Data-driven Assessment of Business Process Resilience

Alexander Kraus[1*], Jana-Rebecca Rehse[1] and Han van der Aa[2]

[1]University of Mannheim, Germany.
[2]University of Vienna, Austria.

*Corresponding author(s). E-mail(s): alexander.kraus@uni-mannheim.de;
Contributing authors: rehse@uni-mannheim.de;
han.van.der.aa@univie.ac.at;

Abstract

Process resilience represents a core competence for organizations in light of an increasing number of process disruptions, such as sudden increases in case arrivals or absences in the workforce. It reflects an organization's ability to restore a process to its acceptable performance level after a disruption. In this regard, the first key step for organizations towards achieving resilience is to understand how resilient their processes actually are. Although recognized as important, few works focus on such resilience assessment in a data-driven manner, thus barring organizations from gaining the necessary insights into how much their processes are affected by disruptions and how long it takes them to recover. To address this problem, we propose an approach for automated resilience assessment, based on recorded event data. Our approach interprets relevant process characteristics, such as the average lead time or arrival rate, as time series, which capture the development of the process execution over time. Based on these time series, it uses statistical modeling, specifically a vector autoregressive model, to determine the inter-relations between those characteristics and assess how the process performance responds to a disruption, i.e., a significant and temporal change in one of the process characteristics. We validate our approach by comparing its accuracy with a what-if analysis using a simulation model and demonstrate its effectiveness by assessing the resilience of the same process to diverse disruptions across different organizations.

Keywords: Business process resilience, resilience assessment, process mining, vector autoregressive modeling, impulse-response analysis.

1 Introduction

Organizations operate in dynamic environments that are subject to frequent changes. This includes the occurrence of disruptions, such as peak in case arrivals, equipment failure, or absences in the workforce. Such disruptions can have a (temporary) negative impact on process performance. As shown in Figure 1, a process disruption can cause performance to initially worsen, e.g., resulting in an increased lead time, before slowly returning to its pre-disruption level. *Process resilience* refers to the ability of an organizational process to deal with such a disruption and recover from it [1]. As a means to ensure consistent performance despite the occurrence of disruptions, being resilient is thus a core competence for organizations [2]. In this regard, the first key step for organizations towards achieving resilience is to understand how resilient their processes actually are [3], which involves assessing how they respond to different disruptions.

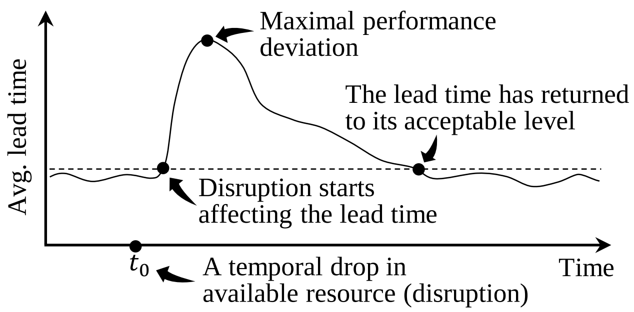


Fig. 1: The impact of a disruption on process performance.

Despite the recognized importance of process resilience and its measurement [2–4], little research actually targets this measurement task, especially not in a data-driven manner. Existing works in this area either provide conceptual characterizations of process resilience [5, 6] and resilient business process management [4], or focus on achieving resilience in a design-time manner [7]. The one exception is a case study by Zahoransky et al. [8], where the authors describe how they assessed process resilience in a specific scenario, based on event data and a known process model. However, they do not propose a generalizable approach that is applicable to other processes. In addition, they strongly depend on domain knowledge (such as having an accurate process model) and manual analysis (for the estimation of key probabilities), which makes their resilience assessment impractical. Hence, the problem of how to properly and conveniently measure the resilience of organizational processes remains unaddressed.

The contribution of this paper is an approach that assesses process resilience in a data-driven and automated manner, based on an event log and a resilience scenario of interest. For a potential disruption, such as a sudden absence of a resource or a peak in cases, our approach quantifies the disruption's impact on process performance according to four resilience measures, which jointly characterize the duration and severity of the disruption's impact. To do this, we establish a statistical model that captures the inter-relations between different process characteristics over time. We then use this model to measure the impact of a potential future disruption on process performance. We evaluate our approach by assessing its accuracy through comparison with a what-if analysis by means of a simulation model. In addition, we demonstrate its effectiveness by assessing the resilience of the same process to diverse disruptions across different organizations.

In the remainder, Section 2 discusses the relevant background information about process resilience and its assessment. Section 3 defines the addressed problem, and Section 4 presents our proposed approach, which we evaluate in Section 5. We consider the advantages and limitations of the proposed solution in Section 6. Finally, Section 7 discusses related work, before Section 8 concludes the paper.

2 Background

In this section, we explore the concept of resilience. We begin by examining resilience in general, considering its definition across various disciplines and existing assessment methods. Next, we delve into the definitions and assessment methods specific to organizational resilience. Finally, we focus on the resilience of business processes as a key layer of organizational resilience, highlighting the current lack of data-driven assessment methods and the need for their development.

Resilience as a concept. The concept of resilience has been discussed across various disciplines [9–11], ranging from psychology to seismology and to material science, as depicted in Table 1. However, a universal definition of resilience has not been established because different disciplines use specific terminology to address their unique needs and challenges. Despite these diverse perspectives, an overarching understanding of resilience can be summarized as the ability of a system to withstand a disruption within acceptable degradation and to recover within a suitable time and reasonable costs [12].

Table 1: Definitions of resilience in different disciplines (based on [9–11])

Discipline	**Definition**
Psychology	The ability of individuals to recover from adversity. Positive ability of individuals to cope with stress and catastrophic events.
Seismology	The ability of the system to reduce the chances of shock, to absorb a shock if it occurs, and to recover quickly after a shock (re-establish normal performance).
Ecology	The magnitude of disturbance that a system can absorb before its structure is redefined by changing the variables and processes that control behavior.
Infrastructure	Ability of infrastructure to reduce the probability of failure, the consequences of such failure, and the response and recovery time.
Material Science	A material's tendency to return to its original form after applying a force or stress that has produced elastic deformation.
Engineering	The ability to sense, recognize, adapt, and absorb variations, changes, disturbances, disruptions, and surprises.
Tourism	Ability of communities (ecosystems) to withstand the impacts of external forces while retaining their integrity and ability to continue functioning.
Networks	The ability of a network to defend against and maintain an acceptable level of service in the presence of challenges.
Society	Capability of a system to maintain its functions and structure in the face of internal and external change and to degrade gracefully when it must.
Economics	Ability to reduce efficiently both the magnitude and duration of deviation from targeted system performance levels given the occurrence of a particular disruptive event.
Sociology	Ability to recover from adversity and become stronger than before.

Resilience assessment methods can be grouped into two categories [10], as visualized in Figure 2. Qualitative approaches involve methods for assessing system resilience

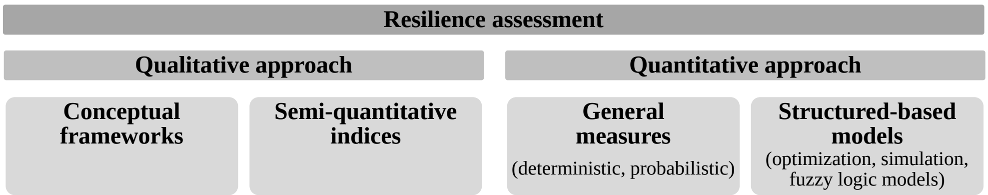


Fig. 2: Resilience assessment approaches (adopted from [10]).

that do not rely on numerical calculations (cf. [13, 14]). These methods include conceptual frameworks that establish best practices, which are the primary qualitative approaches, along with semiquantitative indices that offer expert assessments of different qualitative aspects of resilience. Quantitative methods are categorized into two groups: general resilience approaches, which use deterministic or probabilistic measures to quantify resilience that are independent of specific domains; and structural-based modeling approaches, which employ optimization, simulation, or fuzzy logic models to analyze how a system's structure affects its resilience in specific domains.

Organizational resilience. Over the past decade, research on organizational resilience has gained popularity, leading to the development of numerous approaches, indicators, and methodologies [9, 15, 16]. As shown in Table 2, the underlying conceptual definition generally considers a few key features that characterize an organization's response to a disruption. In the context of an organization, a disruption is an incident that causes an unplanned, negative deviation from the expected delivery of products and services according to the organization's objectives [17].

The importance of measuring organizational resilience is a key requirement to achieving resilience within an organization [5]. Consequently, various approaches have been proposed to measure organizational resilience, which can be categorized into six main groups [18]:

- *The systems view on measuring resilience* emphasizes understanding organizations as complex and dynamic entities, advocating a holistic approach that

Table 2: Definitions of organizational resilience (based on [9, 15])

Features	Definition
Absorption & Flexibility	Enterprise capacity to absorb changes and ruptures through flexibility without affecting its profitability.
Absorption & Transformation	The capacity to absorb shocks effectively, develop situation-specific responses to, and engage in transformative activities.
Anticipation & Adaptability	Ability to anticipate and adapt to key events related to emerging trends and to recover quickly after disasters and crises.
Robustness & Flexibility	Ability of an organization to strengthen the creation of robust and flexible processes in a proactive way.
Recovery & Avoidance	Ability not only to recover from disruptions but to avoid them completely.
Resistance & Recovery	Resistance to shocks, renewal, and recovery or bounce back from shocks.
Resistance & Adaptation	The ability to resist systematic discontinuities and the capability to adapt to new risk environments.
Vulnerability & Adaptation	The ability to manage vulnerabilities and adaptive response in a turbulent environment.
Withstanding & Anticipation	Reactive ability of the company to withstand an external event and active ability to anticipate events.
Withstanding & Adaptation	The ability to withstand systematic discontinuities as well as the capability to adapt to new risky environments.

considers interconnected components, such as stakeholders and environment, to accurately assess resilience.

- *Resilience as an emergent feature of the system* underscores the importance of identifying inherent enterprise attributes that contribute to its resilience.
- *Inherent and adaptive characteristics of resilience* consider resilience under normal operating conditions and the deployment of resourcefulness and extra effort in crisis situations.
- *Resilience as a continuous process* views resilience as the outcome of continuous processes, including planning, responding to threats, and taking adaptive actions to recover.
- *Measuring resilience against disruptions* focuses on preventing and recovering from disruptive events.
- *Measuring resilience using adaptive capacity and time dimension* highlights the importance of the time taken for a system to respond and recover in understanding its resilience.

Next, we delve into existing research on the conceptualization and assessment of business process resilience, a crucial layer of organizational resilience [6].

Table 3: Process characteristics (based on [15, 19]).

Characteristic	Definition
Absorption	The ability to dampen the impact of disruptive events.
Adaptability	The capability to respond to and adapt to the changing environment.
Agility	The ability to rapidly respond to changing conditions.
Flexibility	The ability to change and to adapt to new or complex situations.
Redundancy	The extra capacity to withstand potentially high-impact disruptions.
Robustness	The capability to withstand a given level of stress without significant loss.
Recovery	The ability to quickly resume operations at a desired performance level.
Resourcefulness	The ability to diagnose problems and to initiate solutions.
Rapidity	The ability to react fast to changes in its environment.

Business process resilience. In general, *process resilience* can be summarized as the ability of a process to withstand and recover from a disruption. Similar to organizational resilience, the resilience of a business process can be defined by various characteristics outlined in Table 3, as enhancements in any of these characteristics ultimately strengthen the process's resilience.

Table 4: Overview of existing research on resilience assessment, highlighting the gap in quantitative approaches (general measures) for process resilience assessment.

Resilience assessment level	Resilience assessment method			
	Qualitative approach		Quantitative approaches	
	Conceptual frameworks	Semi-quantitative indices	Structured-based models	General measures
Across disciplines	✱✱✱	✱✱	✱✱✱	✱✱✱
➥ Organizations	✱✱	✱	✱✱	✱✱
➥ Processes	✱	(✱)	✱	ϟ

Legend: "✱✱✱" - Abundant research, "✱✱" - Moderate research, "✱" - Little research, "(✱)" - Indirect research (applicable from other levels), "ϟ" - Missing research (addressed gap).

As highlighted in Table 4, there has been little research conducted to assess resilience at the process level. Furthermore, the existing solutions—discussed in more detail in Section 7.1—predominantly rely on conceptual (qualitative) approaches, such as business continuity analysis [20] or value tree frameworks [21]. There are a few techniques that attempt to quantitatively assess process resilience with actual implementation and validation [3], such as methods that assess the resilience of business process architectures [8] or estimate various levels of process model resilience using a collaboration-oriented modeling language for processes [7, 22]. However, these approaches are structured-based methods primarily utilized to assess process resilience at *design-time*, i.e., during the phase when a process is developed and planned before implementation. They do not evaluate the resilience of a process at *run-time*, i.e., when it is actually being performed. Run-time resilience assessment is crucial because it evaluates how well a process can respond disruptions that may not have been anticipated during the design phase. Furthermore, to avoid the need to rely on expectations about how a process is executed, run-time resilience assessment should be performed in a data-driven manner, which is not done by existing works. Therefore, there is a need for data-driven resilience assessment approaches that evaluate process resilience at run-time using process behavior recorded in event logs.

In the next section, we operationalize the problem of data-driven assessment of process resilience addressed in our work.

3 Problem Statement

In this work, we propose an approach for the data-driven assessment of business process resilience. Our approach assesses such resilience in terms of the ability of a process to withstand and recover from disruptions. Specifically, it analyzes historic data about a process to estimate how its performance will be affected by future (previously unseen) disruptions. It thereby allows us to draw conclusions about the process's overall resilience based on an excerpt of its execution data, as captured in an event log. We operationalize the problem tackled by our approach as follows:

As input, our approach takes an *event log*, capturing data on a process of interest:

Definition 1 (Event log, trace)**.** *An event log L is a collection of events, where each event $e \in L$ has a* case ID*, an* activity*, and a* timestamp*. A* trace *is a sequence of events from L with the same case ID, ordered by their timestamps.*

An event log might contain additional information that can be relevant for assessing business process resilience, such as resource details for each event or trace attributes indicating different *case types* (such as regular or premium customers).

Our approach captures the state and progress of a process through *process features*. We define a process feature as follows:

Definition 2 (Process feature)**.** *A process feature is a characteristic of a business process that is relevant for resilience assessment and measurable as a real-valued numerical metric over time, using information recorded in an event log L.*

We distinguish between *input or in-system features*, such as the numbers of case arrivals or active resources during a specific period, and *output features*, such as the number of completed cases or average lead time during a period. These latter features also encompass *Process Performance Indicators* (PPIs), which are quantifiable metrics that capture a process's effectiveness or efficiency [23]. They can relate to various dimensions, such as time, cost, and quality.

Our work assesses the resilience of a process by quantifying how a PPI will react to a *process disruption*, which we define as follows:

Definition 3 (Process disruption)**.** *A process disruption D is a change in the value of an input or in-system process feature that is* <u>*significant*</u>*, exceeding the range of feature values that are normally observed for that business process, and* <u>*temporary*</u>*, meaning that the feature values return to the normally observed range after a given period of time.*

Examples of process disruptions are a drop in available resources or a rapid increase in case arrivals during a week.[1]

[1] Although an increase in case arrivals can be positive from a business perspective, it may still have negative effects on performance, such as the lead time of a process.

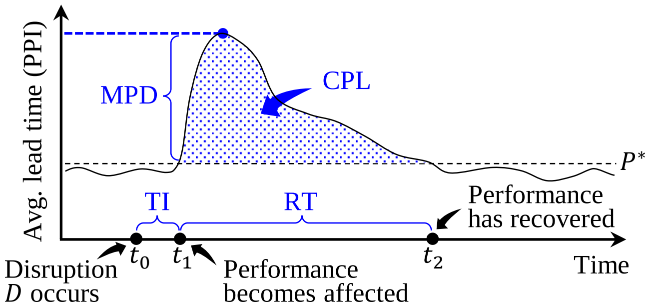


Fig. 3: Four measures of process resilience.

We quantify the impact of a disruption D on a PPI P in terms of four measures derived from established concepts [10], as visualized in Figure 3, assuming an *acceptable performance level* $P^* \in \mathbb{R}$ as the threshold that PPI P must not exceed:

1. *Time-to-impact* ($\texttt{TI}_{[D,P]}$) captures the time between a disruption D, occurring at time t_0, and the next moment t_1 at which process performance P first goes beyond its acceptable performance level P^*.
2. *Recovery time* ($\texttt{RT}_{[D,P]}$) captures the time it takes for PPI P to return to its acceptable performance level P^* (at time t_2) after the initial impact at t_1.
3. *Maximal performance deviation* ($\texttt{MPD}_{[D,P]}$) captures the largest (negative) deflection of PPI P from its acceptable performance level P^* during the recovery period, i.e., between t_1 and t_2.
4. *Cumulative performance loss* ($\texttt{CPL}_{[D,P]}$) aggregates the total performance loss of P incurred during the recovery period, i.e., between t_1 and t_2.

Together, these measures provide in-depth insights into process resilience. Time-to-impact and maximal performance deviation quantify the ability of a process to withstand a disruption. Recovery time assesses the ability of a process to recover from a disruption. Finally, cumulative performance loss summarizes these two abilities by providing an aggregated value.

Given the above, we then define the resilience assessment problem that our work addresses as follows:

Definition 4 (Business Process Resilience Assessment Problem)**.** *Business process resilience assessment uses historic process data in an event log L to estimate the impact of a possible future disruption D on a PPI P under normal process operations, quantified in terms of the time-to-impact, recovery time, maximal performance deviation, and cumulative performance loss.*

In the context of this problem, it is important to note that resilience assessment is not limited to past disruptions recorded in an event log, instead, it considers the impact of possible future process disruptions that can occur. In the next section, we present an approach that can assess process resilience for possible future process disruptions.

4 Approach

In this section, we provide a detailed description of our resilience assessment approach. As depicted in Figure 4, our approach takes as input an event log and a user-defined resilience scenario that specifies the PPIs, process disruptions, and further relevant process characteristics. Based on this scenario, our approach first generates process features as time series to capture the progression of each of the process characteristics over time. Then, we use these time series to establish a VAR model, which captures the linear inter-relations between the different process characteristics. In the final step, we use the obtained VAR model to conduct impulse-response analysis to compute the four resilience measures depicted in Figure 3.

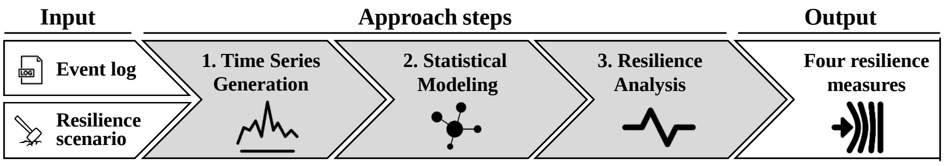


Fig. 4: Overview of the main steps of our resilience assessment approach.

In the remainder, we describe our approach's input (Section 4.1) and three steps (Sections 4.2–4.4) in detail.

4.1 Resilience Scenario

Our approach takes as input a user-defined *resilience scenario* S. The resilience scenario captures the process features that are crucial for assessing process resilience from the user's perspective. Such a scenario is defined in terms of three sets of process features, $S := (F_P, F_D, F_A)$:

- *Performance measures (F_P):* This set contains the output features that are used as PPIs, such as the *average lead time of completed cases* or the *number of rejected orders*. F_P must contain at least one PPI, though a user can also specify multiple ones, so that the resilience of a process is considered from the perspective of different performance measures.
- *Disruption types (F_D):* This set defines process disruptions as input or in-system features that increase or decrease to cause disruptions, such as an increase in the *number of case arrivals* or a decrease in the *number of available resources*. F_D must contain at least one disruption type, though multiple disruptions can be considered in isolation or simultaneously.
- *Additional features (F_A):* Finally, a user can optionally define additional process features that may help to capture the inter-relations between features in F_D and F_P (akin to *mediating variables*). For example, when determining the impact of deviations in the number of case arrivals (in F_D) on the average lead time (in F_P), F_A might include *the number of available resources* as an additional factor.

As a running example in this section, we will use as input the 5th log from the BPI Challenge 2015 [24] and a default resilience scenario S_d, with the *average lead time* as a PPI, i.e., $F_P = \{l\_t\}$, disruption types involving increases in the numbers of *case arrivals*, and *active cases*, and decreases in the *available resources*, i.e., $F_D = \{\uparrow arr\_c, \uparrow act\_c, \downarrow avl\_r\}$, and $F_A = \emptyset$. However, we stress that our approach can work with any process features measurable for an event log. For instance, a resilience scenario can include log-specific features, such as the *number of active cases by premium customers* or the *number of available specialists*.

4.2 Time Series Generation

In the first step, we generate time series that represent the evolution of process features within a resilience scenario. Time series generation is commonly used in process mining to address problems such as concept drift detection and explanation [25], or to assess process complexity and its impact on performance [26]. In the following, we detail the three steps of time series generation used in our approach: windowing, time series construction, and warm-up and cool-down phase detection.

Windowing. First, we split the period of an event log L using time-based tumbling windows of fixed length l [27]. This gives a series of non-overlapping time windows $W_l := \langle w_1, \ldots, w_n \rangle$ of equal length, such that each event $e \in L$ belongs to exactly one window w_t for $t \in \{1, \ldots, n\}$. The first window w_1 starts at the earliest event in L, whereas the last event is in w_n. Our approach can consider different options for window lengths, e.g., l_1, l_2, etc., resulting in a set of window sequences $\mathcal{W} := \{W_{l_1}, W_{l_2}, ...\}$. Based on the modeling results in the next step, our approach automatically selects a window length l that best describes the evolution of the process features over time.

Time series construction. Following convention [28], we define a time series as a set of observations $\{y_t \mid t \in \{t_0, \ldots, t_n\}\}$, where each $y_t \in \mathbb{R}$ represents an observation made at a time t ($\{y_t\}$ in short). We particularly consider series over discrete, equally-spaced time intervals, which means that each y_t captures an observation made for a specific period, e.g., a day or a week.

For all window sequences $W_l \in \mathcal{W}$, we construct time series for each process feature $f_k \in F := F_P \cup F_D \cup F_A$, with $k \in \{1, ..., |F|\}$ as an index. For this, we first calculate the feature value $y_{k,t} \in \mathbb{R}$ for the feature f_k for each window $w_t \in W_l$. Then, we combine these sequential values into a time series $\{y_{k,t}\}$, which captures the evolution of f_k over the windows in W_l. These two steps are repeated for each feature $f_k \in F$, resulting in a set $Y_l := \{\{y_{k,t}\}\}$. Repeating this for each window sequence, we obtain a set $Y := \{Y_l \mid W_l \in \mathcal{W}\}$. Figure 5 shows the week-based time series Y_{week} for our running example.

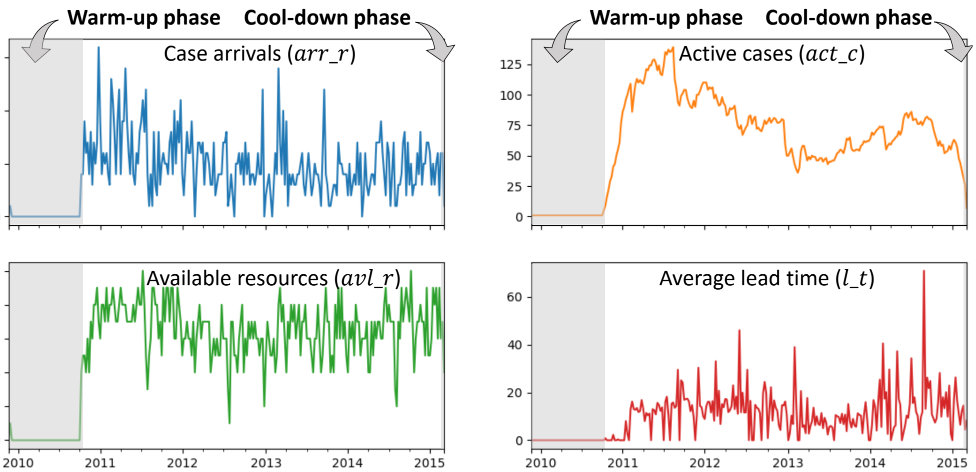


Fig. 5: Process features derived for the running example with indicated warm-up and cool-down phases.

Warm-up and cool-down detection. Finally, we check if the beginning and end of the time series should be truncated. This can be necessary because event logs are data snapshots, consisting of cases associated with a particular time frame. For instance, the running example's log appears to contain all cases that were active between early 2011 and March 2015. However, since some of these cases started much earlier than 2011, the event log has a lengthy *warm-up phase*, in which there are few active cases in the system, as shown in Figure 5. Similarly, if a log consists of cases that started in a certain period, there will be a *cool-down phase* at the end, in which only a few to-be-completed cases remain in the log.

To detect warm-up and cool-down phases, we use the time series $\{y_{k,t}\}$ that describes the evolution of the number of active cases over time (act_c).[2] We use the corresponding frequency distribution of the values $y_{k,t}$ and a percentile δ to define a threshold. We then use this threshold to detect a sequence of windows w_t at the beginning (warm-up) and at the end (cool-down) of W_l, where the feature values $y_{k,t}$ are all below this threshold. Finally, we remove feature values that belong to the detected windows from all features in the set Y_l, resulting in truncated time series Y_l^*. Repeating this procedure for each window sequence W_l, we get $Y^* := \{Y_l^*\}$. In Figure 5, the detected warm-up and cool-down phases are highlighted in gray. For the 1st percentile, $\delta = 1$, we removed 48 windows at the beginning and two at the end.

4.3 Statistical Modeling

In the second step, our approach uses the obtained time series to build a collection of VAR models, from which an optimal model is then selected. In this section, we first introduce VAR models and then explain the model generation and selection steps.

Vector autoregressive models. VAR models are among the most successful and flexible statistical modeling techniques for analyzing multivariate time series. Initially developed by Sims [29], they have proven to be helpful in describing and forecasting process dynamics in many domains. VAR models relate current observations of a variable to past observations of the same and other variables (called *lags*) via a linear combination [30].

Model definition. Given a set of K features, we define a K-dimensional vector $\mathbf{y}_t := (y_{1,t}, \ldots, y_{K,t}) \in \mathbb{R}^K$, which captures the time series values of the features for a given window w_t. Using this vector notation, the standard VAR model is defined as a system of K linear equations:

$$\mathbf{y}_t = A_0\mathbf{d} + \sum_{i=1}^{p} A_i\mathbf{y}_{t-i} + \mathbf{e}_t. \tag{1}$$

[2] We use this feature even if it is not part of the resilience scenario of interest.

The vector $\mathbf{d} \in \mathbb{R}^{m\times 1}$ holds m deterministic parameters, which can be used to model, e.g., an intercept ($\mathbf{d} = 1 \in \mathbb{R}$) or a linear trend ($\mathbf{d} = (1, t) \in \mathbb{R}^2$). The matrix $A_0 \in \mathbb{R}^{K\times m}$ holds the parameters used in $\mathbf{d}$. The value $p \in \mathbb{N}$ is the model order, defining the number of lags that are considered by the model via the vectors $\mathbf{y}_{t-i}$, for $i \in \{1, \dots, p\}$. Matrices $A_i \in \mathbb{R}^{K\times K}$ hold model coefficients that show the per-lag impact of the features on one another. Finally, $\mathbf{e}_t := (e_{1,t}, \dots, e_{K,t}) \in \mathbb{R}^K$ is a vector of error terms.

As an illustration, we consider a first-order VAR model ($p = 1$) with three time series ($K = 3$): $y_{1,t}$, $y_{2,t}$, and $y_{3,t}$, assuming that each model equation includes a linear trend, i.e., $\mathbf{d} = (1, t) \in \mathbb{R}^2$. Then, Equation 1 takes the following form:

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \\ y_{3,t} \end{bmatrix} = \begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \\ a_3 & b_3 \end{bmatrix} \begin{bmatrix} 1 \\ t \end{bmatrix} + \begin{bmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} \end{bmatrix} \begin{bmatrix} y_{1,t\text{-}1} \\ y_{2,t\text{-}1} \\ y_{3,t\text{-}1} \end{bmatrix} + \begin{bmatrix} e_{1,t} \\ e_{2,t} \\ e_{3,t} \end{bmatrix}. \tag{2}$$

Model assumptions. An adequate VAR model should fulfill two assumptions:

1. A VAR model's error terms $\mathbf{e}_t$ should be *white noise*, which is the case when the expectation of the error terms is zero, $E[\mathbf{e}_t] = 0$, the contemporaneous covariance matrix of error terms is nonsingular, $E[\mathbf{e}_t \mathbf{e}_t^T] = \Sigma_{\mathbf{e}}$, the error terms are uncorrelated, $E[\mathbf{e}_t \mathbf{e}_s^T] = 0$ for $t \neq s$, and all fourth moments $E[\mathbf{e}_t^4]$ exist and are bounded (large outliers are unlikely) [31].
2. A VAR model should be stable, which holds when its reverse characteristic polynomial has no roots in and on the complex unit circle [31]. Stability of a VAR model also implies stationarity [31, p. 25], which is necessary for impulse-response analysis (Section 4.4).

In general, VAR models typically assume that all time series are stationary, meaning their statistical properties such as mean and variance remain stable over time. However, this assumption can be overly constraining in practical scenarios, especially within business process analysis. Unlike domains such as finance, where significant trends or abrupt changes are common, features capturing process characteristics often exhibit less variation. Nonetheless, business processes can still undergo subtle or seasonal drifts. In this scenario, concept drift detection could be regarded as an additional preprocessing step to identify different versions of the process and independently assess the resilience of each version before extracting the relevant time series. Alternatively, if seasonality is a permanent characteristic of a business process, standardized techniques for seasonality removal, such as seasonal-trend decomposition based on loess [32], can be applied after time series are extracted.

Model estimation. In situations where these assumptions are fulfilled, the model parameters in Equation 1 can be obtained using multivariate least squares (LS) estimation [31, p. 74]. The resulting LS estimates then have two key asymptotic properties: consistency and normality, which are important to obtain reliable estimation for model parameters [33].

Model generation and selection. Since the optimal configuration of a VAR model cannot be known a priori for a given event log, our approach first generates a collection of candidate VAR models using various parameter settings. Then, it checks the validity of these candidate models regarding the aforementioned model assumptions and, from the valid models, selects the one that has the best fit according to an established information criterion.

Candidate-model generation. We first generate a set of candidate VAR models $\mathcal{M}$, so that each model $M \in \mathcal{M}$ has a unique combination of four parameters:

1. A window sequence $W_l \in \mathcal{W}$.
2. A choice for the deterministic parameter $\mathbf{d}$ from a set of options, which commonly contains $\mathbf{d} = 0$ (no trend or intercept), $\mathbf{d} = 1 \in \mathbb{R}$ (an intercept), $\mathbf{d} = (1, t) \in \mathbb{R}^2$ (a linear trend).

3. A model order $p \in \{1, ..., p_{max}\}$, where p_{max} is the highest possible model order, which can be heuristically estimated[3] by the total number of observations $n = |W_l|$ as $12 \sqrt[4]{n/100}$.
4. A set of features F_M to be included in the model. F_M contains all PPIs from F_P and disruption types from F_D, plus a (possibly empty) subset of additional features from F_A. By including different subsets of F_A, we test whether including extra features leads to a better-fitting model.

Assumption checking. Next, for each of the candidate VAR models in $\mathcal{M}$, we check if it meets the assumptions of noise whiteness and model stability. Given a candidate model $M \in \mathcal{M}$ of order p, we check whether the error terms fulfill the white noise assumption using the multivariate Ljung–Box portmanteau test [35]. The test's null hypothesis is that there is no overall significance in the auto-correlations of the error terms. If we cannot reject the null hypothesis at the significance level $\alpha = 0.05$ and for $p + 1$ lags, then we conclude that there is no evidence to reject the white noise assumption [31, p. 169]. We determine the stability of a model using the eigenvalues of its companion-form representation of order 1, which exists for any model M and provides a more compact form [31]. If all absolute eigenvalues of the companion-form matrix are less than one, the model M is stable [31, p. 15]. We add the models that meet both assumptions to a set of accepted models, $\mathcal{M}_a \subseteq \mathcal{M}$. If no models are found that fulfill the assumptions, we extend the search space and consider models with any set of features F_M that contains at least one performance measure and one disruption type.

Optimal model selection. Finally, we select the optimal model M_a^* as the model in $\mathcal{M}_a$ with the best (i.e., lowest) Akaike Information Criterion (AIC) [28] score. AIC is a widely-employed metric that evaluates a statistical model's ability to fit the data at hand, while accounting for the model's complexity [31]. Given the objective of

[3]This heuristic is applied by the Python library we employ in our implementation [34].

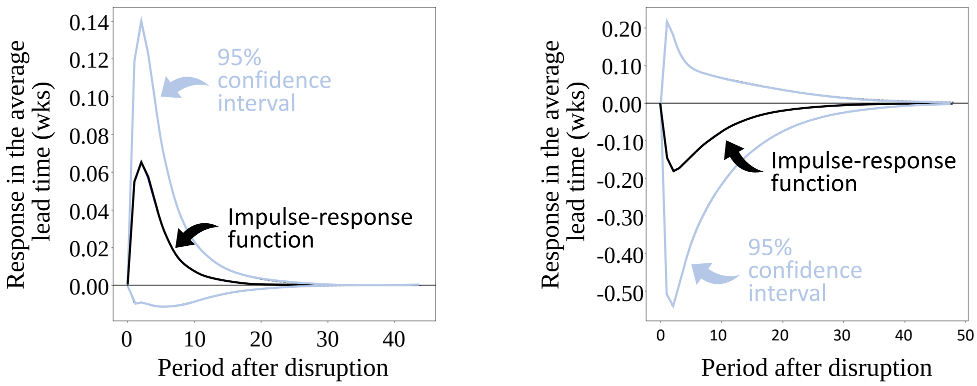
the model to capture as much information as possible from various factors to better describe dependencies during the impulse-response analysis, AIC is a better choice compared to other information criteria, such as the Bayesian Information Criterion (BIC) or the Hannan-Quinn Information Criterion (HQIC), which impose heavier penalties for model complexity. Moreover, AIC tends to perform well with limited data points, a common situation when working with time series data extracted from event logs.

In general, optimal model selection is a challenging task. However, the use of AIC provides a simplification that has proven useful and practical for the addressed problem. Nevertheless, the proposed model selection can be improved by incorporating additional measures or procedures, such as cross-validation [36], or more complex techniques [37, 38].

4.4 Resilience Analysis

In the third step, we assess process resilience using impulse-response analysis, which estimates the expected impact of disruptions on process performance. This corresponds to a form of what-if analysis, which can be derived directly from the obtained optimal VAR model. We first introduce its concept and then explain how we estimate the four resilience measures (time-to-impact, recovery time, maximal performance deviation, and cumulative performance loss) for a given disruption and the corresponding response in a PPI.

Impulse-response analysis. A VAR model's properties are typically assessed through structural analysis such as Granger causality, impulse responses, and forecast error variance decompositions, since examining individual model coefficients alone does not offer a comprehensive understanding of the interactions among the variables in a model. For the purpose of assessing business process resilience, impulse-response analysis serves as the right instrument.



(a) Impulse corresponds to a one-unit increase in the total number of active cases.

(b) Impulse corresponds to a one-unit increase in the total number of available resources.

Fig. 6: Two examples of impulse-response functions (incl. 95% confidence interval), showing the expected response in the average lead time (in weeks) to a one-unit increase in the impulse.

Intuition. Impulse response analysis is a technique for interpreting VAR models. It is based on a *impulse-response function* that predicts how one variable responds to a sudden change in another variable [31], given the identified inter-relations among model variables in Equation 1. To ensure a clean assessment of the response in one variable to a change in another, the analysis is conducted after centering each variable around 0, by subtracting its expected value. Next, a one-unit change is introduced in one variable, and its impact on another variable is observed over subsequent periods, accounting for detected inter-dependencies between the model variables. A one-unit change can always be considered since all time series in a VAR model are numerical and real-valued. Given that the introduced impulse is a one-time change and the VAR model is stable (an assumption checked during the search for the optimal model), the model will return to its centered state after the initial impulse has propagated through the system. This way, impulse-response analysis facilitates a deeper understanding of the dynamic interactions among variables over time, proving valuable in assessing process resilience.

We illustrate the concept of impulse-response analysis using two exemplary impulse-response functions, depicted in Figure 6. The first impulse-response function (Figure 6a) demonstrates the expected increase (black line) in the average lead time following a disruption in the total number of active cases during the periods $h \in H := 1, 2, ..., h_{max}$. The disruption occurs in period $h = 0$ and corresponds to a one-time, one-unit increase in the total number of active cases. The maximum expected deviation in the lead time is observed in the second period after the impulse, elevating the average lead time above its average value by about 0.06 units. The impulse-response analysis also provides confidence intervals (blue lines) indicating where the response in the lead time can be expected with a 95% probability. The second impulse-response function (Figure 6b) illustrates the anticipated change in the average lead time after a one-unit increase in the number of available resources. In this scenario, the expected lead time decreases since more resources are available, making it more likely for the lead time to decrease.

Mathematical derivation. For each feature combination $f_d \in F_D$ and $f_p \in F_P$, we obtain the corresponding impulse-response function $\texttt{IRF}[f_d, f_p](h)$, showing the impact of a potential disruption in feature f_d on performance indicator f_p at the moment h after the disruption, by transforming the (stable) VAR model into its infinite *moving average representation* [31]. To illustrate this transformation, we consider a VAR model of order 1 since any VAR model of order p can be rewritten to that, using the following matrix form [31]:

$$
\underbrace{\begin{bmatrix} \mathbf{y}_t \\ \mathbf{y}_{t-1} \\ \mathbf{y}_{t-2} \\ \vdots \\ \mathbf{y}_{t-p+1} \end{bmatrix}}_{=:Z_t} = \underbrace{\begin{bmatrix} \mu \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{=:\Gamma_0} + \underbrace{\begin{bmatrix} A_1 & A_2 & \cdots & A_{p-1} & A_p \\ I & 0 & \cdots & 0 & 0 \\ 0 & I & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I & 0 \end{bmatrix}}_{=:\Gamma_1} \underbrace{\begin{bmatrix} \mathbf{y}_{t-1} \\ \mathbf{y}_{t-2} \\ \mathbf{y}_{t-3} \\ \vdots \\ \mathbf{y}_{t-p} \end{bmatrix}}_{=:Z_{t-1}} + \underbrace{\begin{bmatrix} \mathbf{e}_t \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{=:\Upsilon_t}.
$$

This matrix form of any VAR model of order p can be simplified using $Z_t, \Gamma_0, \Gamma_1, \Upsilon_t$:

$$
Z_t = \Gamma_0 + \Gamma_1 Z_{t-1} + \Upsilon_t.
$$

Using the initial notation used in Equation 1 for $p = 1$, we can further simplify the obtained equation:

$$
\mathbf{y}_t = \mu + A_1 \mathbf{y}_{t-1} + \mathbf{e}_t,
$$

with $\mu := A_0 \mathbf{d}$. Now, we can derive its infinite moving average representation by using recursive substitution [31]:

$$
\begin{aligned}
\mathbf{y}_t \;=\; & \big(1 + A_1 + A_1^2 + \ldots + A_1^j\big)\mu + A_1^{j+1}\mathbf{y}_{t-(j+1)} \\
& + \big(\mathbf{e}_t + A_1\mathbf{e}_{t-1} + A_1^2\mathbf{e}_{t-2} + \ldots + A_1^j\mathbf{e}_{t-j}\big).
\end{aligned} \tag{3}
$$

For a stable VAR process, the following results hold [31]:

$$(I + A_1 + \ldots + A_1^j)\mu \to (I - A_1)^{-1}\mu \text{ and } A_1^{j+1}\mathbf{y}_{t-j-1} \to 0, \text{ as } j \to \infty.$$

Hence, Equation 3 reduces to:

$$\mathbf{y}_t = \varphi + \sum_{j=0}^{\infty} A_1^j \mathbf{e}_{t-j},$$

where $\varphi := (I - A_1)^{-1}\mu$. Now, by rewriting $B_j := A_1^j$ and $B_0 := I$, we get:

$$\mathbf{y}_t = \varphi + \sum_{j=0}^{\infty} B_j \mathbf{e}_{t-j}. \tag{4}$$

Finally, the obtained representation is used to track the impact of a disturbance in one variable on the other. This is done by initiating the impulse vector $\mathbf{e^t} := (\mathbf{0}, \ldots, \mathbf{0}, \mathbf{1}, \mathbf{0}, \ldots \mathbf{0})$, where the 1 in $\mathbf{e^t}$ corresponds to the position of the feature f_d in the VAR model. Afterward, the impact on any variable within the system for a given future time period t can be assessed by examining the outcome of the computations in Equation 4, given by the right-hand side of the equation. The 95% confidence interval for each period t and pair of impulse and response variables is derived using the associated covariance matrix of $\varphi \sum_{j=0}^{\infty} B_j$, where j ends at t.

Resilience assessment. To assess process resilience, we consider the resilience scenario $S = (F_P, F_D, F_A)$ and the selected optimal VAR model M_a^*. For each feature combination $f_d \in F_D$ and $f_p \in F_P$, we use the corresponding impulse-response function $\texttt{IRF}[f_d, f_p](h)$, which estimates the expected impact of a potential disruption in feature f_d on performance indicator f_p.

Resilience aspects. We account for the following aspects when using an impulse-response function to assess business process resilience:

1. *Disruption size*: By default, an impulse-response function considers the impact of one additional unit, which may be insufficient for modeling an actual disruption. However, as a VAR model is linear, the initial impulse can be rescaled by multiplication with a scalar of interest. A common choice is one, two, or three standard deviations of the impulse variable. This rescaling is widely employed as it facilitates the comparison of disruptions' impacts across variables with different scaling [31]. Consequently, the values of the impulse-response function are rescaled as well, maintaining the same shape. Using the impulse-response function in Figure 6a as an example, if the standard deviation of the number of active cases is 10, and the average lead time is 3 weeks, then a disruption, represented by 2 standard deviations, will increase the lead time by $10 \times 2 \times 0.06 = 1.2$ weeks in the second period after the disruption. In other words, the maximum expected increase in the average lead time is approximately 40% (1.2 wks / 3 wks).
2. *Disruption direction*: Depending on the resilience scenario, we may be interested in disruptions that increase or decrease the impulse variable. For example, when assessing resilience to disruptions in the number of active cases, we want to evaluate the impact of increases, while in the case of available resources, we want to observe the effect of decreases. To obtain the impact of a decreasing disruption, we multiply the values of the impulse-response function by -1. In the context of the impulse-response function in Figure 6b, where we aim to observe the response to a decrease in the number of available resources, this action will mirror the obtained curves along the x-axis.
3. *PPI's negative deviation*: When examining the response in a PPI, our focus lies on negative deviations, which we denote as $\texttt{IRF}[f_d, f_p](h)^-$. Depending on the PPI, a negative impact can be linked to either an increase or decrease in the PPI (or both if the PPI needs to remain within a specified range). In the scenario of average lead time, an increase is deemed negative as it results in prolonged execution

time, commonly viewed as an undesirable outcome. Moreover, deviations in a PPI are commonly tolerated up to a certain level. For this reason, we consider a specific *tolerated deviation*, P_{tol}. This defines the accepted deviation of a PPI from its average (target) value, i.e., $P^* = \bar{f}_p + P_{tol}$ (or $-P_{tol}$ for PPIs where higher is better). If the response to a disruption is less than P_{tol}, the disruption is not considered to impact the PPI.

4. *PPI's deviation confidence*: When assessing the expected PPI deviation after a disruption, we aim to measure process resilience in terms of the worst-case scenario. In other words, we want to ensure that the performance deviation does not exceed a specific threshold, like 2 weeks for average lead time, with a 95% probability. For this, we are more interested in the deviations given by the upper and lower confidence bounds (blue lines) in Figure 6 around the expected response given by the black curve. Depending on the resilience scenario, we denote $\texttt{IRF}^*[f_d, f_p](h)^-$ as the relevant negative extreme bound for the PPI's deviation.

Next, we explain how we define resilience measures using the relevant impulse-response function.

Resilience measures. For a given resilience scenario $S = (F_P, F_D, F_A)$ and the tolerated deviation P_{tol}, we define the following four resilience measures:

- *Time-to-impact*: We define time-to-impact as the first period after a disruption where the PPI's negative deviation exceeds P_{tol}:

$$\texttt{TI}_{[f_d, f_p]} := \inf\{h \in H : |\texttt{IRF}^*[f_d, f_p](h)^-| > P_{tol}\}.$$

  If the deviation does not exceed P_{tol}, all remaining resilience measures are 0, since there is no impact on the PPI.

- *Recovery time*: The recovery time is given by the difference between the last time at which the predicted negative deviation is greater than the tolerated deviation and the time-to-impact value:

$$\texttt{RT}_{[f_d,f_p]} := \sup\{h \in H : |\texttt{IRF}^*[f_d, f_p](h)^-| > P_{tol}\} - \texttt{TI}_{[f_d,f_p]}.$$

Both time-to-impact and Recovery time take integer between 0 and h_{max}.

- *Maximal performance deviation*: The maximal performance deviation is given as the greatest absolute negative deviation of the impulse-response function during the recovery period:

$$\texttt{MPD}_{[f_d,f_p]} := \sup\{|\texttt{IRF}^*[f_d, f_p](h)^-|,\ h \in \{\texttt{TI}_{[f_d,f_p]}, \ldots, \texttt{TI}_{[f_d,f_p]} + \texttt{RT}_{[f_d,f_p]}]\}\}.$$

- *Cumulative performance loss*: To quantify the cumulative performance loss, we sum up the gains and losses incurred between initial impact and recovery:

$$\texttt{CPL}_{[f_d,f_p]} := \sum_h \texttt{IRF}^*[f_d, f_p](h),\ h \in \{\texttt{TI}_{[f_d,f_p]}, \ldots, \texttt{TI}_{[f_d,f_p]} + \texttt{RT}_{[f_d,f_p]}\}.$$

Both $\texttt{MPD}_{[f_d,f_p]}$ and $\texttt{CPL}_{[f_d,f_p]}$ takes a value in $\mathbb{R}^+$.

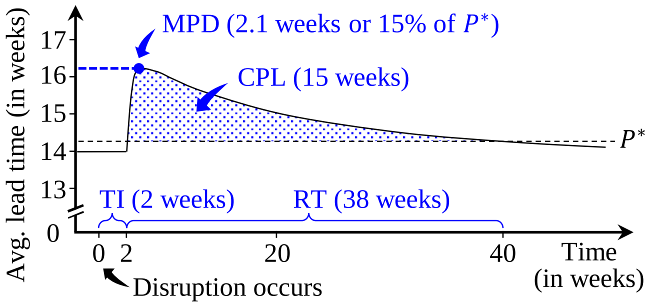


Fig. 7: Process resilience in terms of the average lead time in response to a disruption in the number of active cases for the running example.

Example. To illustrate the output obtained in this manner, Figure 7 depicts an impulse-response function $\texttt{IRF}^*[f_d, f_p](h)$ for the running example. It shows the response of the average lead time (l_t, in weeks) to a disruption of two standard deviations in the number of active cases ($\uparrow act\_c$), using a tolerated deviation P_{tol} of 1 percent (about

0.15 week). With a probability of 95%, the actual increase in the lead time is not going to exceed the predicted response $\texttt{IRF}^*[f_d, f_p](h)$ (black curve). As shown, the disruption impact occurs quickly (time-to-impact of just two weeks) and increases the average lead time from about 14 weeks to a maximum of 16.1 (an increase of 15%). After this initial spike, the process starts recovering gradually, though it takes a total of 40 weeks before it returns to its acceptable performance level. Overall, the cumulative performance loss is 15 weeks, which can be used when comparing the impact of this disruption to other types or when comparing the process' resilience to other versions. We demonstrate this, among others, in the evaluation experiments performed next.

5 Evaluation

This section reports on two experiments conducted to evaluate our approach. In the first experiment, we assess the validity of our approach through a controlled process simulation. In the second experiment, we evaluate the effectiveness of our approach by applying it to real-life event logs.

Our experiments can be replicated using our repository[4], which contains a Python prototype of our approach, input data, relevant experimental details, and raw results. Our prototype employs functions from the *PM4Py* [39] and *statsmodels* [34] libraries.

5.1 Experiment 1: Approach validity

In the first experiment, we demonstrate the validity of our approach by assessing the accuracy of estimated resilience measures. To accomplish this, we conduct multiple simulations of a business process under a predetermined resilience scenario and compare the actual resilience measures—from the simulation—against the estimations obtained using our approach.

[4]Project repository: https://gitlab.uni-mannheim.de/processanalytics/resilience

5.1.1 Setup

In the following, we discuss the relevant experiment settings, i.e., the simulation configurations, resilience scenario, approach configurations, and evaluation metric.

Simulation scenario. As a basis for this experiment, we consider a process model and its simulation parameters that were used in prior work on business process resilience assessment [8]. The model captures an order-to-cash process of a medium-sized company, depicted in Figure 8.

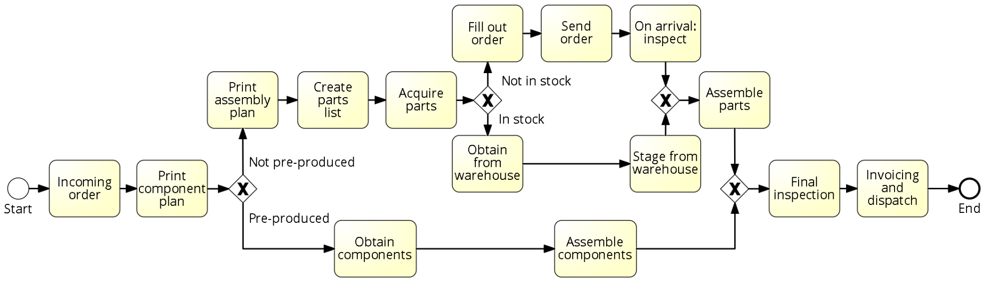


Fig. 8: Simulated order-to-cash process in Experiment 1 (adapted from [8]).

Control-flow, processing times, and resources. The process model consists of 15 activities and two decision points. Following the parameters of the original scenario [8], each decision point has an equal branching probability when simulating the process. Furthermore, the execution times of the activities are given by the distributions shown in Table 5. Finally, each task is managed by one dedicated resource, operating 24/7.

Arrival process. The original paper does not specify how case arrivals were simulated, thus we utilize the Poisson distribution $P(\lambda)$ for the arrival process. We use an arrival rate of $\lambda_n = 0.45$ cases per hour, which leads to stable process utilization without causing queues to grow.

Process disruption. We perform simulation runs that last 10,000 (simulated) hours, with each run involving a disruption in the case arrival rate λ after 70% of the time. This timing ensures the model stabilizes before the disruption, and allows sufficient

Task	Distribution	Parameters	
Incoming order	Log-normal	$\mu = 0.2$	$\sigma = 0.4$
Print component plan	Normal	$\mu = 0.5$	$\sigma = 0.1$
Print assembly plan	Normal	$\mu = 0.6$	$\sigma = 0.15$
Create part list	Gamma	$\alpha = 0.9$	$\beta = 0.7$
Acquire parts	Gamma	$\alpha = 0.8$	$\beta = 0.8$
Fill out order	Log-normal	$\mu = 0.1$	$\sigma = 0.5$
Send order	Gamma	$\alpha = 1.0$	$\beta = 0.5$
Arrival and inspection	Log-normal	$\mu = 0.25$	$\sigma = 0.5$
Obtain from warehouse	Log-normal	$\mu = 0.07$	$\sigma = 0.3$
Stage from warehouse	Gamma	$\alpha = 0.8$	$\beta = 0.3$
Assemble parts	Log-normal	$\mu = 1.3$	$\sigma = 0.4$
Obtain components*	Log-normal	$\mu = 0.4$	$\sigma = 0.4$
Assemble components	Log-normal	$\mu = 0.4$	$\sigma = 0.4$
Final inspection	Normal	$\mu = 1.0$	$\sigma = 0.4$
Invoicing and dispatch	Normal	$\mu = 0.8$	$\sigma = 0.3$

* No specifications were provided for "Obtain components", we used the same as for "Assemble components".

Table 5: Simulation of execution durations (hours) in the process depicted in Figure 8 (adopted from [8]).

time afterward to observe changes in the PPI. The increase in the arrival rate persists for exactly one window, returning then to its normal level of $\lambda_n = 0.45$ cases per hour.

We simulate disruptions by increasing the number of case arrivals by up to 20 standard deviations. This is achieved by multiplying the arrival rate by a factor ranging between 2 and 4 times the normal arrival rate. For each tested increase in the arrival rate, we conduct 1000 simulations, resulting in 10000 simulation runs in total.

Implementation. To simulate the resilience scenario, we built a simulation model using the CIW library [40], an open-source library for conducting discrete event simulations[5]. This library allows us to simulate the process as a network of queues, providing the functionality and simulation logic required to model the desired simulation. The total computation time for his experiment took about 3 hours on 50 CPUs[6].

Resilience scenario. We consider a resilience scenario $S = (F_P, F_D, F_A)$, where the disruption type F_D is given by an increase in the number of case arrivals. The performance measure F_P is given by the average lead time derived from the lead time of all cases that finished during the window. Finally, the additional feature F_A is given

[5] Available online: https://ciw.readthedocs.io/en/latest/index.html
[6] Intel Xeon CPU E5-2698 v4 @ 2.20GHz.

by the number of active cases determined by the number of cases that were in progress during the window[7].

Approach configuration. We use the following settings when applying our approach. In Step 1, we use a fixed window length l of 70 hours (appr. 3 days). In this controlled experiment, this window size leads to a balanced representation of the process features as time series over the total simulation duration. It is neither too large, ensuring a sufficient length of the time series before the disruption to fit VAR models, nor too small, preventing any of the derived process features from having zero values. We test five percentiles for the warm-up and cool-down phase detection parameter δ, i.e., $\delta \in \{1\%, 5\%, 10\%, 15\%, 20\%\}$. In Step 2, we fit VAR models without an intercept, with an intercept, and with an intercept and a linear trend, i.e., $\mathbf{d} \in \{0, 1, (1, t)\}$ and use a significance level $\alpha = 0.05$ for the multivariate Ljung–Box portmanteau test. In Step 3, we consider a response horizon h_{max} of 10 windows after the disruption moment, which corresponds to about 10% of the simulated time before the disruption. This response horizon covers enough time to capture the response in the post-disruption PPI. We consider a tolerated deviation P_{tol} of 30% of the standard deviation of the average lead time observed before the disruption.

Accuracy assessment. To evaluate our approach, we assess how often each resilience measure falls within the interval estimated by our approach for each simulation run. To derive the interval estimated by our approach, we apply our approach to the portion of the event log that occurs before the disruption. This way our approach does not observe the disruption in the data nor does it learn how the process reacts to it. Then, we derive lower R_l^{est} and upper R_u^{est} bounds using the impulse-response functions that denote the corresponding 99% confidence interval, resulting in the interval $I := [R_l^{est}, R_u^{est}]$. To derive the actual resilience measure R^{act} from the simulation, we consider the response in the PPI following the disruption.

[7]In this simulated experiment, we do not consider the number of available resources. This is because each activity has its own dedicated resource, and all resources are consistently available to process tasks once they are completed with the current one.

Given R^{act} and I for each resilience measure, we evaluate our approach by considering its accuracy, denoted Acc, as the proportion of simulations where the actual value R^{act} falls within an interval of interest $I = [R_l^{est}, R_u^{est}]$ (correct detections) divided by the total number of conducted simulations N (all detections) for a given disruption size. It is defined as follows:

$$Acc := \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_{I_i}(R_i^{act}), \quad \text{where} \quad \mathbf{1}_{I_i}(R_i^{act}) := \begin{cases} 1, & \text{if } R_i^{act} \in I_i, \\ 0, & \text{otherwise.} \end{cases}$$

The accuracy ranges between 0 and 1. The higher the value, the more accurate our approach is.

5.1.2 Results

Table 6 shows the results of this experiment, depicting the accuracy of our approach obtained for different disruption sizes in the number of case arrivals.

Table 6: Results of Experiment 1, showing the accuracy obtained for different disruption sizes.

Disruption size	**Moderate (5-10 std.)**	**Severe (10-15 std.)**	**Extreme (15-20 std.)**	**Average**
Time-to-impact	0.75	0.83	0.68	0.76
Recovery time	0.86	0.88	0.89	0.88
Maximal deviation	0.65	0.78	0.88	0.77
Cumulative perf. loss	0.70	0.72	0.72	0.71
Average	**0.74**	**0.80**	**0.79**	**0.78**

Overall, our approach demonstrates a satisfactory level of accuracy in estimating resilience measures, achieving an average accuracy across all disruption sizes and resilience measures of 0.78. With respect to different disruption sizes, the average accuracy ranges from 0.74 to 0.80 (without a clear trend), indicating consistency in accuracy largely regardless of the disruption's severity.

Concerning different resilience measures, the accuracy varies between 0.71 and 0.88. Notably, the approach achieves its highest accuracy of 0.88 for recovery time estimation, maintaining consistency across various disruption sizes. While the accuracy for cumulative performance loss also demonstrates consistency, it is the lowest among the resilience measures with an average accuracy of 0.71. However, this aligns with expectations, as the accuracy of cumulative performance loss is influenced by the combined accuracy of other resilience measures. The accuracy for the time-to-impact measure spans from 0.68 to 0.83, without any notable trend relative to disruption size. Lastly, maximal performance deviation demonstrates good accuracy, surpassing 0.78 for severe and extreme disruption sizes. However, it tends to slightly underestimate the severity of the impact in cases of severe disruption.

The results of this experiment thus show that our approach is able to accurately estimate the disruption size in a complex scenario. After this confirmation of the validity of our approach, we conduct the next experiment to show the effectiveness of our approach using real-life data.

5.2 Experiment 2: Approach effectiveness

After confirming the validity of our approach, we aim to showcase the effectiveness of our approach in evaluating business process resilience using the introduced four resilience measures. Since we do not have access to the actual resilience ground truth of the processes recorded in an event log, our evaluation centers on real-life event logs describing the execution of the same process across various organizations. This allows us to compare findings about process resilience, providing relative insights into strengths and weaknesses across different organizations and demonstrating the utility of our analysis.

5.2.1 Setup

Next, we discuss the setup and obtained results of the conducted experiment.

Data collection. To perform the experiment, we used data from the BPI Challenge 2015 [24], which consists of five real-world event logs (M1-M5), each capturing a permit application process at a Dutch municipality. The nature of these logs makes them highly suitable for our work because they allow us to assess and compare the resilience of the same process across different organizations, accounting for differences in its implementation. The event logs are comparable in terms of the covered period (about 5 years) and size, including numbers of events, event classes, and traces, as shown previously shown in Table 7. However, we also observe considerable differences in their lead times (averages ranging approximately from 9 to 23 weeks), which strongly fluctuate, as evidenced by the high standard deviation (from 14 to 24 weeks).

Table 7: Properties of the considered real-life event logs.

Event log	**Period (wks)**	**N. of events**	**Event classes**	**N. of traces**	**Trace variants**	**Lead time Avg./SD (wks)**
BPIC15 M1	252	52 217	398	1199	1170	14/17
BPIC15 M2	244	44 354	410	832	828	23/24
BPIC15 M3	270	59 681	383	1409	1349	9/14
BPIC15 M4	276	47 293	356	1053	1049	17/15
BPIC15 M5	275	59 083	389	1156	1153	14/15

Approach configuration. We use the following settings when applying our approach. In Step 1, we use a fixed window length to achieve comparability of resilience insights between municipalities. We select semi-monthly time windows because they were found to be the overall best window size across the logs, given their length and fluctuations in the process features. We test the same five percentiles for the warm-up and cool-down phase detection parameter δ, as in the first experiment. In Step 2, we fit VAR models with the same configurations for trend options, keeping again the significance level $\alpha = 0.05$ for the multivariate Ljung–Box portmanteau test. In Step 3, we use a tolerated deviation P_{tol} of 10% of the standard deviation of the PPI and consider a response horizon h_{max} equal to 50% of the window sequence length. Finally, we let each disruption correspond to an increase or decrease of two standard deviations of the respective feature (using the standard deviation across the time-series values).

Using our implementation with the specified configuration, the total computation time for this experiment on an Intel i7-9750H CPU (2.60GHz) with 16 GB RAM took about 40 minutes.

Experiments. We conducted two experiments to assess process resilience[8]:

Experiment 2.1: Overall resilience. We assess the overall resilience per municipality using the default resilience scenario described in Section 4.1. Specifically, we consider the average lead time as the PPI, i.e., $F_P = \{l\_t\}$, and disruptions in the numbers of case arrivals, active cases, and available resources, i.e., $F_D = \{\uparrow arr\_c, \uparrow act\_c, \downarrow avl\_r\}$. We do not consider additional features, i.e., $F_A = \emptyset$.

Experiment 2.2: Resilience per case type. Given that the process under investigation covers a broad range of permit applications, we use this experiment to examine if the process is specifically prone to increases in certain case types. To do this, we consider a resilience scenario with disruptions in terms of the three most common case types per municipality (based on a case's "parts" attribute), $F_D = \{\uparrow act\_c\_type_x, \uparrow act\_c\_type_y, \uparrow act\_c\_type_z\}$, where x, y, and z differ per municipality, due to varying commonality of the case types. Other scenario options are left the same as in Experiment 2.1, i.e., $F_P = \{l_t\}$ and $F_A = \emptyset$.

5.2.2 Results

In this section, we report on the results obtained from our experiments.

Experiment 2.1: Overall resilience. To contextualize resilience insights, Table 8 presents the main characteristics of diverse time series in terms of the average values, standard deviations, and coefficients of variation, obtained through semi-monthly windows. Here, we find that the processes are rather stable in terms of the available

[8]In the second experiment with real-life data, we determine the average lead time by considering the lead time for each window of the last 3% of all cases completed before the end of that window to prevent situations where the lead time is inflated by a few lengthy cases completed within a specific window.

resources (avg. coefficient of variation of 19%) and active cases (22%) but fluctuate in terms of case arrivals (46%). Per municipality, we see considerable differences, with, e.g., M3 having much less fluctuation in its active cases (just 14%) than others (up to 32%).

Table 8: Statistics of the time series per each process feature from the default resilience scenario on a semi-monthly windowing.

Event log	Case arrival			Active cases			Avl. resources			Lead time		
	Avg	**Std**	**CV**	**Avg**	**Std**	**CV**	**Avg**	**Std**	**CV**	**Avg**	**Std**	**CV**
M1	11.3	5.6	49%	88.2	12.4	14%	7.4	1.7	22%	6.5	2.1	33%
M2	8.0	4.8	60%	94.9	21.0	22%	5.3	1.1	21%	9.7	3.4	35%
M3	13.2	4.9	37%	69.1	9.7	14%	6.7	0.8	12%	3.7	1.2	32%
M4	9.9	4.2	42%	91.4	25.3	28%	5.2	1.2	22%	7.4	2.3	31%
M5	10.9	4.7	43%	82.3	26.1	32%	8.1	1.3	16%	6.1	2.4	40%

Table 9 provides an overview of the overall resilience results. We observe that the processes have the worst resilience to increases in the active cases, with the longest average recovery time across all municipalities and the greatest average maximum increase in the average lead time, causing a notable cumulative performance loss. Although resilience to the other two disruptions is generally comparable in terms of maximal impact, the average recovery time after decreases in the number of available resources takes 40% longer compared to increases in the case arrival, which doubles the observed average cumulative performance loss.

Table 9: Results of Experiment 2.1, showing the resilience for the three disruption types (semi-monthly windows).

Disruption	**Increase in case arrivals**					**Increase in active cases**					**Decrease in avl. resources**				
Municipality	M1	M2	M3	M4	M5	M1	M2	M3	M4	M5	M1	M2	M3	M4	M5
Time-to-impact	1	1	1	1	9	1	1	1	1	1	0	1	1	1	1
Recovery time	4	17	3	1	11	8	28	8	22	29	0	28	8	11	5
Max. deviation	0.5	0.6	0.7	0.4	0.2	0.3	2.5	0.4	1.2	0.9	0	0.9	0.4	0.8	0.5
Cum. perf. loss	1.4	5.7	1.3	0.4	2.6	2.2	25.8	2.3	15.9	16.4	0	13.5	2.2	5.6	1.8

When comparing the resilience among the municipalities, we also gain several interesting insights. Municipality M1 exhibits the highest overall resilience. It demonstrates strong resilience to disruptions in case arrivals and active cases. Furthermore, it is not anticipated to experience any performance drops from disruptions in available resources. In other words, the expected increase in the average lead time does not exceed the tolerated deviation P_{tol}. Municipality M3 also displays strong resilience; however, a potential disruption in available resources affects the lead time, compared to M1. In contrast, municipality M2 shows the worst resilience in all disruption scenarios, showing the highest cumulative performance loss across all municipalities. Finally, municipalities M4 and M5 exhibit relatively low resilience to an increase in the number of active cases, remaining competitive with other municipalities in other disruptions. Based on such insights, municipality M2 could conduct further analyses to understand why their resilience on the same process is so different compared to M1 and M3, as a basis for future improvement.

Experiment 2.2: Resilience per case type. Table 10 shows the resilience insights obtained for disruptions in the three common application types per municipality. To contextualize the resilience insights, we include the corresponding frequency and average lead time for each case type. We observe that the three most common application types account for 55 to 70 % of all cases. Interestingly, type A is most common for all municipalities, accounting for 40 to 50% of the total cases, and type B always is the second or third most common type. The lead times differ remarkably between case types and municipalities. For M1, for instance, type A takes about 11 weeks on average (5 semi-monthly windows), whereas cases of type C take 40 weeks (19 semi-monthly windows).

Based on the computed resilience measures, we observe several differences in resilience to the three most common case types compared to the overall resilience. First of all, we see that municipalities M1 and M3 show the highest resilience to all frequent case types. This aligns with insights from an overall analysis of resilience to disruptions in the number of active cases, depicted in Table 6. The resilience measures for common case types at M2 are comparable to the overall resilience in terms of recovery time and cumulative performance loss. However, municipalities M4 and M5 show remarkably lower resilience to increases in the three most common case types compared to the overall resilience. In both cases, the maximum impact doubles, and with a long recovery time reaching the maximum considered horizon, the cumulative loss is significant.

Table 10: Results of Experiment 2.2, showing process resilience (semi-monthly windows) per municipality to disruptions in the three most common application types and the corresponding statistics, obtained based on the initial event logs.

Municipality	**M1**			**M2**			**M3**			**M4**			**M5**		
Increase in case type	*A*	*B*	*C*	*A*	*B*	*C*	*A*	*B*	*D*	*A*	*B*	*E*	*A*	*B*	*D*
Case frequency (%)	50	8	5	39	11	7	39	19	6	44	15	9	50	6	10
Avg. lead time	5	5	19	9	4	20	4	2	6	8	6	7	5	6	7
Time-to-impact	1	1	2	2	1	5	3	3	1	1	4	1	1	1	1
Recovery time	8	12	14	36	31	45	10	4	26	52	49	52	50	50	50
Max. deviation	0.5	0.9	0.3	1.3	1.6	0.8	0.5	0.2	0.6	3.2	0.9	3.4	2.0	2.3	1.6
Cum. perf. loss	2	6	3	29	27	29	3	0.8	8	147	30	151	82	102	73

Case type descriptions:
A: "Bouw"("Construction"), B: "Kap" ("Felling (Cutting down trees)"), C: "Milieu (vergunning)" ("Environment (permit)"), D: "Bouw, Handelen in strijd met regels RO" ("Construction, Acting against spatial planning rules"), E: "Handelen in strijd met regels RO" ("Acting against spatial planning rules")

We also observe notable differences in resilience for the same case type across different municipalities. For example, resilience to the most common case type A is high at M1 and M3 but worse at other municipalities, especially M4. A similar trend is observed for type B, with lower resilience at M5. Looking at type C at M1 and M2, there's a notable difference in resilience despite the same case frequency and average lead time. A similar trend is noted for type D in municipalities M3 and M5.

Overall, when considering resilience per case type and municipality, we see the importance of examining resilience by case type. The resilience insights obtained from this experiment expand upon the findings from the overall resilience analysis and can guide further investigation into improving process resilience in municipalities. For instance, it is worth investigating why the resilience to the most common case types at M4 and M5 differs significantly from their overall resilience or whether batching causes low resilience at M4.

6 Discussion

In this section, we first discuss the advantages of our approach for business process resilience assessment in comparison to the alternative solution. Then, we examine the limitations of our approach, identifying potential areas for improvement.

Advantages of the proposed solution. The proposed approach for resilience assessment offers several advantages that make it particularly well-suited for the task at hand. First, the VAR model estimates how process performance responds to disruptions using impulse-response analysis, which closely aligns with the idea of measuring resilience against disruptions (see Section 2). Moreover, the impulse-response analysis provides insights into the magnitude and duration of disruption impacts, enabling direct identification of all four resilience measures. Second, the VAR model assumes linear relationships between relevant process features, which have been proven to provide reliable and accurate estimates of the resilience of a business process under the short-term impact of disruptions (see Section 5.1), even when process disruptions have not been previously observed in the data. Finally, our approach is automated, requiring only an event log and a user-defined resilience scenario as input, and can be used out-of-the-box.

Comparison with alternative solution. Our approach offers significant advantages over the alternative solution based on (auto-mined) process simulations. Concerning the latter, we note that constructing a simulation model that adequately replicates process behavior is highly complex and time-consuming [41, 42]. Simulation models discovered in an automated manner [43–45] lack the functionality needed to simulate disruption scenarios essential for resilience assessment, requiring additional customization and hindering automated resilience assessment. In contrast, our approach eliminates the need to create simulation models from scratch or to modify automatically generated process simulations to include specific functionalities.

Limitations. The use of a VAR model to assess the resilience of a business process from event data also has limitations, which we aim to address in the future.

The linearity of the VAR model allows for robust estimation of the short-term impact of a disruption on process performance. However, it might lead to an underestimation of the impact if the disruption size is extremely large. This occurs because a process might behave non-linearly under extreme disruptions, making rescaling the detected response in process performance ineffective.

The first step in our approach generates process features as time series, which describe characteristics relevant for resilience assessment. Although these process features are typically stationary, maintaining a consistent mean and variance, they can exhibit significant changes due to factors such as concept drift. In such cases, preprocessing steps are necessary. For instance, concept drift detection can be applied to an event log to identify these changes before assessing the resilience of the process version of interest. Furthermore, if a process is highly seasonal, the time series may need to be de-seasonalized to eliminate the impact of seasonality on the resilience assessment results.

Finally, our approach employs the Akaike information criterion for selecting optimal model settings. Despite its acceptable results, this is a relatively simplistic strategy, whereas selecting (truly) optimal models is a complex task.

7 Related Work

Our work relates to resilience assessment approaches specific to business processes and some well-established problems in process mining. It also relates to broader research areas in other fields.

7.1 Process Resilience Assessment

In the following, we discuss the two main research directions in process resilience assessment and explore related problems within process mining.

Main research directions. The assessment of business process resilience diverges into two directions: resilience assessment at design-time and at run-time.

Resilience assessment at design-time. Resilience assessment at design-time refers to the evaluation of a process' resilience capabilities during the initial design phase. A recently proposed method introduces four levels of process model resilience based on a data-centric, collaboration-oriented modeling language for processes [22]. Based on this method, a resilience-aware maturity model for modeling multi-party business processes allows precise quantification of model compliance concerning the different resilience levels [7]: no resilience awareness, failure awareness, data resilience, milestone resilience, and process resilience.

Resilience assessment at run-time. Resilience assessment at run-time involves evaluating a process's ability to withstand disruptions and maintain performance during its operational phase, with a focus on monitoring its response to unexpected events. Our paper focuses on data-driven resilience assessment at run-time, for which little research in business process management and process mining has been conducted so far. The only research that addresses this problem is the work by Zahoransky et al., who have presented a decision support framework [6]. This framework collects and quantifies metrics and indicators for assessing the run-time resilience of a process based on the ex-post analysis of event logs. They also conducted a case study [8], investigating the temporal aspect of process resilience, done by applying process mining to create probability distributions on the time behavior of business processes, using historic information in an event log. However, their approach is not generalizable because it depends on domain knowledge and manual intervention and requires a known process model for building a simulation model, making process-centered and data-driven resilience assessment impractical.

Connection to other problems in process mining. In the following discussion, we position the problem of assessing business process resilience alongside other well-established problems in process mining, highlighting the unique aspects of resilience assessment.

Anomaly detection. Anomalies can be defined as deviations in process behavior from normal or expected behavior according to an established process model [46]. Numerous supervised and unsupervised techniques have been proposed to detect these deviations within event logs, as highlighted in recent surveys [47, 48]. However, anomaly detection cannot be used to address the problem of process resilience assessment due to its different scope and approach level. Specifically, process resilience assessment aims to estimate the potential impact of future process disruptions, which indicate changes characteristics of a process at the system level. In contrast, anomaly detection focuses on identifying past deviations recorded in an event log, either at the trace level or within segments of traces.

Concept drift detection. Concept drift in process mining occurs when a process changes while being analyzed [49]. The objective of concept drift detection is to identify and describe these changes using data from an event log. A temporary disruption, such as a sudden increase in case arrivals, should not be mistaken for concept drift. Unlike concept drift, which indicates a shift to a new version of the process model, disruptions are temporal and are more linked to anomalies. Furthermore, concept drift detection focuses on detecting changes in the recorded data rather than predicting how a process might react to future disruptions. Therefore, resilience assessment presents a distinct challenge compared to concept drift detection.

Causal process mining. Causal process mining involves identifying and understanding the cause-and-effect relationships within different features of a business process. Causal analysis has been successfully used to automate the discovery of factors that impact process performance [50] or to mine process dependencies for control-flow decisions taken during the execution of process models [51].

The problem of resilience assessment can be viewed as a problem of causal analysis, where the disruption is the cause and the decline in process performance is the effect. However, in this context, identifying the cause and effect is redundant since they are already known, making the quantification of the impact the primary concern. Existing methods for causality detection, such as those based on Granger causality, can quantify the impact of a disruption on process performance. However, these methods do not consider the evolution of this impact over time or its relation to other process features that are also affected and may contribute to the performance decline. In contrast, a vector autoregressive model can address these challenges more effectively. It can analyze the dependencies between relevant process features in a single run, considering all features and their inter-relations simultaneously. This allows for a comprehensive prediction of the response in process performance under various disruption scenarios, taking into account the complex interactions between different process features.

7.2 Resilience assessment in other research fields

The problem of resilience assessment is linked to various established research fields. In the following discussion, we consider these fields and discuss their differences in comparison to business process resilience assessment.

Risk assessment. Risk assessment is a mature discipline that has been developed in the past 40 years to help understand and control the risk of accident events [52]. Many principles and methods are developed for how to conceptualize, assess, and manage risk [53], i.e., the potential occurrence of undesirable consequences for certain entities or situations.

Risk assessment and resilience assessment complement each other [52]. Risk assessment focuses on identifying and managing potential risks, while resilience assessment focuses on the ability to withstand and recover from disruptions. Together, they provide a comprehensive approach to understanding and mitigating threats, ensuring both the prevention of issues and the capability to recover when they occur. However, existing data-driven techniques for risk assessment cannot describe the ability of a process to withstand and recover from possible disruptions, but process resilience assessment can enhance awareness of potential risks in business process management.

Business continuity management. The continuity of a business process and process resilience are closely related concepts that complement each other, though they differ slightly in scope and focus. Business continuity is defined as the ability of an organization to continue delivering products or services at acceptable levels after a disruption [17]. It primarily aims at ensuring immediate operational stability by developing specific, tactical plans and procedures to prevent disruptive events from causing unexpected, unwanted interruptions in production or service activities [52]. Business continuity focuses on building processes and procedures to navigate a single disturbance by establishing alternative suppliers, maintaining backup inventory, and creating detailed plans for switching suppliers if the primary supplier is unavailable. In contrast, process resilience emphasizes the process itself, aiming for long-term strength and the ability to handle any number of disruptions, even unforeseen ones, by integrating flexible manufacturing practices that enable process activities to be executed under various conditions.

Statistical quality control. Statistical quality control is a collection of tools and techniques useful in achieving quality improvement by establishing process stability [54]. The enhancement of process quality stems from the reduction of variability in its outcomes, implying that a decrease in variability in the critical characteristics of a product results in an increase in product quality.

Process resilience assessment and statistical process control share some similarities but differ in scope. Statistical process control aims to maintain process performance within acceptable levels and alert management when deviations occur, typically utilizing control charts that display upper and lower bounds for a given performance indicator. In contrast, process resilience not only addresses deviations from desired levels but also seeks to predict the expected impact and recovery time in the event of disruptions.

Sensitivity analysis. Sensitivity analysis evaluates how variations in input parameters affect a model or system's output [55, 56]. It is used to identify critical factors influencing outcomes, quantify their influence, and inform decision-making by highlighting areas where improvements or interventions may be most effective.

Still, sensitivity analysis provides an incomplete view of the resilience assessment problem. Existing sensitivity analysis methods, e.g., Sobol sensitivity analysis [57] and PRIM (Patient Rule Induction Method) analysis [58], allow quantification of the impact of changes in one variable on another. In resilience assessment, this can be used to assess the maximal expected impact on process performance based on the initial size of the disruption. However, they can only be used to estimate the anticipated maximum impact of a disruption, whereas they do not provide insights into the impact of a disruption over time, such as the time-to-impact or recovery time.

8 Conclusion

This paper presented an approach for assessing process resilience using event log data. We measure process resilience in terms of the expected deviation of the process performance to disruptions in different process characteristics, such as arrival rate, active cases, and available resources. Our approach represents process characteristics as time series, constructs a vector autoregressive model that captures the statistical relationship between them, and conducts impulse-response analysis. Evaluation experiments demonstrate the accuracy and effectiveness of our approach in quantifying overall process resilience and its weak and strong points concerning different disruption types. The obtained insights help to understand the resilience of business processes, which is the first critical step towards achieving better resilience.

In future research, we plan to improve and further develop our work in several manners. First, our employed VAR models assume a linear relationship between the process features, which is not always the case. In our evaluation, we saw that the linear models generally performed well. Nevertheless, future research should investigate other VAR models, for instance, with time-varying coefficients or nonlinearity. Second, our approach employs the Akaike information criterion for selecting optimal model settings. Despite its acceptable results, this is a relatively simplistic strategy, whereas selecting (truly) optimal models is a complex task. Therefore, in the future, we aim to investigate state-of-the-art selection strategies that best suit the objective of resilience assessment using event data. Third, we aim to improve the statistical properties of the generated time series by accounting for, e.g., trends with drifts and seasonal effects, which may yield more accurate models. Finally, we aim to study how countermeasures can mitigate the expected negative impact of a disruption on a PPI. We seek to determine the appropriate timing and amount of these countermeasures to keep the negative impact within defined boundaries.

9 Statements and Declarations

Funding. No funding was received to assist with the preparation of this manuscript.

Competing Interests. The authors have no relevant financial or non-financial interests to disclose.

Code availability. The implementation, experimental details, and obtained raw results are available through the repository linked in Section 5.

References

[1] Sheffi, Y.: Resilience reduces risk. Logistics Quarterly **12**(1), 12–14 (2006)

[2] Annarelli, A., Nonino, F.: Strategic and operational management of organizational resilience: Current state of research and future directions. Omega **62**, 1–18 (2016)

[3] Müller, G., Koslowski, T.G., Accorsi, R.: Resilience-a new research field in business information systems? In: BIS, pp. 3–14 (2013). Springer

[4] Antunes, P., Mourão, H.: Resilient business process management: framework and services. ESWA **38**(2), 1241–1254 (2011)

[5] Dalziell, E.P., McManus, S.T.: Resilience, vulnerability, and adaptive capacity: implications for system performance. IFED (2004)

[6] Zahoransky, R.M., Brenig, C., Koslowski, T.: Towards a process-centered resilience framework. In: ARES, pp. 266–273 (2015). IEEE

[7] Marrella, A., Mecella, M., Pernici, B., Plebani, P.: A design-time data-centric maturity model for assessing resilience in multi-party business processes. Inf Syst **86**, 62–78 (2019)

[8] Zahoransky, R.M., Koslowski, T., Accorsi, R.: Toward resilience assessment in business process architectures. In: SafeComp, pp. 360–370 (2014). Springer

[9] Sanchis, R., Canetta, L., Poler, R.: A conceptual reference framework for enterprise resilience enhancement. Sustainability **12**(4), 1464 (2020)

[10] Hosseini, S., Barker, K., Ramirez-Marquez, J.E.: A review of definitions and measures of system resilience. Reliab. Eng. Syst. Saf. **145**, 47–61 (2016)

[11] Bhamra, R., Dani, S., Burnard, K.: Resilience: the concept, a literature review and future directions. Int. J. Prod. Res. **49**(18), 5375–5393 (2011)

[12] Haimes, Y.Y.: On the definition of resilience in systems. Risk Analysis: An International Journal **29**(4) (2009)

[13] Ungar, M.: Qualitative contributions to resilience research. Qualitative social work **2**(1), 85–102 (2003)

[14] Sarre, S., Redlich, C., Tinker, A., Sadler, E., Bhalla, A., McKevitt, C.: A systematic review of qualitative studies on adjusting after stroke: lessons for the study of resilience. Disability and rehabilitation **36**(9), 716–726 (2014)

[15] Muhammedamin Hussen saad, G.v.d.V. Geoffrey Hagelaar, Omta, S.W.F.: Conceptualization of smes' business resilience: A systematic literature review. Cogent Business & Management **8**(1), 1938347 (2021)

[16] Conz, E., Magnani, G.: A dynamic perspective on the resilience of firms: A systematic literature review and a framework for future research. European Management Journal **38**(3), 400–412 (2020)

[17] International Organization for Standardization: ISO 22301:2019: Security and resilience - Business continuity management systems.

[18] Erol, O., Henry, D., Sauser, B., Mansouri, M.: Perspectives on measuring enterprise resilience. In: IEEE Intern. Systems Conference, pp. 587–592 (2010)

[19] Erol, O., Sauser, B.J., Mansouri, M.: A framework for investigation into extended enterprise resilience. Enterprise Information Systems **4**(2), 111–136 (2010)

[20] Winkler, U., Gilani, W., Guitman, A., Marshall, A.: Models and methodology for automated business continuity analysis. In: 2012 IEEE 17th International Conference on Engineering of Complex Computer Systems, pp. 57–64 (2012)

[21] Stolker, R., Karydas, D., Rouvroye, J.: A comprehensive approach to assess operational resilience. In: Proceedings of the Third Resilience Engineering Symposium, pp. 247–253 (2008)

[22] Plebani, P., Marrella, A., Mecella, M., Mizmizi, M., Pernici, B.: Multi-party business process resilience by-design: a data-centric perspective. In: CAiSE, pp. 110–124 (2017). Springer

[23] del-Río-Ortega, A., Resinas, M., Cabanillas, C., Ruiz-Cortés, A.: On the definition and design-time analysis of process performance indicators. Information Systems **38**(4), 470–490 (2013)

[24] Dongen, B.F.: BPI challenge 2015. 4tu. Centre for Research Data. Dataset (2015)

[25] Adams, J.N., Zelst, S.J.v., Quack, L., Hausmann, K., Aalst, W.M., Rose, T.: A framework for explainable concept drift detection in process mining. In: Business Process Management, pp. 400–416 (2021). Springer

[26] Vidgof, M., Wurm, B., Mendling, J.: The impact of process complexity on process performance: A study using event log data. In: Business Process Management, pp. 413–429 (2023). Springer

[27] Van Der Aalst, W.M.P., Carmona, J. (eds.): Process Mining Handbook. Lecture Notes in Business Information Processing, vol. 448. Springer, Cham (2022)

[28] Brockwell, P., Davis, R.: Time Series: Theory and Methods. Springer, Berlin, Heidelberg (2009)

[29] Sims, C.: Macroeconomics and reality. Econometrica (1980)

[30] Kearney, C., Monadjemi, M.: Fiscal policy and current account performance: International evidence on the twin deficits. J. Macroeconomic **12**(2) (1990)

[31] Lütkepohl, H.: New Introduction to Multiple Time Series Analysis. Springer, Berlin, Heidelberg (2005)

[32] Krake, T., Klötzl, D., Hägele, D., Weiskopf, D.: Uncertainty-aware seasonal-trend decomposition based on loess. IEEE Transactions on Visualization and Computer Graphics (2024)

[33] Wasserman, L., Wasserman, L.: Bayesian inference. All of Statistics: A Concise Course in Statistical Inference, 175–192 (2004)

[34] Seabold, S., Perktold, J.: statsmodels: Econometric and statistical modeling with python. In: 9th Python in Science Conference (2010)

[35] Ljung, G., Box, G.: On a measure of lack of fit in time series models. Biometrika **65**(2) (1978)

[36] Bates, S., Hastie, T., Tibshirani, R.: Cross-validation: what does it estimate and how well does it do it? Journal of the American Statistical Association, 1–12 (2023)

[37] Korobilis, D.: Var forecasting using bayesian variable selection. Journal of Applied

Econometrics **28**(2), 204–230 (2013)

[38] Brüggemann, R., Krolzig, H.-M., Lütkepohl, H.: Comparison of model reduction methods for var processes (2002)

[39] Berti, A., Zelst, S.J., Aalst, W.: Process Mining for Python (PM4Py): Bridging the Gap Between Process- and Data Science (2019)

[40] Palmer, G.I., Knight, V.A., Harper, P.R., Hawa, A.L.: Ciw: An open-source discrete event simulation library. Journal of Simulation **13**(1), 68–82 (2019)

[41] Dumas, M., La Rosa, M., Mendling, J., Reijers, H.: Fundamentals of Business Process Management vol. 2. Springer, Berlin, Heidelberg (2018)

[42] Oberle, L., Aa, H.: DDPS: A Project Methodology for Data-Driven Process Simulation. In: AMCIS (2023)

[43] Camargo, M., Báron, D., Dumas, M., González-Rojas, O.: Learning business process simulation models: a hybrid process mining and deep learning approach. Information Systems **117**, 102248 (2023)

[44] López-Pintado, O., Halenok, I., Dumas, M.: Prosimos: Discovering and simulating business processes with differentiated resources. In: International Conference on Enterprise Design, Operations, and Computing, pp. 346–352 (2022). Springer

[45] Abel, M.: Lightning fast business process simulator. Master's thesis. Institute of Computer Science, University of Tartu (2011)

[46] Bezerra, F., Wainer, J., Aalst, W.M.: Anomaly detection using process mining. In: International Workshop on Business Process Modeling, Development and Support, pp. 149–161 (2009). Springer

[47] Ko, J., Comuzzi, M.: A systematic review of anomaly detection for business process event logs. Business & Information Systems Engineering **65**(4), 441–462 (2023)

[48] Sarno, R., Sinaga, F., Sungkono, K.R.: Anomaly detection in business processes using process mining and fuzzy association rule learning. Journal of Big Data **7**(1), 5 (2020)

[49] Bose, R., Aalst, W., Žliobaitė, I., Pechenizkiy, M.: Handling concept drift in process mining. In: CAiSE, pp. 391–405 (2011). Springer

[50] Hompes, B.F., Maaradji, A., La Rosa, M., Dumas, M., Buijs, J.C., Aalst, W.M.: Discovering causal factors explaining business process performance variation. In: Advanced Information Systems Engineering: 29th International Conference, CAiSE 2017, Essen, Germany, June 12-16, 2017, Proceedings 29, pp. 177–192 (2017). Springer

[51] Leemans, S.J., Tax, N.: Causal reasoning over control-flow decisions in process models. In: International Conference on Advanced Information Systems Engineering, pp. 183–200 (2022). Springer

[52] Zio, E.: The future of risk assessment. Reliability Engineering & System Safety **177**, 176–190 (2018)

[53] Aven, T.: Risk assessment and risk management: Review of recent advances on their foundation. European journal of operational research **253**(1), 1–13 (2016)

[54] Montgomery, D.C.: Introduction to Statistical Quality Control, 8th edn. John Wiley & Sons, Hoboken, NJ (2019)

[55] Saltelli, A., Tarantola, S., Campolongo, F., Ratto, M., *et al.*: Sensitivity Analysis

in Practice: a Guide to Assessing Scientific Models vol. 1. Wiley Online Library, Hoboken, NJ (2004)

[56] Ferretti, F., Saltelli, A., Tarantola, S.: Trends in sensitivity analysis practice in the last decade. Science of the total environment **568**, 666–670 (2016)

[57] Sobol, I.M.: Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates. Mathematics and computers in simulation **55**(1-3), 271–280 (2001)

[58] Friedman, J.H., Fisher, N.I.: Bump hunting in high-dimensional data. Statistics and computing **9**(2), 123–143 (1999)