

Detection of branching points in noisy processes

Michael Beer · Martin Liebscher

Received: 7 January 2009 / Accepted: 3 November 2009
© Springer-Verlag 2009

Abstract Processes in engineering mechanics often contain branching points at which the system can follow different physical paths. In this paper a method for the detection of these branching points is proposed for processes that are affected by noise. It is assumed that a bundle of process records are available from numerical simulations or from experiments, and branching points are concealed by the noise of the process. The bundle of process records is then evaluated at a series of discrete values of the independent process coordinates. At each discrete point of the process, the associated point set of process values is investigated with the aid of cluster analysis. The detected branching points are verified with a recursive algorithm. The revealed information about the branching points can be used to identify the physical and mechanical background for the branching. This helps to better understand a mechanical system and to design it optimal for a specific purpose. The proposed method is demonstrated by means of both a numerical example and a practical example of a crashworthiness investigation.

Keywords Processes · Noise · Time series · Data analysis · Branching points · Bifurcation · Cluster analysis · Process classification

List of symbols

b	Branching point
C	Cluster (point set)
\tilde{C}	Fuzzy cluster (discrete fuzzy set)
$ C $	Cardinality of cluster C
$d(\underline{z}_1, \underline{z}_2)$	Distance between \underline{z}_1 and \underline{z}_2
\mathbb{D}	Real-valued domain of process coordinates
\mathbf{f}, \rightarrow	Mapping
G	Quality measure for cluster configuration
L_U, L_{SC}	Length of record for check of uncertainty and silhouette coefficient, respectively
M	Set
$O(.)$	Set of outliers
$p(.)$	Representative process record
$P(.)$	Process group
SC	Silhouette coefficient
U	Uncertainty
v	Velocity
$\underline{x}(.)$	Mechanical input quantities
$\underline{z}(.)$	Process variables
X, Z	Random variables
$X \sim (.)$	Specification of the distribution of X according to $(.)$
\mathbb{Z}	Real-valued domain of process variables
$ $	For which the following holds
ε	Noise
$\underline{\theta}$	Process coordinates
μ	Membership value according to fuzzy set theory
τ	Single process coordinate
(a, b, c, \dots)	Vector with elements a, b, c, \dots
(a, b)	Open interval with limits a and b
(z, μ)	Value pair with elements z and μ
$[a, b]$	Closed interval with bounds a and b
$\{a, b, c\}$	Set with elements a, b, c

M. Beer (✉)
Department of Civil Engineering, National University of Singapore,
Singapore, Singapore
e-mail: cvebm@nus.edu.sg

M. Liebscher
Dynamore GmbH, Stuttgart, Germany
e-mail: martin.liebscher@dynamore.de

1 Introduction

Processes appear and must be analyzed in many engineering disciplines and industrial applications. The goal of these analyzes is the engineering interpretation of the mathematical characteristics of the processes. In a general case, a process may be affected by randomness of input quantities, environmental and boundary conditions, system properties, and numerical operations (or experiments). A modeling as random process $\underline{Z}(\theta)$ according to [30] is, thus, appropriate. In this context the process is an indexed family of random variables \underline{Z} defined over a domain \mathbb{D} with $\theta \in \mathbb{D}$ being the real-valued parameters or coordinates of the process. The value of the process is described—at each θ —in terms of the variable z , which is, thus, referred to as the process variable. Each single process realization $\underline{z}(\theta)$ —as an outcome of a single analysis or measurement—is, then, a deterministic process record (state of the process). It represents one realization of the family of random variables \underline{Z} for all θ in form of a function of the process variable z in dependence on the process coordinates θ .

Physically, processes can be of various types; they may describe the behavior of any physical quantities in dependence on any physical coordinates such as time, humidity, temperature, or spatial location. Herein, our focus is on those processes in engineering mechanics, in which the process variables z represent mechanical quantities, such as displacements v , forces F or energy E , which result from a numerical analysis or experiment. The results from these analyses or experiments are usually calculated or recorded in dependence on time τ and/or spatial coordinates ξ_1, ξ_2, ξ_3 , which, thus, represent the process coordinates θ in the subsequent discussion. In this set-up the process $\underline{Z}(\theta)$ may describe, for instance, the behavior of a mechanical system in dependence on time. And the analysis or the experiment to generate the process $\underline{Z}(\theta)$ may be understood as a mapping

$$\mathbf{f} : \underline{X}(\theta) \rightarrow \underline{Z}(\theta), \quad (1)$$

with $\underline{X}(\theta)$ summarizing mechanical input quantities such as structural parameters, loads, environmental conditions, and boundary conditions. The quantities contained in $\underline{X}(\theta)$ may be deterministic or random and may or may not depend on the process coordinates θ .

An important issue in the analysis of those processes is the detection of branching points, at which the mechanical system can follow different physical paths. The identification of these paths and the associated physical and mechanical background is essential to understand a mechanical system and to design it for a specific purpose. Branching points can be associated with mechanical effects such as local buckling in shell structures [8, 14, 24], multiple failure kinematics in structural collapse after blasting [13, 21], initialization of sub- or super-harmonic vibrations [17] or even chaotic behavior

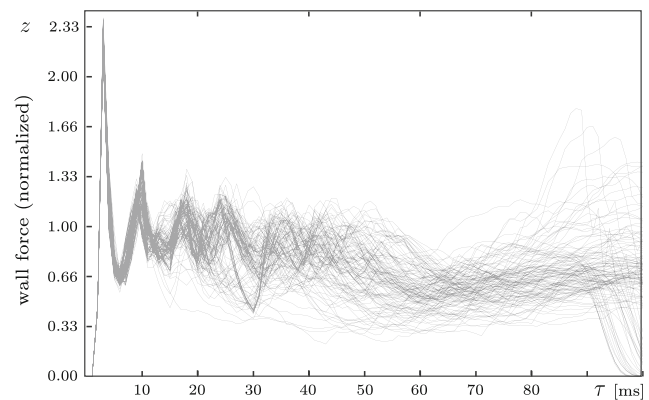


Fig. 1 Crash analysis, time history of wall-force $z(\tau)$, $m = 200$ process records

[19]. Numerical approaches for respective investigations are provided, in particular, on the basis of finite element analysis [1, 2].

Problems in the detection of branching points appear if the process is superimposed by noise. Investigations can be carried out with numerical methods of stochastic mechanics [6, 11, 28]. If the available information is not sufficient for a reliable stochastic modeling, an analysis can also be pursued based on other uncertainty models [22, 23]. In any case, a bundle of noisy process records is obtained. If the signal-to-noise ratio is low, branching points may be concealed by noise making the various physical paths of the mechanical system no longer distinguishable clearly. This is a typical problem in crashworthiness investigations. Both experiments and numerical analyses are characterized by significant noise [9, 20, 25, 27]. An example of a stochastic numerical crash analysis is shown in Fig. 1.

A vehicle component is moved against a rigid wall, and the time history of the wall-force $z(\tau)$ is computed in a nonlinear dynamical analysis with the tool LS-DYNA [12]. Distinctive paths cannot be recognized. But it can be observed that the uncertainty of the process grows significantly with increasing time. This indicates the existence of branching points and various physical paths of the system. An identification of the branching points is, however, not feasible with a standard approach. The dynamical crash analysis is associated with large structures in terms of degrees of freedom and requires a large number of time steps. The analysis of one process record typically takes several hours. Consequently, a detection of branching points based on mechanical principles is not reasonable due to a tremendous computational cost. Further, the number of analyzed process records is usually too small to obtain reliable results from a statistical approach. The problem becomes worse in multi-dimensional and multi-variate cases, which exclude a visual support of the evaluation.

These difficulties counter the industrial ambition to detect the branching points as a basis for an optimum

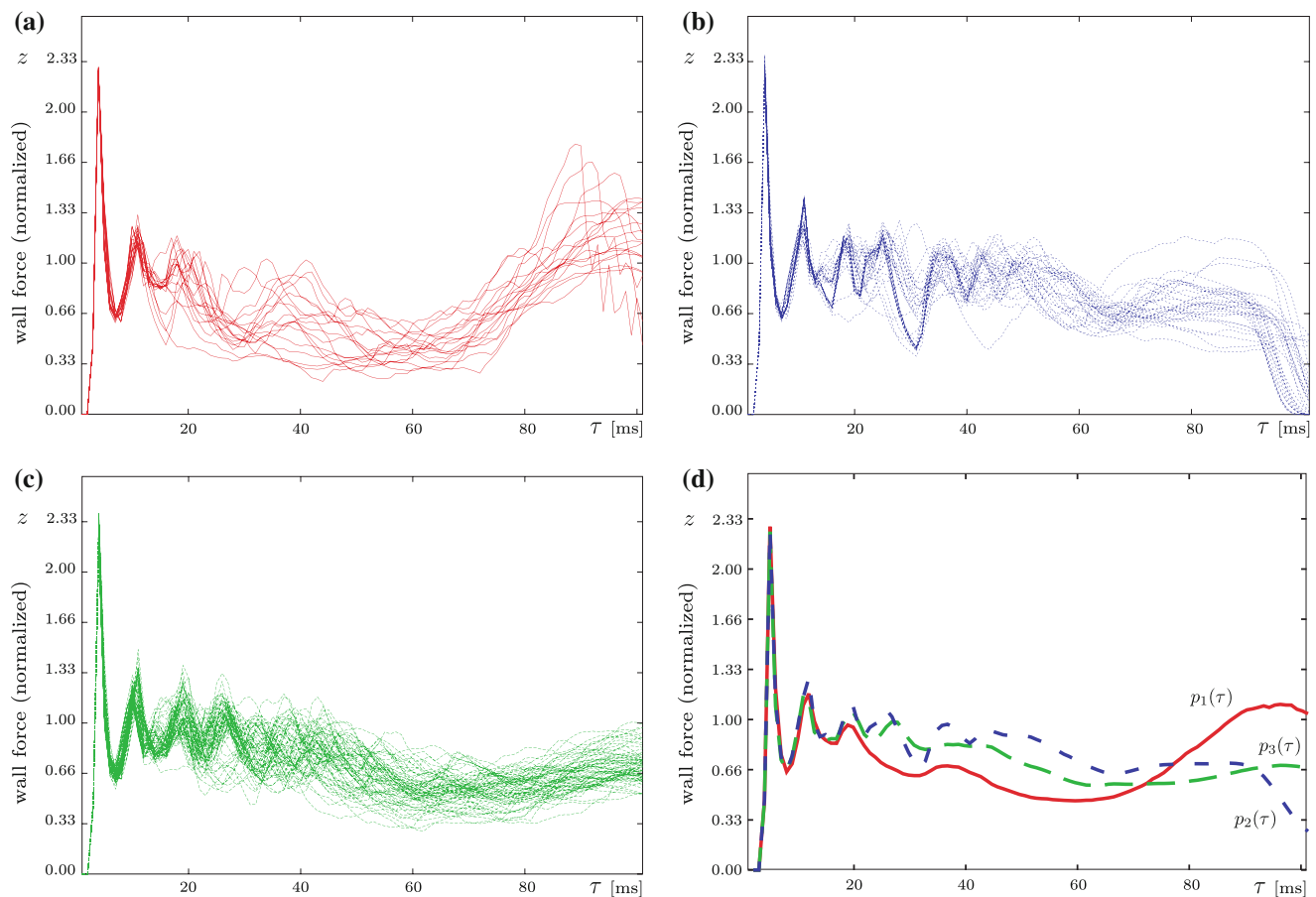


Fig. 2 Distinctive process groups $C_i(\tau)$ and representative process records $p_i(\tau)$. **a** Process group $C_1(\tau)$. **b** Process group $C_2(\tau)$. **c** Process group $C_3(\tau)$. **d** Representative processes records $p_i(\tau)$

crashworthiness design of vehicle structures [25]. Significant effort is made to reveal the mechanical background for the branching of the time histories to identify characteristic failure modes. If the branching point have been detected, a systematic mechanical investigation with limited numerical effort can be conducted. This permits a detailed interpretation of the structural behavior within the time domain. Based on this information constructive measures can be applied to eliminate certain structural behavior modes. This reduces the uncertainty of the system response and, consequently, the failure probability. For instance, the design of stiffeners and crimping can be adjusted to prevent certain buckling modes of local parts of sheets.

A proper numerical method for the solution of this conflict is not available to the authors's best knowledge. Previous extensive research on the branching of processes, see [31,32], involved stochastic process models, the construction of which is not feasible in the cases considered herein. An attempt for solution is, thus, made following the approach in [18]. Methods of cluster analysis [7,15] are employed as the basis. This approach is purely mathematical and

independent of the physical problem. It may, thus, be applied to various similar problems, in which the involved processes can be understood as generated in the fashion of Eq. (1). The potential of the approach pursued in [18] becomes obvious in Fig. 2a–c. Three distinctive process groups have been identified, each associated with its own failure mechanism of the vehicle structure. A representative process record has been extracted for each group Fig. 2d.

Subsequently, the concept and the algorithm for the detection of branching points in noisy processes are elucidated in Sect. 2. Then, the selected cluster method and the quality measure for assessing the cluster configuration are discussed in Sect. 3. Section 4 provides a numerical example as well as a practical example from car industry. Finally, the paper is summarized in a conclusions section.

2 Concept and algorithm

The method for the detection of branching points is developed for noisy processes generated according to Eq. (1).

It is assumed that a bundle or set

$$M = \{z_r(\underline{\theta}), r = 1, \dots, m\} \quad (2)$$

of process records $z_r(\underline{\theta})$ are available from numerical simulations or from experiments. Each process record $z_r(\underline{\theta})$ is known as a series of vectors $z_r(\underline{\theta}_i)$ for discretized process coordinates $\underline{\theta}_i$,

$$z_r(\underline{\theta}) = \{z_r(\underline{\theta}_i) \forall i\}, \quad r = 1, \dots, m. \quad (3)$$

The discretization of the process coordinates is the same for all process records. That is, a point set

$$M_i = \{z_r(\underline{\theta}_i), r = 1, \dots, m\} \quad (4)$$

constituted by process variables $z_r(\underline{\theta}_i)$ is available for each coordinate vector $\underline{\theta}_i$. The idea is to detect branching points by observing the point set Eq. (4) in dependence on a varying coordinates $\underline{\theta}_i$. At a branching point b with $i = i_b$ and $\underline{\theta}_i = \underline{\theta}_{ib}$, the set M_i is split into several subsets

$$M_{1,i} \subseteq M_i, \dots, M_{k,i} \subseteq M_i, \dots, M_{n_i,i} \subseteq M_i, \quad i = i_b, \quad (5)$$

and with restriction to classical set theory

$$M_i = \bigcup_{k=1}^{n_i} M_{k,i}, \quad i = i_b. \quad (6)$$

This splitting provides suitable information to subdivide the set M of process records from Eq. (2) into homogeneous process groups

$$P_k(\underline{\theta}) = \{z_r(\underline{\theta}) \mid z_r(\underline{\theta}_i) \in M_{k,i}\}, \quad k = 1, \dots, n_i, \quad i = i_b, \quad (7)$$

$$M = \bigcup_{k=1}^{n_i} P_k(\underline{\theta}), \quad i = i_b. \quad (8)$$

This principle is shown in the upper part of Fig. 3 with a respective indexing for the handling of various branching points as described below.

In crashworthiness investigations the process records $z_r(\underline{\theta})$ typically represent time-dependent quantities such as forces or energy measures. The processes are highly dynamical and involve strong geometrical and physical nonlinearities. Mechanical effects such as kinematics with large deformations, buckling, material plastification, and fracture appear and must be considered in the numerical analysis. Consequently, the results of both experiments and numerical analyses are very sensitive with respect to even very small changes in the structural parameters and boundary conditions. In addition, noise appears in the results. Numerical noise is caused by, even small, approximation errors, round-off errors, and random effects in the CPU. Experimental noise comes from uncontrollable small changes in the experimental set-up and in environmental conditions. The process records obtained from test series or from repeated calculations within a Monte Carlo simulation, then, usually show a non-uniform

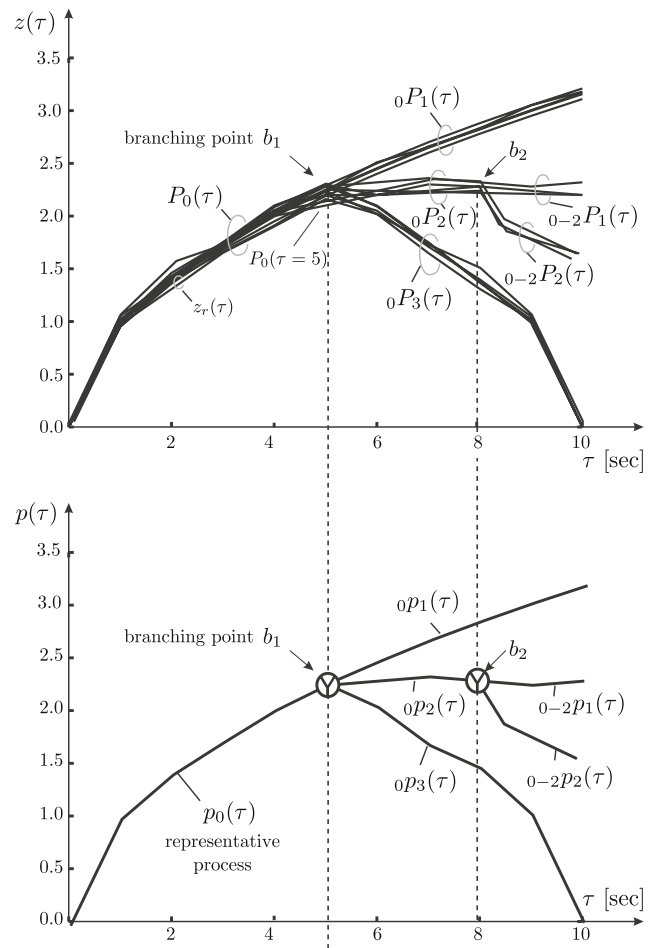


Fig. 3 Illustration of two branching points; b_1 —branching of process group $P_0(\tau)$ into the three process subgroups ${}_0P_1(\tau)$, ${}_0P_2(\tau)$, ${}_0P_3(\tau)$; b_2 —branching of the process subgroup ${}_0P_2(\tau)$ into the two process subgroups ${}_{0-2}P_1(\tau)$, ${}_{0-2}P_2(\tau)$; associated representative processes $p(\tau)$

behavior, which is only partly caused by mechanical effects. That is, the separation of the set M_i is interfered by noise. And the subsets $M_{1,i}, \dots, M_{k,i}, \dots, M_{n_i,i}$ cannot be identified clearly.

The splitting of the set M_i can, however, be detected with the aid of cluster analysis. The pattern of the set M_i is investigated numerically for all coordinate vectors $\underline{\theta}_i$ to identify the number and the arrangement of the subsets $M_{1,i}, \dots, M_{k,i}, \dots, M_{n_i,i}$. Let τ be a component of the coordinate vector $\underline{\theta}$ and i represent the counter for the coordinate discretization in τ -direction. Then, only the component τ is changed from τ_{i-1} to τ_i , all other components in $\underline{\theta}$ are frozen. For convenience in notation (List of symbols), the indices of the frozen components are abandoned subsequently so that the vectors $\underline{\theta}_{i-1}$ and $\underline{\theta}_i$ correspond to the values τ_{i-1} and τ_i , respectively. In the overall investigation this is applied to all coordinate vectors $\underline{\theta}_i$ and in a rotating fashion with respect to all components of $\underline{\theta}$ selected as τ . With this convention

the following criteria are formulated for the existence of a branching point b at $\underline{\theta}_{i-1}$:

1. The uncertainty of the set M_i , grows significantly in the step from $\underline{\theta}_{i-1}$ to $\underline{\theta}_i$.
2. A structure of subsets is formed within the set M_i in the step from $\underline{\theta}_{i-1}$ to $\underline{\theta}_i$.

Criterion 1 can be evaluated with the aid of a suitable measure of uncertainty \mathcal{U} . Depending on the characteristics of the uncertainty of the process different options for the selection of the measure \mathcal{U} exist. If the uncertainty is clearly of stochastic nature, the variance is an appropriate choice for \mathcal{U} . This may be applied, for instance, to the individual coordinates in the set M_i (not the coordinates of the process) or to the distances of its points with respect to the center of gravity. If the uncertainty of M_i is contaminated by non-stochastic components such as chaotic or unclear influences, set theory provides a variety of alternative energy and entropy measures for \mathcal{U} . Criterion 1 is satisfied if

$$\mathcal{U}(M_{i-1}) \ll \mathcal{U}(M_i). \quad (9)$$

Criterion 2 requires an evaluation of the sets M_i in a cluster analysis; see Sect. 3. It is investigated if the structure of the set M_i can be represented appropriately by a set of subsets $M_{k,i}$. In terms of cluster analysis, each subset $M_{k,i}$ is identified as a cluster $C_{k,i} = M_{k,i}$. And the unity of all clusters represents the set M_i . The set of all clusters is referred to as cluster configuration

$$\mathcal{C}_i = \{C_{k,i}, k = 1, \dots, n_i\}, \quad (10)$$

and clustering of the set M_i refers to the mapping

$$M_i \rightarrow \mathcal{C}_i. \quad (11)$$

As the clustering does usually not lead to a unique result, the various possible cluster configurations \mathcal{C}_i are assessed with a quality measure

$$G: \mathcal{C}_i \times \mathcal{C}_i \rightarrow \mathbb{R} \quad (12)$$

to identify the most appropriate number n_i of clusters. Herein, the silhouette coefficient is selected as a quality measure, $G = SC(n_i)$; see Sect. 3.2. An appropriate cluster configuration can then be identified by a silhouette coefficient $SC(n_i)$ larger than a predefined minimum value SC_{\min} . In a series of practical applications the minimum value $SC_{\min} = 0.75$ has been found as suitable, see also Table 1. The numerical formulation of Criterion 2 is then

$$SC(n_{i-1}) < SC_{\min} \forall n_{i-1} > 1 \\ \wedge SC(n_i) \geq SC_{\min}, n_i > 1. \quad (13)$$

In the case of a low signal-to-noise ratio Criteria 1 and 2 according to Eqs. (9) and (13) can be too sensitive and may

Table 1 Interpretation of the silhouette coefficient after [16]

SC	Cluster configuration \mathcal{C} describes the structure of the dataset ...
0.71–1.00	Very well
0.51–0.70	Well
0.26–0.50	Poorly; may be spurious
≤ 0.25	No applicable configuration

lead to spurious results. It is thus reasonable to extend both criteria to a consideration of a series of process coordinate values $\tau_{i-1}, \tau_i, \tau_{i+1}, \dots$ of specific length L . That is, Eqs. (9) and (13) are rewritten as

$$\mathcal{U}(M_{i-1}) \ll \mathcal{U}(M_j), \quad j = i, \dots, i-1 + L_U, \quad (14)$$

and

$$SC(n_{i-1}) < SC_{\min} \forall n_{i-1} > 1 \\ \wedge SC(n_j) \geq SC_{\min} n_j > 1, \\ j = i, \dots, i-1 + L_{SC}. \quad (15)$$

The values for L_U and L_{SC} have to be selected in dependance on the particular problem. In the investigation of processes from crash analysis the setting $L_U = L_{SC} = 4$ has been found as appropriate in most cases. Another phenomenon is the occurrence of outliers with respect to the sets $M_{k,i}$. This is associated with a low signal-to-noise ratio, too. A numerical identification of outliers and their exclusion from the sets $M_{k,i}$ can be achieved with the aid of fuzzy cluster analysis; see Sect. 3.1. The identified outliers are summarized in a separate set O_i , and the union property from Eq. (6) is changed to

$$M_i = \left(\bigcup_{k=1}^{n_i} M_{k,i} \right) \cup O_i, \quad i = i_b. \quad (16)$$

The separation of outliers prevents a bias in the sets $M_{k,i}$. Such bias may, otherwise, lead to problems in subsequent mechanical evaluations.

For an efficient detection of branching points the specific features of Criteria 1 and 2 can be exploited. While an increase of uncertainty (Criterion 1) is a strong indicator but not a condition, the formation of subsets (Criterion 2) is a sufficient condition for the existence of a branching point. The numerical effort for the evaluation of Criterion 2 is, however, significantly higher compared to the evaluation of Criterion 1. The latter applies, in particular, when searching for all branching points b_q , $q = 1, \dots, n_b$, in a noisy process, and when the number of process records and of discretization points $\underline{\theta}_i$ is large. It is, thus, proposed to apply Criterion 1 in a first fast run to pre-select parts of the process with potential branching points. The application of Criterion 2 can then

Algorithm 1 Algorithm for the detection of branching points**Iteration over process group until all groups are analyzed completely**

1. Select process group: $\dots P_k(\underline{\theta})$ (initially, $P_0(\underline{\theta})$)
If all process groups are analyzed, then the detection of branching points is completed.
2. Optional: pre-select process parts with potential branching points (Criterion 1, Eq. (14))
3. Specify the values for SC_{\min} and L_{SC} for Criterion 2 (Eq. (15))
4. **Iteration over $\underline{\theta}_i$ until Criterion 2 (Eq. (15)) is satisfied or end of process is reached**
 - (a) Specify the starting point $\underline{\theta}_{i-1}$ (systematic increase of the components of $\underline{\theta}$)
If the entire process group $\dots P_k(\underline{\theta})$ is analyzed, then proceed with the next process group $\dots P_k(\underline{\theta})$, step 1
 - (b) Select the component τ of $\underline{\theta}$ (search direction)
If all search directions are evaluated, then proceed with a new starting point $\underline{\theta}_{i-1}$, step 4a
 - i. Evaluate Criterion 2
 - Build the point sets $M_j = \{z_r(\underline{\theta}_j) \mid z_r(\underline{\theta}) \in \dots P_k(\underline{\theta})\}$ at $\underline{\theta}_j$ for $j = i, \dots, i-1 + L_{SC}$
 - Determine the clustering $M_j \rightarrow C_j(n_j)$ for $n_j = 2, 3, 4, \dots$ clusters and $j = i, \dots, i-1 + L_{SC}$ (Sect. 3.1)
 - Calculate the silhouette coefficient $SC(n_j)$ for all determined cluster configurations $C_j(n_j) = \{C_{l,j}, l = 1, \dots, n_j\}$ (Sect. 3.2)
 - Check Eq. (15)
 - If Eq. (15) is not satisfied, then set $i = i + 1$
 - If end of process is exceeded in τ -direction, then proceed with the next component τ of $\underline{\theta}$, step 4b
 - Otherwise, evaluate Criterion 2 for the new $\underline{\theta}_{i-1}$, step 4(b)i
 - If Eq. (15) is satisfied, then
 - Register the branching point b_q as detected at $\underline{\theta}_{i-1}$
 - Build the subsets $M_{l,i} = C_{l,i}$ from the cluster configuration $C_j(n_j) = \{C_{l,j}, l = 1, \dots, n_j\}$ with the largest $SC(n_j)$ for $j = i$
 - Determine the process subgroups $\dots_k P_l(\underline{\theta}) = \{z_r(\underline{\theta}) \mid z_r(\underline{\theta}_i) \in M_{l,i}\}, l = 1, \dots, n_i$
 - Subdivide the current process group $\dots P_k(\underline{\theta}) \rightarrow \dots_k P_1(\underline{\theta}), \dots, \dots_k P_l(\underline{\theta}), \dots, \dots_k P_{n_i}(\underline{\theta}), \dots_k O_i(\underline{\theta})$
 - ii. Terminate investigation of process group $\dots P_k(\underline{\theta})$
 - iii. Add the determined process subgroups $\dots_k P_1(\underline{\theta}), \dots, \dots_k P_l(\underline{\theta}), \dots, \dots_k P_{n_i}(\underline{\theta})$ to the investigation portfolio
 - iv. Proceed with the analysis of the next process group $\dots P_k(\underline{\theta})$, step 1

be limited to the pre-selected process parts to verify the existence of the branching points and to detect their position more specifically.

When a branching point has been detected, the set of process records is subdivided into homogeneous process groups according to Eq. (7). The cluster analysis in Criterion 2 provides sufficient information for this subdivision in form of the subsets $M_{k,i} = C_{k,i}$ and the separated outliers O_i . The identified homogeneous process groups are then investigated separately to detect further branching points. This is realized in a systematic analysis in a recursive fashion as summarized in Algorithm 1. The analysis starts at a point of the process, at which only one homogeneous process group $P_0(\underline{\theta})$ exists. This can be the initialization point of the process or a pre-selected starting point specified with the aid of Criterion 1. It is proceeded point by point with increasing i until the first branching point b_1 is detected at $\underline{\theta}_i$, $i = i_{b1}$. The process group $P_0(\underline{\theta})$ is then subdivided in the subgroups

$$P_0(\underline{\theta}) \rightarrow {}_0P_1(\underline{\theta}), \dots, {}_0P_k(\underline{\theta}), \dots, {}_0P_{n_i}(\underline{\theta}), {}_0O_i(\underline{\theta}), \quad i = i_{b1}, \quad (17)$$

$$P_0(\underline{\theta}) = \left(\bigcup_{k=1}^{n_i} {}_0P_k(\underline{\theta}) \right) \cup {}_0O_i(\underline{\theta}), \quad i = i_{b1}, \quad (18)$$

with ${}_0P_k(\underline{\theta}) = P_k(\underline{\theta})$ according to Eq. (7) and $M_{k,i} = C_{k,i}$ from cluster analysis at b_1 . The separated outliers O_i are summarized in a special process group ${}_0O_i(\underline{\theta})$.

The detection procedure is then repeated separately for each subgroup ${}_0P_k(\underline{\theta})$, $k = 1, \dots, n_i$, $i = i_{b1}$, again starting from the initialization point of the process or from a pre-selected point determined with Criterion 1. When a branching point b_q is detected in subgroup ${}_0P_k(\underline{\theta})$, further subdivision of ${}_0P_k(\underline{\theta})$ is carried out,

$${}_0P_k(\underline{\theta}) \rightarrow {}_{0-k}P_1(\underline{\theta}), \dots, {}_{0-k}P_l(\underline{\theta}), \dots, {}_{0-k}P_{n_i}(\underline{\theta}), {}_{0-k}O_i(\underline{\theta}), \quad i = i_{bq}, \quad (19)$$

$${}_0P_k(\underline{\theta}) = \left(\bigcup_{l=1}^{n_i} {}_{0-k}P_l(\underline{\theta}) \right) \cup {}_{0-k}O_i(\underline{\theta}), \quad i = i_{bq}. \quad (20)$$

Then, all the subgroups ${}_{0-k}P_l(\underline{\theta})$, $l = 1, \dots, n_i$, $i = i_{bq}$ are investigated. It is proceeded in this recursive fashion until no more branching points are found; see Algorithm 1. Finally, a representative process record p —with respective indexing—is selected from each process subgroup P as illustrated in the lower part of Fig. 3. For example, the record with the smallest distance to the mean of the process subgroup, in an integral sense, may be selected as the representative record p . The mechanical background of the branching can then be investigated by means of the representative records p . A separate investigation of the outliers from the process groups $O_i(\underline{\theta})$ may be carried out, in addition, to identify rare but momentous mechanical effects and possible spurious behavior modes.

It is important for the overall analysis to separate the subgroups of processes immediately after the identification of a branching point. This ensures that crossovers and spurious reunifications of process groups can be identified as such; see first example in Sect. 4. A misinterpretation of these phenomena may, otherwise, lead to wrong mechanical conclusions or unnecessary mechanical investigations. Furthermore, the immediate separation of subgroups prevents the multiple detection of branching points which are associated with the same mechanical effects. If $\underline{\theta}$ has more than one component, branching occurs along a parameter curve in the space of the process coordinates. It is then sufficient to identify the branching at one point on this curve to separate the subgroups. But the subsequent mechanical investigation of the associated representative process records p includes the entire domain of branching. If the branching characteristics change at some point on such parameter curve, this is identified as a new branching point in one of the process subgroups.

The results from the proposed algorithm are not unique for the following reasons. First, the adjustment of the thresholds to activate Criteria 1 and 2 (sufficiently large difference in U and value of SC_{\min} , respectively) influence the result. This includes the specification of L_U and L_{SC} in Eqs. (14) and (15). Second, if the process depends on more than one process coordinate, the order of selection of τ out of $\underline{\theta}$ may change the result. Thus, sensitivities of the result with respect to the thresholds and to the order of selection of the coordinates in the set M_i (not the coordinates of the process) should always be investigated to verify the results.

3 Cluster analysis of the process variables

3.1 Determination of clusters

Cluster analysis is a general numerical method to identify structures in data; its major field of application is pattern recognition [5, 10, 15, 16]. The goal is to subdivide a given set of elements into homogeneous subsets, which are referred to as clusters. And homogeneity is understood as a high degree of similarity between the elements of each subset. Herein, the given set is the point set M ; see Eqs. (10), (11) and Algorithm 1. And its elements are process variables $\underline{z}_r \in \underline{Z} \subseteq \mathbb{R}^n$; see Eq. (4). The similarity between elements \underline{z}_r and \underline{z}_s can then be expressed as a distance $d(\underline{z}_r, \underline{z}_s)$ with

$$d : \underline{Z} \times \underline{Z} \rightarrow \mathbb{R}. \quad (21)$$

The most popular choice for d is the Euclidean distance. The goal of the cluster analysis can be expressed in terms of the distance d with the following two components:

1. The degree of similarity between elements within each particular cluster C_k is maximum,

$$\sum_{\underline{z}_r, \underline{z}_s \in C_k} d(\underline{z}_r, \underline{z}_s) \Rightarrow \text{Min } \forall C_k, \quad k = 1, \dots, n. \quad (22)$$

This represents homogeneity within the clusters.

2. The degree of similarity between elements from different clusters C_k and C_l is minimum,

$$\sum_{\underline{z}_r \in C_k, \underline{z}_s \in C_l} d(\underline{z}_r, \underline{z}_s) \Rightarrow \text{Max} \\ \forall C_k, C_l, \quad k, l = 1, \dots, n, \quad k \neq l. \quad (23)$$

This characterizes heterogeneity between different clusters.

Solving Eqs. (22) and (23), simultaneously, cluster analysis is a two-criteria optimization problem. A variety of solution methods and procedures are available, which lump the two criteria together in one functional and determine a Pareto-optimal cluster configuration; see [5, 10, 15, 16].

A basic distinction between cluster analysis methods can be made by means of the manner in which the elements are assigned to the clusters. This assignment can be either unique, which is referred to as deterministic clustering, or it can be uncertain, which includes stochastic clustering and fuzzy clustering. In the case of *deterministic clustering*, each element is assigned to exactly one cluster. This is reasonable if the cluster structure is obvious as shown in Fig. 4a. In many cases, a deterministic clustering does, however, not comply with the data structure and may lead to a loss of information. A typical example is overlapping of clusters as appears in the noisy process data. In Fig. 4b, the element in the center can neither be uniquely assigned to the left, nor to the right cluster. If this element is assigned uniquely to one of the clusters, the information that the data set is almost symmetric is lost. Moreover, the cluster result does not provide any information on how typical the particular elements are for the cluster. In this manner, outliers cannot be identified. Thus, an uncertain clustering is more appropriate for our application.

In the investigations *fuzzy cluster analysis* [15, 26] has provided the most suitable basis for the clustering of noisy process data. Fuzzy cluster algorithms consider a smooth transitions between clusters (see Fig. 5) and take this into account when solving the optimization Eqs. (22) and (23). A gradual assignment of the elements \underline{z}_r to the clusters C_k is described with characteristic values $\mu_{k,r} \in [0, 1]$ (membership values) according to fuzzy set theory [33],

$$\mu_{k,r} = \mu(\underline{z}_r \in C_k). \quad (24)$$

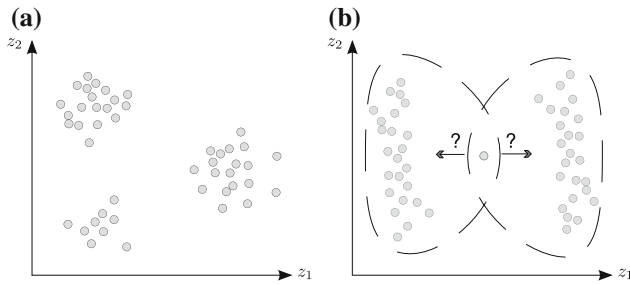


Fig. 4 Example point sets. **a** Point set with three obvious clusters. **b** Problem of unique assignment of elements to clusters

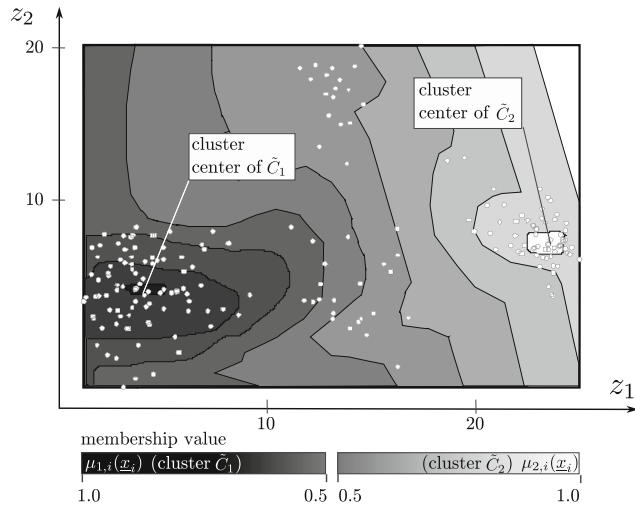


Fig. 5 Result of a fuzzy c-means clustering

The clusters C_k become discrete fuzzy sets

$$\tilde{C}_k = \{(\underline{z}_r, \mu_{k,r})\}. \quad (25)$$

A membership of $\mu_{k,r} = 1$ indicates a definite assignment of object \underline{z}_r to cluster \tilde{C}_k , whereas $\mu_{k,r} = 0$ stands for assigning \underline{z}_r to \tilde{C}_k on no account. These features are provided by the fuzzy c-means algorithm (FCM) [4, 15], which has been selected for the present work. An example of a result from FCM is shown in Fig. 5. Two clusters were detected. The degree of membership of the elements with respect to the clusters \tilde{C}_1 and \tilde{C}_2 , respectively, decrease with increasing distance from the cluster centers.

The problem in Fig. 4 is now solved by assigning the membership value $\mu_{k,r} = 0.5$ to the element in the center with respect to both clusters. The symmetry property remains visible. Further, the other elements would obtain higher membership values so that the central element can be identified as less typical for both clusters. Outliers can be identified by their low membership values and filtered out. With the requirement of a minimum cluster membership $\mu_{k, \min}$ sets of typical elements for the clusters

$$C_{k, \text{typ}} = \{\underline{z}_r \mid \mu_{k,r} > \mu_{k, \min}\}. \quad (26)$$

and, hence, groups of typical process records, can be found. This strategy has been adopted from the design of structures under uncertainty proposed in [3].

3.2 Quality assessment of cluster configuration

The determination of clusters requires a predefined number n of clusters; see Eqs. (22) and (23). This number n is, however, not known in advance. A common method for its specification is the search for a cluster configuration $\mathcal{C} = \{C_k, k = 1, \dots, n\}$ of optimal quality. In Fig. 4 it is obvious that a cluster configuration with $n = 3$ clusters is most appropriate and thus of optimal quality for the given data set. In general, the optimum number of clusters n_{opt} must be identified numerically. The cluster analysis is repeated for a series of numbers of clusters $n = 2, 3, \dots$. Then, the quality of the cluster configuration is evaluated for each n , and n_{opt} is selected for the final clustering. In practice, the series $n = 2, 3, \dots$ is terminated

- if n reaches a reasonable predefined maximum number $n = n_{\text{max}}$,
- if a peak value of the quality has been clearly identified, or
- if all possible cluster configurations have been investigated, which is indicated by n attaining the number of elements in the data set.

If the quality of \mathcal{C} is low for all $n \geq 2$, no suitable subdivision of the point set M exists so that only one cluster appears, and $n = 1$ is retained. This approach is called unsupervised clustering as no user interaction is involved.

The quality of a cluster configuration \mathcal{C} is evaluated by means of a quality measure G according to Eq. (12). In literature, a variety of quality measures G are proposed, which possess different features and characteristics [15, 16]. For the present application, the silhouette coefficient SC has been found as most suitable. This is available for the evaluation of deterministic cluster configurations. With $|C_k|$ denoting the cardinality of cluster C_k , the silhouette coefficient is defined as

$$SC = \frac{1}{n} \sum_{k=1}^n \frac{1}{|C_k|} \sum_{\underline{z}_r \in C_k} \frac{b(\underline{z}_r) - a_k(\underline{z}_r)}{\max[a_k(\underline{z}_r), b(\underline{z}_r)]}, \quad (27)$$

in which

$$a_k(\underline{z}_r) = \frac{1}{|C_k|} \sum_{\underline{z}_h \in C_k} d(\underline{z}_r, \underline{z}_h) \mid \underline{z}_r \in C_k, \quad (28)$$

$$b(\underline{z}_r) = \min_{\underline{z}_h \in C_k, k=1, \dots, n, h \neq r} [a_k(\underline{z}_h)]. \quad (29)$$

It measures the homogeneity and heterogeneity of a cluster configuration \mathcal{C} , simultaneously, which makes it

particularly effective. This quality measure varies in the interval $SC \in [-1, 1]$. Large values $SC \approx 1$ indicate a high quality of the cluster configuration. That is, the data set possesses a clear structure with n clusters of elements. In [16] an interpretation of the values SC is proposed according to Table 1.

The silhouette coefficient in Eq. (27) is only applicable to deterministic cluster configurations. As, herein, the fuzzy c-means algorithm is selected for clustering, an extension of the silhouette coefficient is formulated to deal with fuzzy clusters. This includes the membership values $\mu_{k,r}$ from Eqs. (24) and (25) in the evaluation of the fuzzy cluster configuration $\tilde{C} = \{\tilde{C}_k, k = 1, \dots, n\}$. Let m be the total number of elements z_r in the entire cluster configuration \tilde{C} . The silhouette coefficient SC_F for fuzzy clusters is proposed as

$$SC_F = \frac{1}{m} \sum_{k=1}^n \sum_{r=1}^m \left(\frac{b_{Fk}(z_r) - a_{Fk}(z_r)}{\max[a_{Fk}(z_r), b_{Fk}(z_r)]} \cdot \mu_{k,r} \right), \quad (30)$$

with

$$a_{Fk}(z_r) = \frac{\sum_{h=1}^m d(z_r, z_h) \cdot \mu_{k,h}}{\sum_{h=1}^m (\mu_{k,h}, r \neq h)}, \quad (31)$$

$$b_{Fk}(z_r) = \min_{l=1, \dots, n, l \neq k} \left[\frac{\sum_{h=1}^m d(z_r, z_h) \cdot \mu_{l,h}}{\sum_{h=1}^m (\mu_{l,h}, r \neq h)} \right]. \quad (32)$$

The extended silhouette coefficient SC_F for fuzzy cluster configurations takes on values in the interval $[-1, 1]$. For deterministic cluster configurations, SC_F according to Eq. (30) and SC from Eq. (27) lead to identical results. For the interpretation of the values of SC_F Table 1 remains applicable.

4 Examples

4.1 Numerically generated process

The algorithm for the detection of branching points is demonstrated for a random process with a one-dimensional process variable depending on a single process coordinate and obvious branching points. Specifically,

$$\mathbf{f} : \underline{X}(\tau) \rightarrow Z(\tau), \quad (33)$$

with $\underline{X}(\tau) = (X_1(\tau), \dots, X_5(\tau))$ involving five independent random variables with a standard uniform distribution, $X_q(\tau) \sim (0, 1)$, $q = 1, \dots, 5$. These produce the noise ε of the process. The process is synthesized numerically as a set

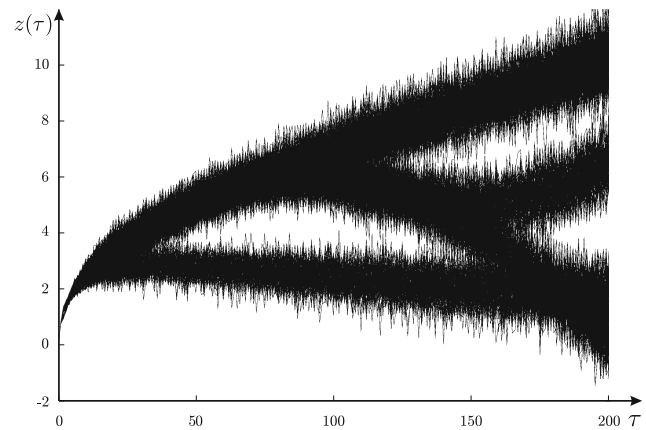


Fig. 6 Process records $z_r(\tau)$

of five functions,

$$\begin{aligned} Z(\tau) = \{ & \sqrt{\tau} + \varepsilon(X_1, \tau), \\ & 1.5 \cdot \ln(\tau) + \left(\frac{\tau}{55}\right) + \varepsilon(X_2, \tau), \\ & 2.0 \cdot \left(\sin\left(\frac{\pi \cdot \tau}{60}\right) + \left(\frac{\tau}{30}\right)\right) + 1.5 + \varepsilon(X_3, \tau), \\ & 5.0 \cdot \left(\sin\left(\frac{\pi \cdot \tau}{100}\right) - \left(\frac{\tau}{500}\right)\right) + 1.5 + \varepsilon(X_4, \tau), \\ & \left(\frac{60}{\tau + 30}\right) \cdot \ln\left(\frac{\tau + 2}{2}\right) + \varepsilon(X_5, \tau) \} \end{aligned} \quad (34)$$

with

$$\begin{aligned} \varepsilon(X_q, \tau) \\ = \begin{cases} \frac{\tau}{10} \cdot (1 - X_q) & \text{if } \tau < 10 \\ \frac{(\tau + 90)}{100} \cdot (1 - 2X_q) & \text{otherwise} \end{cases}, \quad q = 1, \dots, 5. \end{aligned} \quad (35)$$

A set of 100 process records $z_r(\tau)$ are generated for $\tau = 0, \dots, 200$ with a step width of $\Delta\tau = 1$; see Fig. 6.

The records $z_r(\tau)$ appear as a single bundle at the beginning of the process and split into four branches with increasing τ . The process includes three branching points and one crossover.

The numerical detection of the branching points is carried out with Criterion 2 only. The values $SC_{\min} = 0.75$ and $L_{SC} = 4$ are selected for the evaluation of Eq. (15). The cluster analysis is realized with the fuzzy c-means algorithm, and the extended silhouette coefficient SC_F from Eq. (30) is calculated for each cluster configuration. Three branching points are detected at $\tau_{b1} = 38$, $\tau_{b2} = 121$, and $\tau_{b3} = 161$ as shown in Fig. 7. The process is subdivided into four process groups, and representative process records p are identified as a moving average over 11 successive process values. The representative process records p show a good agreement with the actual functions from Eq. (34) without noise; see Fig. 7.

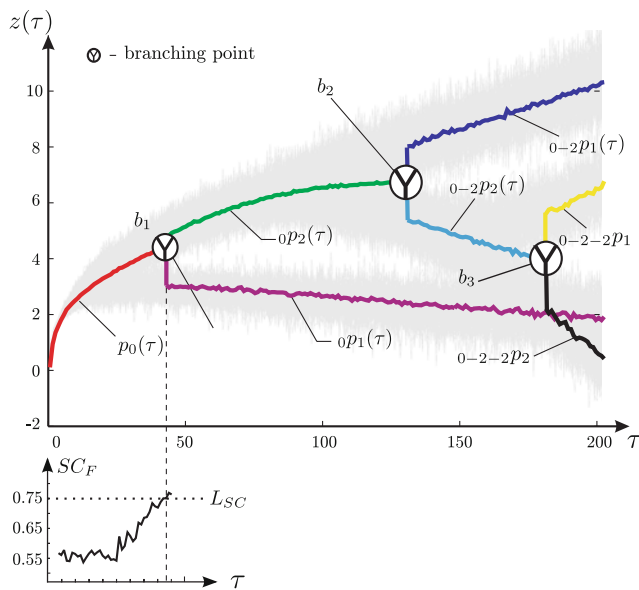


Fig. 7 Detected branching points b , representative process records p , and actual process without noise

The information about the branching points and process groups becomes important not only in the mechanical interpretation but also in the statistical evaluation of the process. Let $\tau = 200$ be a point of interest, for which histograms of the process value $z(\tau)$ are investigated. Without the knowledge about the branching points, the result in Fig. 8a is obtained. The two different components in the first frequency hill in the range $z = -1.3, \dots, 3.5$ cannot be separated. But a joint statistical evaluation is unjustified and may lead to erroneous results and interpretations. In contrast to this, the subdivision of the process in process groups enables the determination of individual histograms for the process groups; see Fig. 8b. The process values in the range $z = -1.3, \dots, 3.5$ now appear in a separated form, ready for a sound statistical evaluation.

4.2 Crash analysis

The results from a stochastic crash simulation of a vehicle component are investigated to reveal the characteristic mechanical behavior modes. In the crash analysis the component shown in Fig. 9 was moved, with a constant velocity v_0 , against a rigid wall. The force acting on the wall was analyzed. This dynamic analysis was carried out with the nonlinear Finite Element code LS-DYNA [12]. A time-step integration was performed over the entire crash process resulting in a force-time dependency. On this basis it was investigated whether the design of the component is appropriate to comply with a required upper limit of the wall-force. Uncertainties and fluctuations of selected structural parameters were taken into account with the aid of random quantities. The sheet thicknesses of the four labeled parts in Fig. 9 and

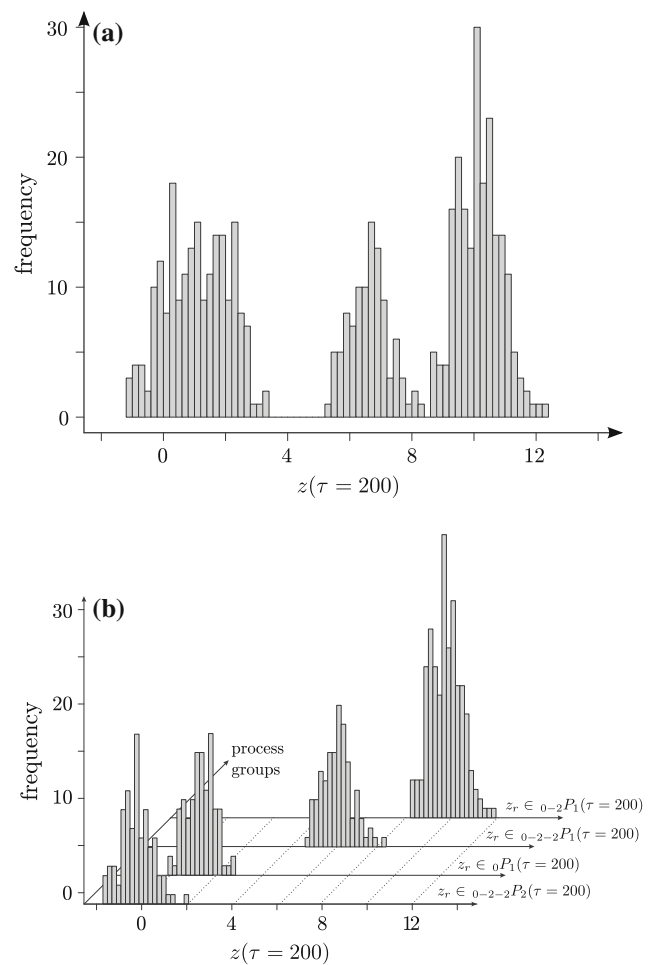


Fig. 8 Histograms for $z(\tau = 200)$. **a** Histogram for $z(\tau = 200)$ without separated process groups. **b** Histogram for $z(\tau = 200)$ with separated process groups

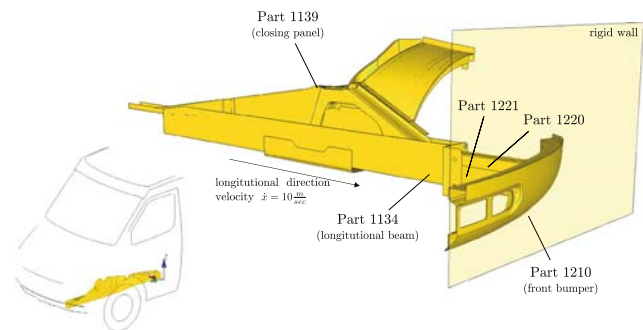
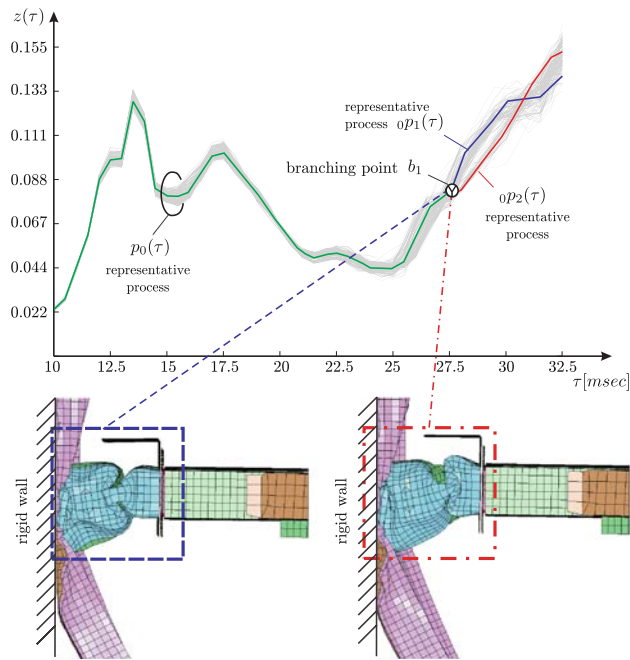


Fig. 9 Vehicle component; courtesy of Daimler AG [29]

a scaling factor for the yield surface were modeled with normal distributions according to Table 2. Then, a direct Monte Carlo simulation was carried out, in which a total of 1,500 process records $z_r(\tau)$ for the force-time dependency were produced; see Fig. 10. And a probability for exceeding the limit of the stonewall force was estimated.

Table 2 Random structural parameters, values standardized

Random variable X_q	Description	Part	Expected value μ_q	Standard deviation σ_q
X_1	Sheet thickness	1,139	10.0	0.20
X_2	Sheet thickness	1,134	10.0	0.21
X_3	Sheet thickness	1,210	10.0	0.62
X_4	Sheet thickness	1,221	10.0	0.50
X_5	Scaling factor	1,134	10.0	0.50

**Fig. 10** Process records $z_q(\tau)$ from Monte Carlo simulation, branching point b_1 , representative process records ${}_0p_1(\tau)$ and ${}_0p_2(\tau)$ and associated deformation characteristics at $\tau_{b1} = 27.5$ ms

In view of a steady product improvement, those investigations raise the following questions: Which mechanical effects are associated with an exceedance of the limit? How can the probability of exceedance be reduced by simple mechanical measures? A detailed investigation of the 1,500 process records is not feasible due to a tremendous effort. This is the starting point for the proposed method. In an initial evaluation of Criterion 1 (Eq. (14)) the process part up to $\tau = 12.5$ ms is found as homogeneous and prospectively not containing branching points. The consideration of Criterion 2 is limited to the process part $\tau \in [12.5, 32.5]$ ms. The parameters for the evaluation of Eq. (15) are selected as $SC_{\min} = 0.75$ and $L_{SC} = 4$. The cluster analysis is carried out with the fuzzy c-means algorithm. The silhouette coefficient is calculated according to Eq. (30). A branching point is detected at $\tau_{b1} = 27.5$ ms, and the process $P_0(\tau)$ is subdivided into two homogeneous process groups ${}_0P_1(\tau)$

and ${}_0P_2(\tau)$. The representative process records ${}_0p_1(\tau) = z_{p1}(\tau) \in {}_0P_1(\tau)$ and ${}_0p_2(\tau) = z_{p2}(\tau) \in {}_0P_2(\tau)$ are identified for the two process groups; see Fig. 10. The mechanical behavior is then investigated in detail for these two process records ${}_0p_1(\tau)$ and ${}_0p_2(\tau)$ with focus on the branching point at $\tau_{b1} = 27.5$ ms. The reason for the branching is revealed as a significant difference in the deformation characteristics associated with ${}_0p_1(\tau)$ and ${}_0p_2(\tau)$. Based on this insight, slight changes in the design of the vehicle component can now be introduced to eliminate the behavior mode which leads to an exceedance of the limit of the stonewall force. This reduces the uncertainty of the structural response as well as of the failure probability, and it increases the product quality.

5 Conclusions

A numerical method has been presented in this paper for the detection of branching points in noisy processes. The development was driven by the desire of the car industry to identify different mechanical behavior modes out of a set of noisy process records from crash analyses. As respective attempts with traditional methods did not lead to satisfying results, a novel approach based on cluster analysis has been pursued. The formulated algorithm has been verified in numerical tests and is applied in practice. The branching of noisy processes from crash simulations and from experiments can so be detected properly, and the associated mechanical behavior modes can be identified. This helps to improve the design and product quality of vehicle components. In this framework, the proposed method has been applied with a very limited dimensionality of process coordinates and process variables.

The proposed detection method is a form of data analysis and is not connected to a physical or mechanical background. It may, thus, also be applied to similar problems in further engineering fields and applications. A potential utilization is seen in the evaluation of time records of loads to filter out the major characteristics. This might be helpful, for example, to adjust design loads based on earthquake records for specific regions. The step to further application fields may also initiate an extension to a higher dimensionality of process coordinates and process variables.

Acknowledgments The authors gratefully acknowledge the elaboration of the numerical example in Section 4.1 by Dipl.-Ing. Christoph Zopf within his Master Thesis at the Institute of Structural Analysis, TU Dresden, Germany. Further appreciation is given to M. Thiele, Dynamore GmbH, Germany, for raising the question for a suitable method to identify branching points in vehicle crash processes. Sincere thanks are expressed to the Daimler AG for providing us with the specific sample structure of the vehicle component for the example in Sect. 4.2.

References

- Argyris JH, Mlejnek H-P (1991) Dynamics of structures, volume V of Texts on computational mechanics. Elsevier, Amsterdam
- Bathe K-J (1991) Finite element procedures, vols 1, 2. Prentice-Hall, Upper Saddle River
- Beer M, Liebscher M (2008) Designing robust structures—a non-linear simulation based approach. Special Issue Comput Struct 86(10):1102–1122
- Bezdek JC, Ehrlich R, Full W (1984) FCM: the fuzzy c-means clustering algorithm. Comput Geosci 10:191–203
- Bezdek JC, Pal SK (eds) (1992) Fuzzy models for pattern recognition—methods that search for structures in data. IEEE Press, Piscataway
- Deodatis G, Spanos PD (eds) (2007) Computational stochastic mechanics. Millpress, Rotterdam
- Duran BS, Odell PL (1974) Cluster analysis—a survey. Lecture notes in economics and mathematical systems. Springer, Berlin
- Eckstein U, Harte R, Krätzig WB, Wittek U (1987) Simulation of static and kinetic buckling of unstiffened and stiffened cooling tower shells. Eng Struct 9(1):9–18
- Gandhi UN, Hu SJ (1995) Data-based approach in modeling automobile crash. Int J Impact Eng 16(1):95–118
- Gath I, Geva AB (1989) Unsupervised optimal fuzzy clustering. IEEE Trans Pattern Anal Mach Intell 11(7):773–781
- Ghanem RG, Spanos PD (1991) Stochastic finite elements: a spectral approach. Springer, New York. Revised Edition 2003, Dover Publications, Mineola
- Hallquist JO (1998) LS-DYNA theoretical manual. Livermore Software Technology Corporation, Livermore
- Hartmann D, Breidt M, Nguyen VV, Stangenberg F, Höhler S, Schweizerhof K, Mattern S, Blankenhorn G, Möller B, Liebscher M (2008) Structural collapse simulation under consideration of uncertainty—Fundamental concept and results. Comput Struct 86(21–22):2064–2078
- Hilburger MW, Starnes JHJr (2005) Buckling behavior of compression-loaded composite cylindrical shells with reinforced cut-outs. Int J Non-Linear Mech 40(7):1005–1021
- Höppner F, Klawonn F, Kruse R, Runkler T (1999) Fuzzy cluster analysis: methods for classification, data analysis and image recognition. Wiley, Chichester
- Kaufman L, Rousseeuw PJ (1990) Finding groups in data: an introduction to cluster analysis. Wiley, Chichester
- Lee YS, Vakakis AF, Bergman LA, McFarland DM, Kerschen G (2005) Triggering mechanisms of limit cycle oscillations due to aeroelastic instability. J Fluids Struct 21(5–7):485–529
- Liebscher M (2007) Design and assessment of structures under uncertainty—solving the inverse problem with methods of the explorative data analysis. PhD thesis, TU Dresden, Institute for Statics and Dynamics of Structures, vol. 13 (in German), Dresden, Germany
- Litak G, Borowiec M, Friswell MI, Szabelski K (2008) Chaotic vibration of a quarter-car model excited by the road surface profile. Commun Nonlinear Sci Numer Simulation 13(7):1373–1383
- Livermore Software Technology Corp. (ed) (2008) 10th International LS-DYNA users conference, Dearborn, MI
- Möller B, Liebscher M, Schweizerhof K, Mattern S, Blankenhorn G (2008) Structural collapse simulation under consideration of uncertainty—improvement of numerical efficiency. Comput Struct 86(19–20):1875–1884
- Möller B, Beer M (2004) Fuzzy randomness—uncertainty in civil engineering and computational mechanics. Springer, Berlin
- Moens D, Vandeputte D (2005) A survey of non-probabilistic uncertainty treatment in finite element analysis. Comput Methods Appl Mech Eng 194(1):1527–1555
- Reitinger R, Ramm E (1995) Buckling and imperfection sensitivity in the optimization of shell structures. Thin-Walled Struct 23(1–4):159–177
- Roux WJ, Stander N, Günther F, Müllerschön H (2006) Stochastic analysis of highly non-linear structures. Int J Numer Methods Eng 65(8):1221–1242
- Ruspini EH (1969) A new approach to clustering. Inf Control 15(1):22–32
- Safety Test Instrumentation Stds Comm. (2003) Instrumentation for impact test-part 1-electronic instrumentation (j211/1). Technical report, SAE
- Schenk CA, Schuëller GI (2005) Uncertainty assessment of large finite element systems. Springer, Berlin
- Thiele M, Liebscher M, Graf W (2005) 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
- VanMarcke E (1983) Random fields: analysis and synthesis. MIT Press, Cambridge
- Vatutin VA, Zubkov AM (1987) Branching processes. I. J Math Sci 39(1):2431–2475
- Vatutin VA, Zubkov AM (1993) Branching processes. II. J Math Sci 67(6):3407–3485
- Zimmermann H-J (1992) Fuzzy set theory and its applications. Kluwer, Boston