots, C_{n_C} \in \mathcal{C}$  bestimmt. Die Anzahl der Cluster ist jedoch in den wenigsten Anwendungsfällen bekannt. Sie kann für eine optimale Klassifikation der Elemente mit Gütemaßen bestimmt werden. Dafür sind mehrere Klassifikationen mit unterschiedlicher Clusteranzahl  $n_C \in \{n_C^1, n_C^2, \dots, n_C^{\max}\}$  zu untersuchen. Die Obergrenze  $n_C^{\max}$  ist zu schätzen.

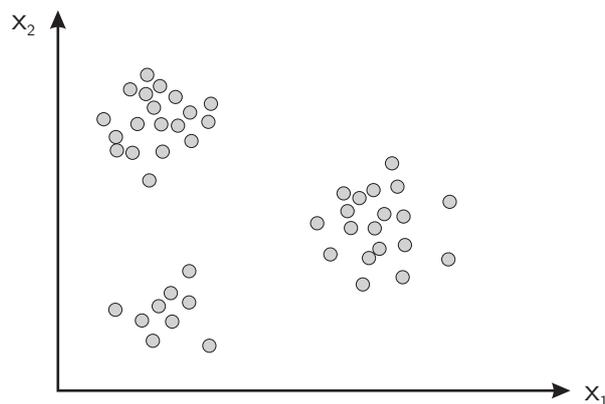


Abbildung 4: Beispieldatensatz

## 2.1 k-Medoid-Clusterverfahren

Das k-Medoid-Clusterverfahren [6] gehört zu den deterministischen, partitionierenden Verfahren. Bei diesem Verfahren wird die Klassifikation realisiert, indem zuerst  $n_C$  repräsentative Elemente  $r_1, \dots, r_k, \dots, r_{n_C} \in R$  aus der Gesamtheit der Elemente (Realisierungen)  $\underline{z}$  des Datensatzes  $M$  ausgewählt werden.

Anschließend werden diesen repräsentativen Elementen nach dem Kriterium größtmöglicher Ähnlichkeit schrittweise weitere Elemente zugeordnet. Jedes repräsentative Element  $r_k$  und die ihm zugeordneten Elemente  $\underline{z}$  bilden den Cluster  $C_k$ . Alle Cluster  $C_1, \dots, C_k, \dots, C_{n_C}$  bilden die Clusterkonfiguration  $\mathcal{C}$ .

Mit der Zielfunktion des k-Medoid-Clusterverfahrens wird der durchschnittliche Abstand von Elementen  $\underline{z}$  zu ihrem repräsentativen Element  $\underline{z}_k$  für alle  $r_k \in R$  minimiert.

$$\sum_{k=1}^{n_C} \sum_{\underline{z} \in C_k} d(r_k, \underline{z}) \Rightarrow \text{Min} \quad (14)$$

Im Gegensatz zu den meisten deterministischen Clusterverfahren wird auf eine Quadrierung der Abstände verzichtet.

Der Algorithmus des k-Medoid-Clusterverfahrens ist in zwei Stufen untergliedert. In der ersten Stufe werden für eine vorgegebene Clusteranzahl  $n_C$  repräsentative Elemente  $r_1, \dots, r_k, \dots, r_{n_C} \in R$  festgelegt, deren Auswahl in der zweiten Stufe verbessert wird.

Stufe 1 beginnt mit der Festlegung des zentralen Elementes  $r_k \in R$  als erstem repräsentativem Element. Das zentrale Element wird durch Minimierung der Summe der Abstände  $d(r_k, \underline{z}_i)$  zu allen anderen Elementen  $\underline{z}_i$  gefunden

$$\sum_{\underline{z}_i \in M} d(r_k, \underline{z}_i) \Rightarrow \text{Min} \quad (15)$$

und der Menge der repräsentativen Elemente  $R$  hinzugefügt. Die weiteren repräsentativen Elemente  $r_1, \dots, r_k, \dots, r_{n_C} \in R$  werden nach Algorithmus 2 ermittelt.

---

#### Algorithm 2 Stufe 1 des k-Medoid-Clusterverfahrens

---

0. Auswahl eines Elements  $r_k$  aus allen Elementen  $\underline{z}_i \in M$  mit

$$\sum_{\underline{z}_i \in M} d(r_k, \underline{z}_i) \Rightarrow \text{Min.}$$

Die weiteren repräsentativen Elemente  $r_2, \dots, r_{n_C}$  werden nacheinander durch folgendes iteratives Vorgehen ermittelt:

1. Auswahl eines Elements  $\underline{z}_i \in M$
2. Auswahl eines Elements  $\underline{z}_j \notin \{R \cup \{\underline{z}_i\}\}$  und Ermitteln des kleinsten Abstandes zu den repräsentativen Elementen  $r_k \in R$

$$f_j = \min_{r_k \in R} d(\underline{z}_j, r_k)$$

3. Bestimmen des Maßes

$$h_{j,i} = \max \{f_j - d(\underline{z}_i, \underline{z}_j), 0\}$$

4. Berechnen der Zielfunktion für die Auswahl  $\underline{z}_i$  und aller verbleibenden  $\underline{z}_j$

$$e_{j,i} = \sum_{\underline{z}_j \in M \setminus \{R \cup \{\underline{z}_i\}\}} h_{j,i}$$

5. Auswahl des Elements  $\underline{z}_i \notin R$ , das die Zielfunktion

$$\max_{\underline{z}_i \in M \setminus R} e_{j,i}$$

maximiert, und Zuordnung des ausgewählten Elements zur Menge der repräsentativen Elemente  $R$

---

In Stufe 2 wird geprüft, ob eine alternative Auswahl repräsentativer Elemente  $r_k$  ein besseres Clusterergebnis liefert. Dazu werden Elementpaare  $r_k, z_i$  ausgewählt. Verbessert sich das Clusterergebnis bei der Auswahl des Elements  $z_i$  als repräsentatives Element anstelle des Elements  $r_k$ , werden die beiden Elemente getauscht. Die Vorgehensweise ist in Algorithmus 3 wiedergegeben.

Alle potentiell möglichen Elementtausche werden berücksichtigt. Damit ist das k-Medoid-Clusterverfahren im Gegensatz zum k-Means-Clusterverfahren [1] nicht von der Reihenfolge der Eingangsdaten abhängig.

---

### Algorithm 3 Stufe 2 des k-Medoid-Clusterverfahrens

---

1. Auswahl eines repräsentativen Elements  $r_k \in R$  und eines nichtrepräsentativen Elements  $z_i$

$$r_k \in R \wedge z_i \in M \setminus R$$

2. Auswahl eines Elements  $z_j \notin \{R \cup \{z_i\}\}$

Berechnen des Maßes  $h_{k,i,j}$  mit der Fallunterscheidung

- a) Element  $z_j$  ist von  $r_k$  und  $z_i$  weiter entfernt als von einem repräsentativen Element  $r \in R$

$$\exists r \in R \setminus \{r_k\} : d(z_j, r) < \min \{d(z_j, r_k), d(z_j, z_i)\} \Rightarrow h_{k,i,j} = 0$$

- b) Element  $z_j$  ist weiter entfernt von  $z_k$  als von mindestens einem anderen repräsentativen Element  $r \in R$ , aber dem Element  $z_i$  näher als jedem anderen Repräsentanten

$$\min_{r \in R \setminus \{r_k\}} d(z_j, r) < d(z_j, r_k) \wedge d(z_j, z_i) < \min_{r \in R \setminus \{r_k\}} d(z_j, r)$$

$$\Rightarrow h_{k,i,j} = d(z_j, z_i) - d(z_j, r_k)$$

- c) Wenn die Bedingungen 2.a) und 2.b) nicht erfüllt sind, dann gilt

$$d_{j,r} = \min_{r \in R \setminus \{r_k\}} d(z_j, r)$$

$$E_j = \min_{r \in R \setminus \{r_k, r\}} d(z_j, r)$$

$$\text{i. } d(z_j, z_i) < E_j \Rightarrow h_{k,i,j} = d(z_j, z_i) - d(z_j, r_k)$$

$$\text{ii. } d(z_j, z_i) \geq E_j \Rightarrow h_{k,i,j} = E_j - d(z_j, r_k)$$

3. Berechnung von

$$T_{r_k, z_i} = \sum_{z_j \in M \setminus \{R, \{z_i\}\}} h_{k,i,j}$$

4. Bestimmen des Elementpaares  $r_k, z_i$ , das  $T_{r_k, z_i}$  minimiert

5. Austausch der Elemente  $r_k$  und  $z_i$ , falls das Minimum einen negativen Wert annimmt

$$R = (R \cup \{z_i\}) \setminus \{r_k\}$$

Diese Iteration wird so lange fortgesetzt, bis kein Austausch mehr erfolgt.

---

### 3 Unüberwachte Klassifikation

Mit der Clusteranalyse wird für eine vorgegebene Anzahl von Clustern  $n_C$  jeweils eine "beste" Klassifikation ermittelt, d. h. eine Zuordnung von Elementen zu den Clustern  $C_1, \dots, C_{n_C} \in \mathcal{C}$ . Die Anzahl der Cluster  $n_C$ , die einen Datensatz zutreffend beschreibt, ist allerdings nicht bekannt. Es existieren unterschiedliche Vorgehensweisen, die die Ermittlung einer zutreffenden Clusteranzahl unterstützen. Ein etablierter Ansatz zur Ermittlung der zutreffenden Clusteranzahl besteht darin, die Klassifikation des Datensatzes

bei unterschiedlicher Clusteranzahl  $n_C$  durchzuführen. Die für jede gewählte Clusteranzahl  $n_C$  erhaltenen Ergebnisse werden anschließend bewertet und die "optimale" Clusteranzahl  $n_{C,opt}$  ausgewählt. Die Bewertung der Klassifikationsergebnisse erfolgt mit Hilfe von (globalen) Gütemaßen. Gütemaße sind das Ergebnis der Abbildung der Clusterkonfiguration  $\mathcal{C}$  auf einen reellen Wert.

$$\mathcal{C} \rightarrow G \in \mathbb{R} \quad (16)$$

Die Klassifikation unter Verwendung von Clusterverfahren und Gütemaßen ohne Nutzerinteraktion wird auch als unüberwachte Klassifikation bezeichnet. Nachfolgend sei exemplarisch das Gütemaß Silhouettenkoeffizient für deterministische Clusterverfahren [6] wiedergegeben.

**Silhouettenkoeffizient.** Der Silhouettenkoeffizient  $SC$  bewertet die Homogenität und Heterogenität einer Clusterkonfiguration  $\mathcal{C}$  und ist damit ein sehr leistungsfähiges Gütemaß. Der Silhouettenkoeffizient ist definiert als

$$SC = \frac{1}{n_C} \sum_{k=1}^{n_C} \frac{1}{|C_k|} \sum_{i=1}^{|C_k|} \frac{b_i - a_i}{\max[a_i, b_i]} \quad (17)$$

mit

$$a_i = \frac{1}{|C_k|} \sum_{h=1}^{|C_k|} d(\underline{x}_i, \underline{x}_h), \quad (18)$$

$$b_i = \min_{h=1, \dots, m, h \neq i} [a_h]. \quad (19)$$

Nähert sich das Gütemaß  $SC$  der oberen Schranke von 1, so existieren eindeutige Strukturen, d. h. es wurde eine gute Clusterung erreicht. In Tab. 3 (nach [6]) ist eine mögliche Interpretation für unterschiedliche  $SC$ -Werte angegeben. Der Mindestwert des Silhouettenkoeffizienten  $SC_{min}$  ist problemspezifisch.

Tabelle 3: Bewertung des Silhouettenkoeffizienten [6]

$SC$	Die Klassifikation $\mathcal{C}$ beschreibt die Struktur des Datensatzes ...
0,71 – 1,00	sehr gut
0,51 – 0,70	gut
0,26 – 0,50	schlecht; die Klassifikation ist möglicherweise künstlich erzwungen
$\leq 0,25$	keine zutreffende Klassifikation

#### 4 Identifikation von Verzweigungspunkten

Der in Abschn. 1 entwickelte Algorithmus zur Detektion von Verzweigungspunkten wird an einem stochastischen Prozeß  $Z(\tau)$  erprobt. Der Prozeß sei durch die Abbildung

$$\mathbf{f}: \underline{X}(\tau) \rightarrow Z(\tau) \quad (20)$$

mit  $\underline{X}(\tau) = \{X_1(\tau), \dots, X_5(\tau)\}$  beschrieben. Die Abbildung  $\mathbf{f}$  ist durch eine Menge von Funktionen

$$\begin{aligned}
 \mathbf{f}(\tau) = & \left\{ \sqrt{\tau} + f(x_1, \tau); \right. \\
 & 1,5 \cdot \ln(\tau) + \left(\frac{\tau}{55}\right) + f(x_2, \tau); \\
 & 2,0 \cdot \left(\sin\left(\frac{\pi \cdot \tau}{60}\right) + \left(\frac{\tau}{30}\right)\right) + 1,5 + f(x_3, \tau); \\
 & 5,0 \cdot \left(\sin\left(\frac{\pi \cdot \tau}{100}\right) - \left(\frac{\tau}{500}\right)\right) + 1,5 + f(x_4, \tau); \\
 & \left. \left(\frac{60}{\tau+30}\right) \cdot \ln\left(\frac{\tau+2}{2}\right) + f(x_5, \tau) \right\}
 \end{aligned} \tag{21}$$

mit

$$f(x_i, \tau) = \begin{cases} \tau < 10 & \frac{\tau}{10} \cdot (1 - x_i) \\ \tau \geq 10 & \frac{(\tau+90)}{100} \cdot (1 - 2x_i) \end{cases} \tag{22}$$

festgelegt. Die Realisierungen  $x_1, \dots, x_5$  sind im Intervall  $[0; 1]$  gleichverteilt. Die ersten einhundert Realisierungen  $z(\tau)$  des stochastischen Prozesses  $\underline{Z}(\tau)$  sind in Abb. 5 dargestellt.

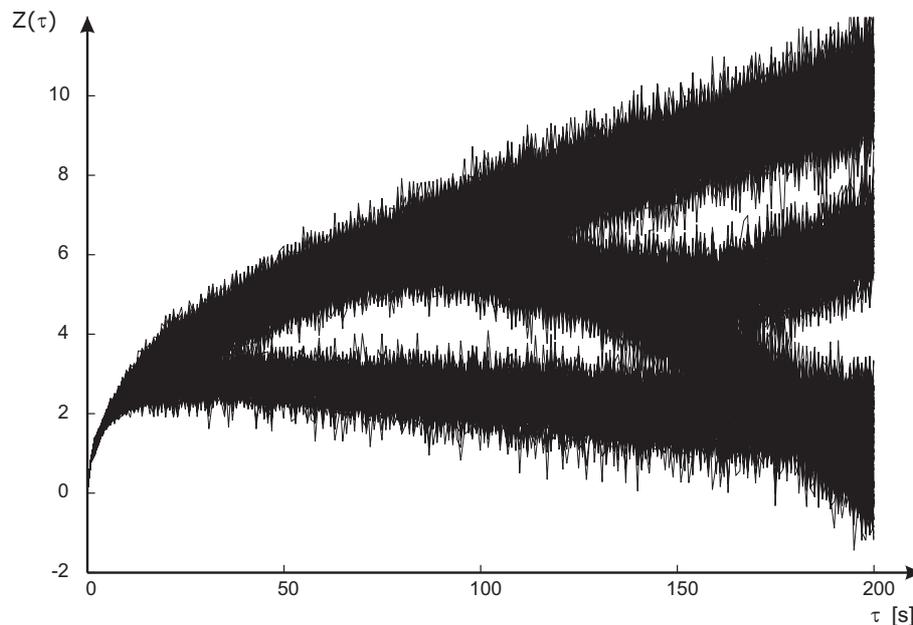


Abbildung 5: Realisierungen  $z(\tau)$  des stochastischen Prozesses  $\underline{Z}(\tau)$

Für den Zeitpunkt  $\tau = 200s$  kann aus den Realisierungen das in Abb. 6 dargestellte Histogramm entwickelt werden.

Für die Detektion von Verzweigungspunkten wird als Mindestwert des Silhouettenkoeffizienten  $SC_{min} = 0,75$  gewählt, s. Tab. 3. Der Parameter  $s_\tau$  wird zu  $s_\tau = 4$  festgelegt, um den Einfluß lokaler Streueffekte zu reduzieren. Dadurch wird die Sensitivität des Algorithmus herabgesetzt. Der Silhouettenkoeffizient  $SC$  muß nun zu vier aufeinanderfolgenden Zeitpunkten  $\tau_{t+1}, \dots, \tau_{t+4}$  den Wert  $SC_{min}$  überschreiten, damit ein Verzweigungspunkt zum Zeitpunkt  $\tau_{t-1}$  detektiert wird. In Abb. 7 ist das Ergebnis der Berechnung mit  $s_\tau = 4$  dargestellt. Es werden jetzt die erwarteten drei Verzweigungspunkte  $\tau_{v1} = 38s$ ,  $\tau_{v2} = 121s$  und  $\tau_{v3} = 161s$  erkannt. Die repräsentativen Tragwerksprozesse sind in der Abb. 7 gekennzeichnet.

Den einzelnen Prozeßgruppen können nun Histogramme der Realisierungen des stochastischen Prozesses zugeordnet werden. Zum Zeitpunkt  $\tau = 200s$  spaltet sich das Histogramm der Abb. 6 in die in Abb. 8 dargestellten Teilhistogramme auf. Jedes Teilhistogramm ist einer Prozeßgruppe zugeordnet.

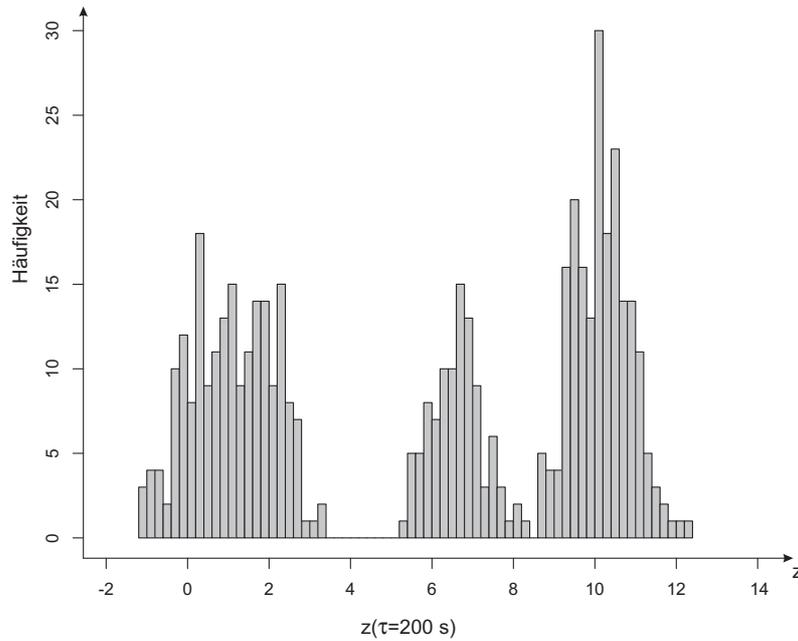


Abbildung 6: Histogramm der Realisierungen des stochastischen Prozesses  $\underline{z}(\tau)$  zum Zeitpunkt  $\tau = 200s$

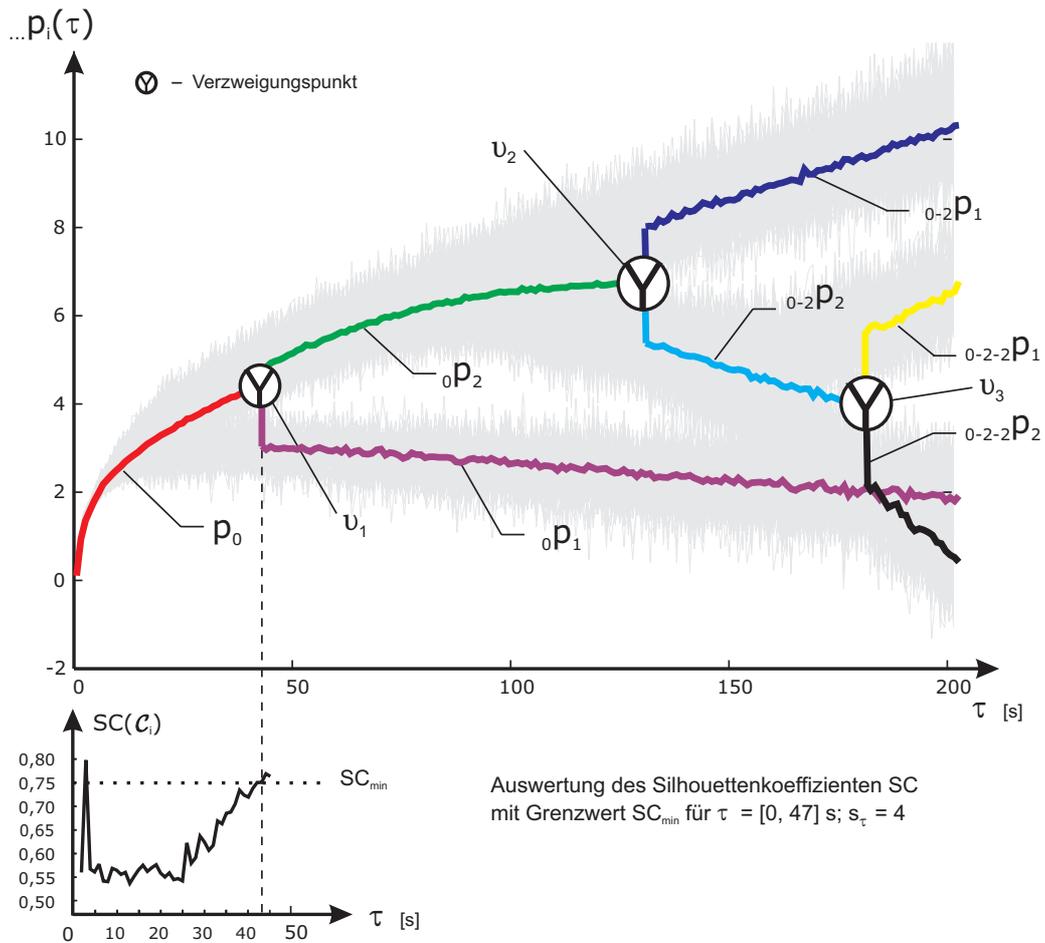


Abbildung 7: Detektierte Verzweigungspunkte  $\Upsilon = \{v_1, v_2, v_3\}$ ;  $SC_{min} = 0,75$  und  $s_\tau = 4$

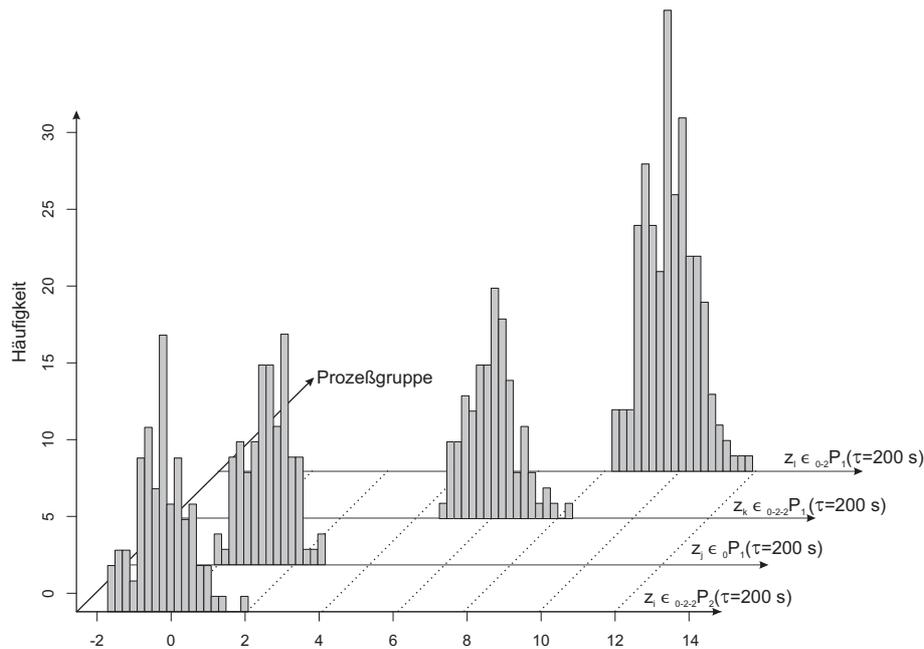


Abbildung 8: Histogramme der Prozeßgruppen ... $P_{...}(\tau)$  zum Zeitpunkt  $\tau = 200\text{s}$

## 5 Zusammenfassung

Verzweigungspunkte kennzeichnen Zeitpunkte innerhalb eines unscharfen Tragwerksprozesses zu denen sich Tragwerksantworten stark diversifizieren. Die Kenntnis von Verzweigungspunkten kann die Analyse eines unscharfen Tragwerksprozesses unterstützen und ermöglicht u.a. die Identifikation mechanischer Ursachen. Nur in Sonderfällen gelingt die visuelle Identifikation von Verzweigungspunkten. Für mehrdimensionale, komplexe Tragwerksprozesse ist die Detektion nur numerisch möglich. Der in diesem Beitrag vorgestellte Algorithmus ermöglicht die (unüberwachte) Detektion von Verzweigungspunkten mit Hilfe von Clusterverfahren und Gütemaßen. Das Verfahren wird anhand eines künstlichen stochastischen Prozesses demonstriert.

## Literatur

- [1] Anderberg, M. R.: *Cluster analysis for applications*. Academic Press, New York, 1973.
- [2] Bacher, J.: *Clusteranalyse: Anwendungsorientierte Einführung*. Oldenbourg, München, 1996.
- [3] Beer, M. and M. Liebscher: *Designing robust structures – a nonlinear simulation based approach*. Special Issue of Computers & Structures, 2007. (in press, available online).
- [4] Bezdek, J. C. and S. K. Pal (eds.): *Fuzzy models for pattern recognition – methods that search for structures in data*. IEEE Press, Piscataway, 1992.
- [5] Höppner, F., F. Klawonn, R. Kruse, and T. Runkler: *Fuzzy Cluster Analysis: Methods for Classification, Data Analysis and Image Recognition*. Wiley, Chinchester New York Weinheim, 1999.
- [6] Kaufman, L. and P. Rousseeuw: *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley, Chinchester New York Weinheim, 1990.
- [7] Thiele, M., M. Liebscher, and W. Graf: *Fuzzy analysis as alternative to stochastic methods – a comparison by means of a crash analysis*. In *Proceedings of the 4th German LS-DYNA Forum*, pp. D–I–45–63, Bamberg, 2005.

