Structural Process Variety and Standardization

Clemens Schreiber
AIFB
Karlsruhe Institute of Technology, Karlsruhe, Germany
clemens.schreiber@kit.edu

Amine Abbad-Andaloussi
Institute of Computer Science
University of St. Gallen, St. Gallen, Switzerland
amine.abbad-andaloussi@unisg.ch

***Abstract*—Structural process variety occurs due to control-flow differences among alternative process execution variants, i.e., trace variants. Structural variety is an important indicator for analyzing process performance, identifying process changes, guiding process model discovery, or sampling process data. Yet, existing measures in the research literature do not capture structural process variety in a consistent and mathematically sound way. In this paper, we identify a property called ultrametric, which ensures that a measure, based on the pairwise comparison of trace variants, leads to a meaningful and unique structural variety value. Thereupon, we propose a measurement framework, that guarantees the ultrametric property and provides a ranking of trace variants within a set of multiple trace variants, according to their contribution to the set's overall structural variety. One important implication of our work is, that our framework can be applied to substantially improve process standardization efforts.**

***Index Terms*—process measurement, process variety, trace variant analysis, process flexibility**

I. Introduction

The analysis of process data often reveals that a process consists of numerous alternative process execution variants, also called trace variants, which differ based on their respective sequence of activities [24]. The overall structural variety of a set of distinct trace variants can be defined based on the structural differences among the trace variants. These differences occur due to distinct activity patterns, e.g., indicating a particular ordering, or a repetition of activities. Structural variety is thereby an important indicator for the analysis of process flexibility and performance [16], for the identification of process changes over time [6], for the selection of process model discovery algorithms [1], [2], or for the sampling of relevant process data [11]. Yet, existing measures in the literature [1], [2], [16], which aim to capture structural process variety, have at least one of the following three drawbacks. First, they are not mathematically sound and therefore do not provide a unique and meaningful outcome. Second, they do not allow to rank the trace variants according to their contribution to the overall structural variety of a set. Hence, it is not possible for a decision-maker to uniquely identify how much a specific trace variant contributes to the variety. Third, they combine different variety perspectives, which can be contradictory. In general, one can distinguish between three different variety perspectives [4], [15], [23]:

(P1) the number of variants,
(P2) the relative frequency of variants, and
(P3) the differences among variants based on their properties.

Depending on the perspective, the evaluation of variety can lead to different results. This becomes apparent when comparing the two sets of process traces (sequences of activities): $L_1 = \{\langle A, B, C\rangle, \langle A, B, C\rangle, \langle A, B, C\rangle, \langle A, C, B, B\rangle, \langle X, B, Y\rangle\}$, and $L_2 = \{\langle A, B, C\rangle, \langle A, B, B, C\rangle, \langle A, B, B, B, C\rangle\}$. According to (P1) the two sets are identical in terms of the number of unique trace variants (both contain three). According to (P2) the variety of L_2 might be considered higher than the one of L_1, since the relative frequency of the trace variants is more equally distributed in L_2, while in L_1 the trace variant $\langle A, B, C\rangle$ is dominating the others. On the contrary, according to (P3) the variety of L_1 might be considered higher than the one of L_2, since the structural differences among the trace variants regarding the number and order of activities are higher in L_1.

While several process variety measures have been proposed in the literature, which focus on the variety based on relative frequency (P2) or a combined form of the three perspectives [1], [2], there only exist a few measures for the analysis of process variety based on the comparison of structural differences among trace variants (P3) [16]. Additionally, the measures proposed to evaluate (P3) have mathematical limitations and cannot provide a meaningful ranking of trace variants.

Based on the foundational work on variety measures by Weitzman [28], and Nehring and Puppe [15] we propose a novel measurement framework, which is capable of measuring the structural variety of a set of process traces, i.e., a log file, based on the pairwise comparison of all contained trace variants (P3). The framework guarantees the uniqueness of the measured variety value and provides a corresponding ranking of the trace variants.

We evaluate this framework in two ways. First, we use artificial process data to show that the calculated structural variety correctly reflects the structural properties of the trace variants within a log file. The framework thereby outperforms alternative variety measures in terms of consistency and correctness. Second, we use real business process data to show that the structural variety is directly related to the flexibility of a process, i.e., the ability of a process to adjust its structure to changes, which is also known as flexibility-by-change [17]. This is particularly useful for process standardization, where the goal is to reduce the number of trace variants, while simultaneously preserving the flexibility of the process. We show that in this regard process standardization can be improved by selecting trace variants based on structural variety, rather than a frequency-based selection.

Overall, the research contribution of this paper is twofold. *First*, we propose a measurement framework to evaluate structural process variety and to rank trace variants accordingly. *Second*, we show the dependency between structural process variety and flexibility-by-change, with major implications for process standardization efforts.

In the following sections, we first introduce basic terminology and some background on variety and process flexibility (Sect. II), then we introduce the framework to measure structural process variety (Sect. III), followed by an extensive evaluation (Sect. IV). Finally, we identify the related work (Sect. V), followed by a discussion of the findings (Sect. VI), and a short conclusion (Sect. VII).

II. BACKGROUD

A. Trace, log, trace similarity

In this paper we rely on the simple trace and event log notion, considering only activities and no other process perspectives [26]. Let $\mathcal{A}$ denote a finite set of activity names and $\mathcal{A}^*$ the set of all finite sequences over $\mathcal{A}$. A *trace* $\sigma \in \mathcal{A}^*$ corresponds to a finite, non-empty sequence of activities. A log L is a multi-set of traces over $\mathcal{A}^*$, e.g., $L_1 = [\langle a,b,b,c\rangle^3, \langle a,b,c\rangle^1]$. $|L| = n$ thereby denotes the number of traces within L. Traces with an identical sequence of activities belong to the same *trace variant* [24]. The *structure* of a trace variant is defined by its set of activities and their order, resulting in structural patterns, such as a particular ordering, or a repetition of activities.

Analyzing the structural variety of a log requires a comparison of the contained trace variants. For this purpose, we will disregard the multiplicity of the traces and only consider the flattened log $f_{flat}(L) = L'$ [2], such that $f_{flat}([\sigma_1^{n_1}, \dots, \sigma_m^{n_m}]) = [\sigma_1, \dots, \sigma_m]$ with $|L'| = m$. Finally, $d : L' \times L' \rightarrow \mathbb{R}_0^+$ denotes a distance function, defining a pairwise distance between all $\sigma \in L'$. A distance function d is considered to be a metric, if for all $\sigma_i, \sigma_j, \sigma_k \in L'$ the following axioms hold:

A1: $d(\sigma_i, \sigma_j) \geq 0$ (non-negativity)
A2: $d(\sigma_i, \sigma_j) = d(\sigma_j, \sigma_i)$ (symmetry)
A3: $d(\sigma_i, \sigma_j) = 0 \Leftrightarrow \sigma_i = \sigma_j$ (identity of indiscernibles)
A4: $d(\sigma_i, \sigma_k) \leq d(\sigma_i, \sigma_j) + d(\sigma_j, \sigma_k)$ (triangle-inequality).

B. Variety

Variety is closely related to the concept of diversity. Diversity is generally considered an important system property across many domains, such as ecology and economics [4], [23]. It thereby refers to the overall dissimilarity of the objects within a system, following the intuitive notion: "the more dissimilar objects are among each other, the more diverse is their totality" [15].

However, we use the distinction in terminology to distinguish between the diversity among object types (e.g., species, product types, and process types) and the variety among instances of an object type (e.g., individuals of a species, products of a product type, and executions of a process type). Nevertheless, diversity and variety can be quantified based on the same measures. We define a process variety measure as follows:

$$V : L \rightarrow \mathbb{R}_0^+ \tag{1}$$

For each of the three variety perspectives (P1-P3) there exists a respective measurement approach [4], [23].

(I) The number of variants (P1) can be derived by counting the number of trace variants within a log:

$$V^C(L) = |L'| \tag{2}$$

Although this simple measure can be calculated efficiently and is also easy to interpret, it ignores the relative frequency and the structural differences among the variants.

(II) Variety measures based on the relative frequencies of variants (P2) can be subsumed under the generalized function [8]:

$$V^E(L) = (\sum_{i=1}^{n} p_i^{\alpha})^{\frac{1}{1-\alpha}} \quad \text{with} \quad \alpha \geq 0 \tag{3}$$

where $p_i = \frac{|\sigma_i|}{n}$ equals the relative frequency of a trace variant $\sigma_i \in L'$, and the parameter α determines how much weight the diversity measure places on the relative frequency of the variants. For $\alpha = 0$, the equation yields (2), the variant count. In this case, the relative frequency does not matter. A common alternative is the Shannon-Wiener entropy [23], which can be obtained from (3) by setting $\alpha = 1$ (a proof is provided by Hill [8]):

$$V^{EH}(L) = e^H \quad \text{with} \quad H = -\sum_{i=1}^{n} p_i \ln p_i \tag{4}$$

While entropy has also been proposed for the analysis of process variety [1], [2], it has several shortcomings. First, it is difficult to find a substantive interpretation of the parameter α [15]. Since (4) with $\alpha = 1$ originates from encoding theory, the interpretation of this measure in terms of process variety analysis has not been discussed in the literature [1], [2]. Second, the measure does not provide a ranking for the trace variants within a set. Third, it does not reflect the structural differences among the trace variants.

(III) Finally, we introduce a variety measure based on the structural differences of the variants (P3). Once, a pairwise distance d is calculated for all trace variants in L', the distance between a trace variant $\sigma_i \in L'$ and a subset of trace variants $Q \subset L'$ can be derived as follows:

$$\delta(\sigma_i, Q) = \min_{\sigma_j \in Q} d(\sigma_i, \sigma_j) \text{ for all } \sigma_i \in L' \setminus Q. \tag{5}$$

This distance indicates the difference in variety between Q and the enlarged set $Q \cup \{\sigma_i\}$. Consequently, the overall structural variety of L' can be derived based on (5) by iteratively adding up distances. To achieve this, Weitzman [28] proposes a recursive formula, starting with an arbitrary trace $\sigma_0 \in L'$ and an arbitrary start value $V^W(\{\sigma_0\})$:

$$\begin{gathered} V^W(Q \cup \{\sigma_i\}) = V^W(Q) + \delta(\sigma_i, Q) \\ \text{for all } \sigma_i \in L' \setminus Q \text{ and } \varnothing \subset Q \subset L'. \end{gathered} \tag{6}$$

In the context of process analysis, it would be plausible to assume that the start value for $V^W(\{\sigma_0\}) = 0$, since there exists no variety if L' consists of a single trace variant. The variety value based on (6) would then incrementally increase by adding the remaining traces $\sigma_i \in L' \setminus Q$ to Q. The variety thereby increases each time based on (5).

One important issue with this recursive approach is that, in general, the outcome of (6) is not unique, i.e., the variety value can differ based on the selection of σ_0. Weitzman [28] shows that this issue can only be overcome if the applied distance function d is an *ultrametric*, i.e., a metric with the additional property that the two greatest distances among three points are always equal, i.e:

A5: $d(\sigma_i, \sigma_k) \leq max\{d(\sigma_i, \sigma_j), d(\sigma_j, \sigma_k)\}$ (ultram.-inequ.).

This is a stronger condition than the triangle-inequality (A4) and restricts the evaluation of structural process variety significantly. Indeed, the work by Nehring and Puppe [15] shows that this restriction generally applies when measuring variety based on the properties of variants (P3).

However, a set of traces L' characterized by ultrametric distances results in a tree structure, which induces a hierarchical ordering between the traces (as depicted in Fig. 3 (3)). Such a tree structure can be derived based on hierarchical trace clustering [9], [22]. Hierarchical clustering thereby guarantees the ultrametric distance between the clustered traces (a proof is provided by Johnson [10]). Hence, to derive a unique structural variety value for a set of trace variants L' requires first to group the traces based on hierarchical clustering and then to derive the variety based on the (ultrametric) distances between the clusters. By applying the recursive formula (6), we additionally ensure that the variety measure is monotonously increasing with respect to an increasing number of trace variants and an increase in structural differences [28], i.e.:

A6: $V(L') < V(L' \cup \sigma_i)$, where $\sigma_i \notin L'$ (monotonicity).

C. Structural Process Variety and Standardization

Structural variety in business processes (BP) is commonly associated with inefficiency and high error probability [14], [16], [25]. At the same time, structural variety is also an indicator for the flexibility of a process and its ability to adapt to changes [16], [25]. BP standardization, a common BP practice, therefore seeks to define standardized processes, which provide a balance between efficiency and flexibility [7], [14], [25]. One way to achieve this goal is to design a standardized process in such a way, that the restricted number of execution paths allow for a high *flexibility-by-change* [17]. Flexibility-by-change assumes that not every process behavior can be anticipated before the execution. Hence, a process should be designed in such a way that its structure can be easily adapted during run-time, without imposing high adaptation costs.

The observed trace variants documented in a log file can serve as a source for such a standardization approach (a similar approach is suggested by Fleig et al. [7]). First, to define a standardized process, a process designer can simply restrict the number of trace variants by selecting a subset $S \subset L'$ according to some selection criteria. Second, to evaluate the flexibility-by-change of the subset S, one can calculate the trace alignment costs [9] that occur due to the removal of the traces $\sigma_i \in L' \setminus S$. Optimal trace alignment is a common technique in process mining to calculate the costs for the adaptation operations that are necessary to transform one trace into another [9]. The alignment costs are thereby calculated based on a scoring function considering deletions and insertions (gaps), mismatches, and matches of activities.

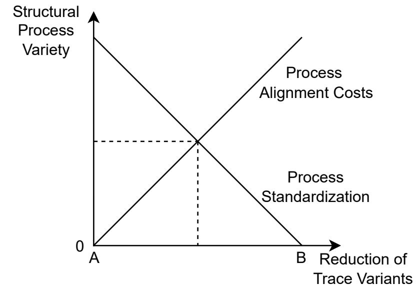

Fig. 1. Trade-off between process standardization and alignment costs.

The trade-off between process standardization and process alignment costs for a given set of trace variants L' is depicted in Fig. 1. When there is no process standardization, i.e., the reduction of trace variants from L' is zero (state A), the structural process variety of L' is at its maximum value (cf. Sect. II-B). Assuming that the process execution is restricted to trace variants contained in L', the non-standardized process does not require any adaptations, i.e., the process alignment costs are zero. However, the more trace variants are removed from L', due to process standardization efforts, the higher are the process alignment costs for the standardized process $S \subset L'$, when the removed variants are eventually executed. The alignment costs reach their highest value, when the reduction of trace variants reaches its maximum and the structural variety of the standardized process S is zero (state B).

A structural variety assessment can support the design of standardized BP threefold. First, it provides an overall evaluation of the process variety based on L', which can serve as a proxy for flexibility-by-change. Second, it can show how individual trace variants contribute to the overall variety and how this correlates with the increase in adaptation costs. Third, this allows to select trace variants in such a way that the balance between efficiency (low number of trace variants) and flexibility-by-change (high variety) can be optimized.

III. Structural Variety Measurement Framework

For the measurement of structural process variety we propose a framework as depicted in Fig. 2. In the following we will elaborate on each step of the framework.

A. Preprocessing

This step involves, in particular, the flattening of the log $f_{flat}(L) = L'$. But it might also involve the removal of outliers and incomplete traces, which might not be considered relevant for the structural variety evaluation.

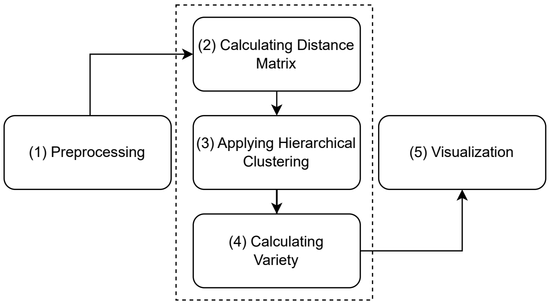

Fig. 2. Proposed framework for the measurement of structural process variety.

B. Calculating Distance Matrix

This step involves the selection and application of a distance function d to calculate the pairwise distances for all $\sigma_i, \sigma_j \in L'$. The selection of the distance function thereby has a significant impact on the variety evaluation. Back and Simonsen [3] propose a set of evaluation measures, which can help to select a fitting distance function for a specific log. However, the evaluation relies on the assumption, that the traces in L' can be clearly distinguished based on class labels, and that the distances among traces within a class are lower than the distances among traces belonging to different classes. In practice, such a distinction based on class labels is not always possible. In this paper, we therefore rely on the Levenshtein distance (LD), which is commonly applied in process analysis to evaluate the distance between traces [3], [6], [16]. Similarly to optimal trace alignment (cf. Sect. II-C) the LD is calculated based on the minimal number of insertions, deletions, and substitutions of activities needed to transform one trace into another. It is indeed the case that the LD and the calculation of the optimal alignment between two traces yield equivalent outcomes (cf. for a proof Sellers [21]). This is in particularly important when analyzing the relationship between structural process variety based on LD and flexibility-by-change based on alignment costs.

Fig. 3 (1-2) provides an example for the calculation of the LD for all $\sigma_i, \sigma_j \in L_0$ resulting in the distance matrix D_{L_0}.

C. Applying Agglomerative Hierarchical Clustering

By applying agglomerative hierarchical clustering (AHC) based on the derived distance matrix, we derive ultrametric distances between the clustered traces (A5) (cf. Sect. II-B). AHC was proposed in the research literature [9], [22] as a clustering approach for traces, since it well reflects hierarchical properties of a process. This hierarchy naturally occurs, since a process often possesses global structural properties (such as a start and end sequences), which apply to a larger subset of traces, as well as local structural properties (such as reoccurring activities), which apply to a smaller subset of traces [9], [22].

The AHC gradually merges traces into clusters based on minimal distances, starting with each trace belonging to an individual cluster [13]. The distance between the clusters is calculated based on a linkage criterion. In this work, we apply the commonly used centroid linkage, which measures the distance between two clusters based on the Euclidean distance between the centroids (the arithmetic mean) of the clusters [13].

L_0
$\sigma_1 = \langle A,B,C,D,E,F,G,H \rangle$
$\sigma_2 = \langle A,B,C,E,F,G,H,D,I \rangle$
$\sigma_3 = \langle A,B,C,E,F,G,H,D \rangle$
$\sigma_4 = \langle A,B,C,E,D,F,G,H \rangle$
$\sigma_5 = \langle A,B,C,E,F,D,G,H,I \rangle$

(1)

$$D_{L_0} = \begin{bmatrix} 0 & 3 & 2 & 2 & 3 \\ 3 & 0 & 1 & 3 & 2 \\ 2 & 1 & 0 & 2 & 2 \\ 2 & 3 & 2 & 0 & 3 \\ 3 & 2 & 2 & 3 & 0 \end{bmatrix}$$

(2)

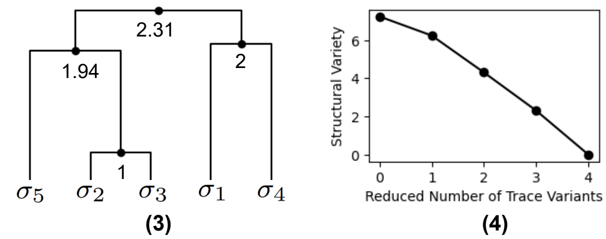

Fig. 3. An example for the application of the proposed measurement framework. (1) depicts the log L_0, (2) shows the respective distance matrix for L_0 based on LD, (3) shows the dendrogram derived from D_{L_0} based on AHC, and (4) shows a visualization of the ranking of the trace variants. L_0 and the clustering are based on an example provided in [9].

When applying AHC based on D_{L_0} (Fig. 3 (2)), each of the five trace variants is considered as an individual cluster in the beginning. Subsequently the following clusters are derived successively based on the centroid linkage: $\{\sigma_2, \sigma_3\}$, $\{\sigma_2, \sigma_3, \sigma_5\}$, $\{\sigma_1, \sigma_4\}$, and $\{\sigma_2, \sigma_3, \sigma_5, \sigma_1, \sigma_4\}$. The respective dendrogram in Fig. 3 (3) depicts the hierarchical clusters, as well as their calculated distances.

D. Calculating Variety

Based on the derived distances between the clustered traces a unique structural variety value can be calculated by applying the recursive formula V^W (6), independent from the selection of the initial trace. Starting with σ_2 from the example in Fig. 3 (1) the structural variety for L_0 can be calculated as follows: $V^W(L_0) = 0 + 1 + 1.94 + 2.31 + 2 = 7.25$.

E. Visualization

Following V^W (6) it is further possible to generate a ranking of the trace variants according to their marginal contributions to the overall structural variety of a log [28]. Ordering such a ranking from the least to the highest marginal contribution can be visualized as a decreasing structural variety curve. This curve shows not only how the variety decreases according to the derived ranking, but also how the structural variety is distributed among the trace variants within a log. If each trace variant equally contributes to the variety, the decrease in variety is linear. If the contributions are more unequally distributed, the decrease in variety has a concave shape.

In the example in Fig. 3 (4), all trace variants contribute rather equally to the log's structural variety, therefore the depicted variety curve follows a rather linear shape. The structural variety values are derived by subsequently removing trace variants from L_0 in the following order: σ_2, σ_3, σ_1, and σ_4. It would be of course also possible to remove the variants in a different order, yielding the same structural variety curve.

IV. EVALUATION

To evaluate the proposed measurement framework we conduct a comprehensive evaluation consisting of two parts. First, we consider artificial log files with different levels of structural variety (low, medium, high). In this way, we can compare the outcome of the framework with alternative variety measures proposed in the research literature and determine, whether the measures correctly evaluate structural variety across different logs. The logs thereby differ based on the number of trace variants and the structural properties they contain, i.e., the ordering, the reocurrence, and the skipping of activities. A similar evaluation approach was also applied by Back et al. [2]. Second, we consider four real-life logs to show that the proposed framework can be used to rank trace variants based on their marginal contribution to the overall log's variety. Furthermore, we evaluate whether the reduction of trace variants according to the derived ranking can indeed provide guidance for process standardization (cf. Sect. II-C). Here, we compare the impact of the step-wise reduction of trace variants on the increase of alignment costs following (a) the ranking derived by the framework, and (b) a frequency-based ranking.. The alignment costs are thereby calculated based on the optimal alignments between the removed trace variants and the remaining trace variants in the log. Overall, the evaluation seeks to answer the following two questions:

- Can the proposed measurement framework consistently evaluate the structural variety across different logs?
- Can the proposed measurement framework rank the trace variants to provide guidance for process standardization?

The data and implementations for the evaluation are available via the following repository: https://github.com/promilab/ProcessVarietyRepo.git. The implementations are Python based and use the pm4py library [5].

A. Structural Variety Comparison

Process Data. For the first part of the evaluation we consider seven artificial log files. The first three logs are depicted in Table I. Each trace variant in these logs consists of five activities, including a fixed start and a fixed end activity. The logs differ based on the following properties: in L_1 the activities B, C, D occur in any possible order, in L_2 the activities B, C, D occur in a fixed order, but in each trace variant at least one activity has to occur twice, and in L_3 the activities B, C, D occur in a fixed order, but in each trace variant at least one activity has to be skipped. In this way the three logs represent three different types of structural properties, which can be considered as relevant for a structural variety evaluation, i.e., the ordering, the recurrence, and the skipping of activities [2], [16], [24]. By combining the logs, we derive three additional logs with medium structural variety: $L_4 = L_1 \cup L_2$, $L_5 = L_2 \cup L_3$, $L_6 = L_3 \cup L_1$, and one log with high structural variety: $L_7 = L_1 \cup L_2 \cup L_3$.

Evaluation Method. The calculation of V^W for the seven logs with low, medium and high structural variety allows to evaluate the consistency of the proposed framework with respect to monotonicity (A6). Additionally, we compare the derived results with five other variety measures, proposed in the literature. The first four measures are based on entropy and involve different types of trace abstractions, the fifth measure simply takes the average LD over all traces in a log:

- V^{EN} – entropy based on n-blocks (n=2) [2],
- V^{EL} – entropy based on Lempel Ziv complexity [2],
- V^{EP} – entropy based on prefixes [2],
- V^{EK} – entropy based on kNN and norm. LD (k=1) [2],
- V^{A} – average LD [16].

Table I. Three artificial logs with distinct structural properties.

L_1	L_2	L_3
$\langle A, B, C, D, E \rangle$	$\langle A, B, B, C, D, E \rangle$	$\langle A, C, D, E \rangle$
$\langle A, B, D, C, E \rangle$	$\langle A, B, C, C, D, E \rangle$	$\langle A, B, D, E \rangle$
$\langle A, C, B, D, E \rangle$	$\langle A, B, C, D, D, E \rangle$	$\langle A, B, C, E \rangle$
$\langle A, D, B, C, E \rangle$	$\langle A, B, C, B, C, D, E \rangle$	$\langle A, D, E \rangle$
$\langle A, C, D, B, E \rangle$	$\langle A, B, C, D, C, D, E \rangle$	$\langle A, B, E \rangle$
$\langle A, D, C, B, E \rangle$	$\langle A, B, C, D, B, C, D, E \rangle$	$\langle A, E \rangle$

Results. The results of the calculated variety values can be found in Table II. It shows that the structural variety measure V^W based on the proposed framework is the only measure, which consistently evaluates the variety of the logs based on the increase in the number of trace variants and structural properties. All other measures fail to consistently evaluate the variety across the logs, due to their sensitivity towards the relative frequencies of the structural properties, i.e., some of the structural properties such as the relative ordering of the activities in L_2 and L_3 repeat more often than others. Additionally, the variety values based on V^W can be more easily interpreted than the other derived values, i.e., one can detect a clear distinction between low (L_1, L_2, L_3), medium (L_4, L_5, L_6) and high structural variety (L_7). It should also be pointed out, that V^W is the only measure providing a ranking of the trace variants according to their marginal contribution to the log's overall variety (cf. Sect. III-E).

Table II. Comparison of log variety measures (M.) based on seven different logs with low (L_1, L_2, L_3), medium (L_4, L_5, L_6), and high structural variety (L_7). Monotonicity violations according to (A6) are marked in bold.

M.	L1	L2	L3	L4	L5	L6	L7
V^{EN}	3,59	2,76	3,13	**3,47**	3,21	3,60	**3,53**
V^{EL}	2,29	1,86	2,16	**2,01**	**2,07**	**2,15**	**2,09**
V^{EP}	4,19	4,14	3,31	4,78	4,38	4,24	**4,78**
V^{EK}	2,15	1,05	1,72	**2,13**	2,08	2,28	2,50
V^{A}	1,67	1,17	1,17	2,00	2,25	1,71	**2,20**
V^{W}	8,50	5,68	5,68	15,74	14,58	14,17	22,93

B. Structural Variety in Relation to Alignment Costs

Process Data. To evaluate the framework in the context of process standardization we consider four real-life log files:

- WABO 2014, environmental permit application process (4.245 trace variants, 465 activities),
- BPI 2015, building permit application process across 4 dutch municipalities (4.209 trace variants, 29 activities),
- BPI 2017, loan application process of a Dutch financial institute (15.930 trace variants, 26 activities),
- PO 2019, a purchase order handling process (10.207 trace variants, 36 activities).

The logs are selected due to their differences in size and complexity. Furthermore, the processes represent examples in a business context, where process standardization could potentially be applied. All four logs are available at the 4TU research data repository: https://data.4tu.nl/.

Evaluation Method. To evaluate to which extend our framework can improve process standardization, we calculate the ranking of the trace variants according to their marginal contribution to the log's structural variety for the four real-life log files. We then incrementally remove the trace variants from the logs, based on the derived ranking (starting with the trace variant that contributes the least to the log's variety). When trace variants have equal variety values, one of them is randomly selected. Additionally, we calculate after each removed trace the optimal alignment costs between all remaining trace variants and all removed trace variants. The alignment costs are thereby an approximation of the flexibility-by-change (cf. Sect. II-C), i.e., the lower the increase in alignment costs, the higher is the flexibility-by-change of the standardized process, consisting of the remaining trace variants. Assuming that flexibility-by-change can be approximated based on the number of mimimal changes necessary to align the removed trace variants with the remaining once, we calculate the optimal alignment with costs of 1 for gaps and mismatches, and rewards of 0 for matches. We compare the reduction of the trace variants according to the framework with a frequency-based reduction of trace variants as a baseline. In the baseline the variants are ranked according to their frequency, starting with the least frequent one, which can be considered as valid practice to improve BP efficiency [7], [25]. Trace variants, which occur equally often are removed in a random order.

To compare the two selection methods in a quantitative way, we apply two different evaluation measures: (a) the approximate integral over the alignment cost curve based on the trapezoid rule, i.e., area under the curve (AUC), and (b) the point of intersection (POI) between the alignment cost curve and the variety curve. The AUC is the most important evaluation measure, since it reflects the increase of the alignment costs based on the reduced trace variants. The lower the AUC, the slower the increase in alignment costs, which is a strong indicator for the improvement of process standardization. The POI can be considered as the equilibrium between standardization and alignment costs (cf. Sect. II-C). The higher the POI, the more trace variants are reduced, again indicating an improvement for process standardization.

Results. The derived structural variety and alignment cost curves based on the framework (a) and based on the frequency-based selection (b) are depicted in Fig. 4. The variety curves based on the framework are for all four logs more concave, indicating that the structural variety decreases more slowly in relation to the reduction of trace variants.

The results of the evaluation measures are depicted in table III. The reduction of trace variants based on the framework yields better results for the AUC and the POI across all four logs, when compared to the frequency-based reduction.

Table III. Comparison of the two trace variant selection approaches based on (a) structural variety and (b) frequency.

Log file	Selection Approach	AUC	POI
WABO 2014	Structural Variety	941,44	3406
	Frequency	1296,23	2698
BPI 2015	Structural Variety	891,98	3377
	Frequency	1278,63	2565
BPI 2017	Structural Variety	2879,23	13516
	Frequency	4416,32	10356
PO 2019	Structural Variety	1698,93	9295
	Frequency	1858,98	7300

V. Related Work

The work by Pentland [16] is one of the first to propose structural process variety measures based on trace variants. Although the author uses the term sequential variety instead of structural variety. Three measures are proposed: average LD, algorithmic complexity, and deviations from random. The average edit distance is the closest related measurement approach to our proposed framework. Even though the measure ensures a unique outcome, it does not consistently reflect a log's structural variety, as we could show in the evaluation (cf. Sect. IV-A). The idea of the algorithmic complexity measure is to apply a compression algorithm, which can detect reoccurring patterns in a sequence. In order to apply this measure, all traces need to be represented by a single sequence. The higher the number of reoccurring patterns in this sequence, the lower its variety. The third proposed measure, deviation from random, compares sequences of activities to a uniform, random Markov process. The higher the deviation, the less random the relative ordering of activities in the sequences and the lower the variety. Both measures, i.e., algorithmic complexity and deviations from random, are not included in the evaluation (cf. Sect. IV-A), since they do not consider the structural differences between traces.

A comprehensive overview of potential process variety measures is provided by Back et al. [2]. Although the authors acknowledge the relevance of structural trace properties for the variety evaluation, all identified measures are based on entropy and are therefore sensitive to the relative frequency of activities and structural patterns. Due to this bias, the proposed measures are not able to consistently evaluate structural variety and are also difficult to interpret (cf. evaluation Sect. IV-A). In a similar vein, Augusto et al. [1] additionally investigate mathematical properties of entropy measures based on prefixes, i.e., sub-traces of traces, which contain all past activities up to a point. One disadvantage of this approach is however, that it is backwards-looking and fails to account for structural similarity that might occur after shared prefixes [2].

Other related measures, which also consider structural variety, are rather focused on the distinctiveness between logs, e.g., for process data sampling [11], and do not provide an overall evaluation of a single log's structural variety. Additionally, none of the existing measures in the literature enables an ordering of traces according to their marginal contribution to structural variety.

Finally, also the purpose of the evaluation of structural vari-

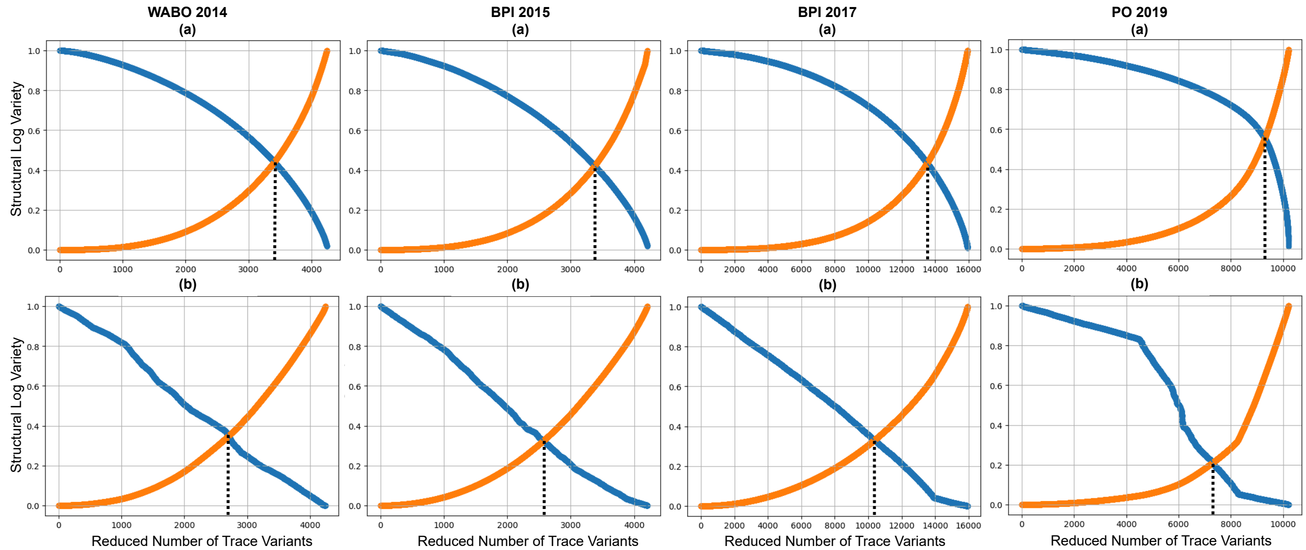

Fig. 4. The structural variety curve (blue) and the alignment costs curve (orange) in relation to the reduced number of trace variants (x-axis) and the structural log variety (y-axis). The diagrams show the results for the evaluation based on four different real-life logs, when ranking the trace variants based on (a) structural variety and (b) relative frequency.

ety differs in the research literature. While the works by Back et al. [2] and Augusto et al. [1] investigate variety measures to evaluate the suitability of process mining algorithms for a given log, the work by Pentland [16] considers process variety for the purpose of process improvement, which is closely related to our own work.

The correlation between process variety and process performance was later confirmed based on an empirical study by Vidgof et al. [27]. Consequently, Wurm et al. [29] confirmed based on expert interviews, that the structural properties of a process, e.g., structural variety, are an important factor for process standardization. Based on these insights Fleig et al. [7] propose a method, which supports process standardization efforts guided by process mining. This method, however, relies on a repository of predefined best-practice process standards and does not involve variety considerations.

Furthermore, our work builds upon our initial proposal to use variety measures for business process performance analysis [18], where we outline different potential performance implications of variety in process executions.

VI. Discussion

The evaluation in Sect. IV has shown, that the proposed framework can measure structural variety in a consistent way (Sect. IV-A), and can provide valuable guidance for process standardization efforts by ranking trace variants according to their marginal contribution to the overall structural variety of a log (Sect. IV-B). The framework could therefore also be applied to provide guidance for the sampling of process traces [11], to evaluate how well a discovered process model reflects the structural variety of a log [1], as well as to analyze process performance [16], [27], and process change [6].

Nevertheless, there are some limitations of the framework, that need to be addressed. First of all, there still remains some uncertainty regarding the derived variety value due to the selection of the distance measure and the clustering algorithm. The clustering only allows for an approximation of the variety. However, evaluation measures can help validate the outcome of the clustering [3] and provide more certainty regarding the variety evaluation.

A similar uncertainty remains regarding the ranking of trace variants. Due to the ultrametric property, the two greatest distances between three trace variants are always equal (cf. Sect. II-B). Hence, even for a small set of three trace variants, there are multiple ways how to rank the variants. Our framework can therefore only guarantee a unique structural variety curve (cf. Sect. IV-B), but not a unique ranking. In practice, this could be overcome by involving experts in the ranking decision, who possess additional insights on the trace variants beyond their distance values.

A third shortcoming, which is at the same time a strength (cf. Sect. I), is that the framework does not consider the relative frequency of traces and activities for the evaluation of variety, but only focuses on the process structure. However, structural variety is especially relevant for BPs, where flexibility plays a more important role than frequency, such as in service work [16]. In addition, the real-life logs used in the evaluation (Sect. IV-B) mainly contain infrequent behavior, i.e., the share of trace variants only occurring once is in all event logs greater than 80% (WABO 2014 - 99%, BPI 2015 - 92%, BPI 2017 - 87%, PO 2019 - 81%). Nevertheless, for some application scenarios the frequency can be of importance, such as for process data sampling [11]. If certain structural properties are more common than others, then this should also be reflected in the sampling of the traces. One possibility to integrate the frequencies into the framework is to adjust the calculation of the pairwise distances (cf. Sect. III-B), such that the distance would increase with the frequency of required edits. However, to avoid the mix up of the different variety perspectives (cf. Sect. I) it would be better to consider them separately based on multiple measures.

A further limitation is that we so far only considered structural properties of the traces based on their control-flow, thereby neglecting other process aspects such as resources or data. In the future, these additional perspectives could also be integrated in the framework by adjusting the notion of a trace variant, as well as the distance function. A similar approach can be found in the work by Kabierski et al. [11].

With respect to the application of the framework as guidance for process standardization, it is also important to notice that we assume that there is no interdependence between the trace variants in terms of executability. In reality, this is of course not necessarily the case. Hence, trace variants cannot simply be removed according to some variety evaluation. Here, it would be interesting to investigate in the future how a variety measure could deal with such interdependence among traces. Furthermore, a qualitative evaluation of the framework would be necessary to further validate its applicability for process standardization in practice.

Finally, the guidance for process standardization is so far only backward-looking, i.e., based on the already observed process variants. But the approach could also be easily applied in future-looking scenarios, e.g., based on simulation [12]. In this way, a process designer could evaluate the adaptation costs of a standardized process based on alternative trace variants, which are not already contained in a log.

VII. Conclusion and Future Work

In this work, we propose a measurement framework, which allows to measure a log's structural variety based on the pairwise comparison of its trace variants. The framework relies on the measurement approach proposed by Weitzman [28] and guarantees a unique and meaningful outcome. It further allows to rank trace variants according to their marginal contribution to the log's structural variety. In Sect. IV we could show that the framework outperforms existing measures, when it comes to the evaluation of structural variety, and that it can significantly improve process standardization efforts. For the latter, we tested four real-life event logs, and showed that the removal of trace variants based on our framework increased the flexibility-by-change of the standardized process in comparison to a frequency-based selection.

In the future we plan to apply this framework in simulation scenarios, to also test process changes, which have not occurred in the past. In this way, processes could be further adjusted in order to become even more flexible towards changes. We further plan to apply the framework to analyze behavioral process data, such as eye-tracking data. This could help to better understand changes in human cognitive behavior, and to develop better user-oriented tool support [19], [20].

References

[1] Adriano Augusto, Jan Mendling, Maxim Vidgof, and Bastian Wurm. The connection between process complexity of event sequences and models discovered by process mining. *Information Sciences*, 598, 2022.

[2] Christoffer Olling Back, Søren Debois, and Tijs Slaats. Entropy as a Measure of Log Variability. *Journal on Data Semantics*, 8(2), 2019.

[3] Christoffer Olling Back and Jakob Grue Simonsen. Comparing trace similarity metrics across logs and evaluation measures. In *CAiSE*, pages 226–242. Springer, 2023.

[4] Stefan Baumgärtner. Measuring the diversity of what? And for what purpose? A conceptual comparison of ecological and economic biodiversity indices. *Available at: https://ssrn.com/abstract=894782*, 2006.

[5] Alessandro Berti, Sebastiaan van Zelst, and Daniel Schuster. Pm4py: a process mining library for python. *Software Impacts*, 17:100556, 2023.

[6] RP Jagadeesh Chandra Bose, Wil MP van der Aalst, Indrė Žliobaitė, and Mykola Pechenizkiy. Handling concept drift in process mining. In *CAiSE*, pages 391–405. Springer, 2011.

[7] Christian Fleig, Dominik Augenstein, and Alexander Maedche. Designing a process mining-enabled decision support system for business process standardization in erp implementation projects. In *BPM Forum*, pages 228–244. Springer, 2018.

[8] Mark O Hill. Diversity and evenness: a unifying notation and its consequences. *Ecology*, 54(2):427–432, 1973.

[9] RP Jagadeesh Chandra Bose and Wil van der Aalst. Trace alignment in process mining: opportunities for process diagnostics. In *BPM*, pages 227–242. Springer, 2010.

[10] Stephen C Johnson. Hierarchical clustering schemes. *Psychometrika*, 32(3):241–254, 1967.

[11] Martin Kabierski, Hoang Lam Nguyen, Lars Grunske, and Matthias Weidlich. Sampling what matters: relevance-guided sampling of event logs. In *ICPM*, pages 64–71. IEEE, 2021.

[12] Nina Kinkelin, Clemens Schreiber, and Josua Reimold. Combining process mining and simulation to guide users towards process improvement in mpmx. In *ICPM Doctoral Consortium/Demo*, 2023.

[13] Sadaaki Miyamoto. *Theory of agglomerative hierarchical clustering*, volume 15. Springer, 2022.

[14] Björn Münstermann, Andreas Eckhardt, and Tim Weitzel. The performance impact of business process standardization: An empirical evaluation of the recruitment process. *BPMJ*, 16(1):29–56, 2010.

[15] Klaus Nehring and Clemens Puppe. Diversity. *The handbook of rational and social choice*, pages 298–322, 2009.

[16] Brian T. Pentland. Conceptualizing and measuring variety in the execution of organizational work processes. *Management Science*, 49(7):857–870, 2003.

[17] Helen Schonenberg, Ronny Mans, Nick Russell, Nataliya Mulyar, and Wil MP van der Aalst. Towards a taxonomy of process flexibility. In *CAiSE Forum*, volume 344, pages 81–84, 2008.

[18] Clemens Schreiber. Exploring the impact of process diversity on business process performance. In *ICPM Doctoral Consortium/Demo*, pages 17–18, 2021.

[19] Clemens Schreiber, Amine Abbad-Andaloussi, and Barbara Weber. On the cognitive effects of abstraction and fragmentation in modularized process models. In *BPM*, pages 359–376. Springer, 2023.

[20] Clemens Schreiber, Amine Abbad-Andaloussi, and Barbara Weber. On the cognitive and behavioral effects of abstraction and fragmentation in modularized process models. *Information Systems*, 125:102424, 2024.

[21] Peter H Sellers. On the theory and computation of evolutionary distances. *SIAM Journal on Applied Mathematics*, 26(4):787–793, 1974.

[22] Minseok Song, Christian W Günther, and Wil MP Van der Aalst. Trace clustering in process mining. In *BPM Workshops*, pages 109–120. Springer, 2009.

[23] Andy Stirling. A general framework for analysing diversity in science, technology and society. *Journal of the Royal Society Interface*, 4(15):707–719, 2007.

[24] Farbod Taymouri, Marcello La Rosa, Marlon Dumas, and Fabrizio Maria Maggi. Business process variant analysis: Survey and classification. *Knowledge-Based Systems*, 211:106557, 2021.

[25] Roger Tregear. Business process standardization. In *Handbook on business process management 2: Strategic alignment, governance, people and culture*, pages 421–441. Springer, 2014.

[26] Wil van der Aalst. *Process Mining: Data Science in Action*. Springer, 2016.

[27] Maxim Vidgof, Bastian Wurm, and Jan Mendling. The impact of process complexity on process performance: A study using event log data. In *BPM*, pages 413–429. Springer, 2023.

[28] Martin L Weitzman. On diversity. *The quarterly journal of economics*, 107(2):363–405, 1992.

[29] Bastian Wurm, Theresa Schmiedel, Jan Mendling, and Christian Fleig. Development of a measurement scale for business process standardization. In *ECIS*, 2018.