



Discussion Paper

Process Forecasting: Towards Proactive Business Process Management

by

Rouven Poll, Artem Polyvyanyy¹, Michael Rosemann², Maximilian Röglinger, Lea Rupprecht

To be presented at: 16th International Conference on Business Process
Management, Sydney, Australia, September 2018

¹ The University of Melbourne, Parkville, VIC, 3010, Australia

² Queensland University of Technology, Brisbane, Australia

Process Forecasting: Towards Proactive Business Process Management

Rouven Poll¹, Artem Polyvyanyy², Michael Rosemann³,
Maximilian Röglinger⁴, and Lea Rupprecht¹

¹ FIM Research Center, University of Augsburg, Augsburg, Germany

² The University of Melbourne, Parkville, VIC, 3010, Australia

³ Queensland University of Technology, Brisbane, Australia

⁴ FIM Research Center, University of Bayreuth, Bayreuth, Germany

{rouven.poll, maximilian.roeglinger, lea.rupprecht}@fim-rc.de
artem.polyvyanyy@unimelb.edu.au, m.rosemann@qut.edu.au

Abstract. The digital economy is highly volatile and uncertain. Ever-changing customer needs and technical progress increase the pressure on organizations to continuously improve and innovate their business processes. The ability to anticipate incremental and radical process changes required in the future is a critical success factor. However, organizations often fail to forecast future business process designs and process performance. One reason is that Business Process Management (BPM) is dominated by *reactive* methods (e.g., lean management, traditional process monitoring), whereas there are only a few future-oriented approaches (e.g., process simulation, predictive process monitoring). This paper supports the shift towards *proactive* BPM by coining the notion of *process forecasting* – an umbrella concept for future-oriented BPM methods and techniques. We motivate the need for process forecasting by eliciting various types of process forecasting from BPM use cases and create a first understanding of its scope by providing a definition, a reference process, showing the steps to be followed in process forecasting initiatives, and a positioning against related BPM sub-areas. The definition and reference process are based on a structured literature review.

Keywords: Process forecasting, proactive business process management, predictive business process management.

1 Introduction

Business processes allow organizations to match existing customer demand with the supply of the resources needed to fulfil this demand. In the digital age, ever-changing customer needs and rapid technical progress cause high volatility and uncertainty. Such ongoing changes in market conditions force organizations to continuously adapt their business processes [1], which involves both the adaptation of resources to changes in the quantitative demand (e.g., the number of incoming customer orders) and the provision of business processes in respect to qualitative demand changes (e.g., customers seeking new digital channels to interact with organizations). The ability to timely anticipate incremental or radical changes of business processes required in the future is a critical success factor for organizations. However, the BPM state of the art does not provide sufficient tools to manage business processes proactively [2].

We argue that a more prevalent usage of future-oriented methods in BPM will lead to an improved and earlier understanding of future process demands and, thus, enable the timely implementation of required process changes [2]. Concretely, a widespread use of these methods could help shift the predominating focus on reactive BPM (e.g., lean management, process monitoring) towards proactive BPM. Proactive BPM is concerned with sensing process changes required in the future timely and effectively and implementing the identified changes before issues occur or opportunities are missed. For instance, as described in [3], instead of the reactive practice of spotting different types of waste, this would entail proactively identifying *waste-in-the-making* (e.g., emerging re-work) leading to an entire new discipline of *proactive lean management*.

The quality of proactive BPM can be measured by the extent to which it reduces *process latency*, i.e., the time of the occurrence of a process problem and its resolution, reducing the accumulated time during which a process is of unsatisfactory design or execution. The economic benefits of forecasting business processes can be seen in selective practices such as Amazon's predictive shipping where goods are delivered in anticipation that customers will order them. Done successfully, this leads to earlier demand satisfaction and revenue, and positive customer experience [4]. On the cost side, predictive maintenance approaches show how dynamically calculating emerging maintenance actions and embedding them into the production schedule minimizes costs related to significant replacements [5]. Predictive shipping and maintenance, however, are still isolated practices, and only a few future-oriented methods exist in BPM [2].

In this light, our paper aims to sensitize for the need for proactive BPM and to trigger a community-wide discussion on the use of forecasting elements into BPM. Thus, we seek to introduce process forecasting as a concept for gaining early insights into and anticipating future business processes by answering the following research questions: *(RQ1) What are use cases of forecasting in the context of BPM? (RQ2) How can process forecasting be defined? (RQ3) What are main steps of a process forecasting initiative?* Our answers to these questions resulted in three conceptual elements proposed in this paper: process forecasting types, definition, and reference process. The remainder of this paper is structured as follows. In Section 2, our methodological approach to develop the three conceptual elements is presented. In Section 3, we present the distinct types of process forecasting, demonstrating its wide range of use cases. After having motivated the need for process forecasting, Section 4 proposes the definition of process forecasting. In Section 5, we propose the reference process for process forecasting initiatives. After discussing the results of the conducted literature review in Section 6, Section 7 positions process forecasting against other BPM sub-areas. Finally, Section 8 concludes the paper by summarizing the findings and pointing to future work.

2 Research Method

To develop the three conceptual elements aiming to coin the notion of process forecasting, we applied the following methodological approach. Firstly, on the basis of BPM literature, we identified different types of process forecasting. Secondly, by means of a structured literature review, we defined process forecasting. Thirdly, we used the knowledge obtained from the literature review to adapt a well-accepted forecasting

reference process to the BPM domain, resulting in the proposed process forecasting reference process. Below, we describe our tripartite research method in detail.

Identifying the types of process forecasting. The different types of process forecasting describe use cases of forecasting in BPM. To identify them, we conducted an in-depth analysis of the BPM life cycle and BPM use cases proposed by Van der Aalst [6]. To do so, we mapped the BPM use cases to the phases of the life cycle. Then, each researcher independently analyzed the BPM use cases with regard to whether they can be supported by forecasting. The resulting unstructured collection of forecasting use cases in the BPM domain was then discussed and consolidated in a joint workshop.

Formulating the definition of process forecasting. With the aim to construct a well-founded definition of process forecasting, which comprises all identified types, we performed a structured literature review. Referring to the guidelines of Vom Brocke et al. [7], before conducting a literature search, the research scope needs to be defined. The topic of concern is the concept of process forecasting with its objective to predict process characteristics. To this end, in our literature review, we aimed for a comprehensive coverage of BPM-related research containing a forecasting component. Based on the terminology used in seminal publications [8–10], we formulated the following search phrases: "predict* [...] business process*", "forecast* [...] business process*", "business process forecast*", "business process prediction", "predictive business process monitoring" and "predictive process monitoring". We applied these to two scholarly databases, i.e., Scopus and Web of Science. Scopus is one of the largest abstract and citation database of peer-reviewed literature and includes scientific journals and books. To account for the fact that in computer science conferences are a significant publication outlet, we also used Web of Science, as this database – besides a large number of journals and books – covers over 180,000 conference proceedings [11]. Thus, a wide coverage of our literature review in information and computer science related topics is ensured. All studies containing at least one of the phrases in the title, keywords, abstract or (for Scopus only) in the full text of the paper were retrieved. Subsequently, we merged and filtered the retrieved papers, i.e., we removed duplicates, manuscripts not written in English, and not published as a journal article, book chapter, proceedings paper or as an article in press. From the resulting 120 papers, 56 were classified as relevant. For these papers, a forward and backward search was conducted, leading to a final set of 65 relevant papers. We selected papers that propose techniques, methods, or approaches supporting the early detection of process issues and opportunities, i.e., the prediction of future values of process characteristics. Examples for exclusions are papers addressing quality and complexity of process models, process discovery, and conformance checking. These were considered irrelevant as they do not contain a forecasting component. As we intended to construct a framework compiling possible input and output parameters of process forecasting, we then classified the methods proposed in the literature by their input and output. After that, we grouped the individual parameters found into categories. The resulting high-level categories of input and output parameters were then conceptualized and defined by means of appropriate literature.

Constructing the process forecasting reference process. To create a reference process to be followed in process forecasting initiatives, we drew from an accepted forecasting reference process (see [12]). The process we chose as our basis resembles other

reference processes in forecasting literature (see e.g., [13]). Making use of forecasting literature is reasonable here, because – in our understanding – process forecasting is a specific type of forecasting and as such, it should inherit its basic, domain-agnostic properties. To instantiate the reference process for BPM, based on the knowledge gained from the literature survey, we carried out domain-specific adaptations to each step.

3 Types of Process Forecasting

The BPM life cycle, as proposed by Van der Aalst [6], includes three phases. In the phase *(re-)design*, a process model is designed. The phase *implement/configure* refers to making a process model executable. Finally, the phase *run and adjust* is concerned with process execution. In the context of the phase *run and adjust*, Van der Aalst [6] emphasizes the need for analysis of expected and past performance, and monitoring of processes (see use cases “analyze performance based on model”, “analyze performance using event data”, and “monitor” [6]). By combining these use cases and taking a future-oriented perspective, we obtain the first type of process forecasting, viz., “solving the execution problem”, which addresses the predictive monitoring and simulation of processes at or shortly before run-time. The second type, “solving the configuration problem”, relates to the phase *implement/configure* and adds value when multiple process model variants exist one of which needs to be selected prior to process execution (see use case “configure configurable model” [6]). Thereby, process forecasting enables to anticipate which model is best suited for upcoming process executions, accounting for the future states of the operating environment. The third type, “solving the design problem”, relates to the phase *(re-)design* and supports the demand-driven design and improvement of process models (see use cases “design model” and “improve model” [6]). Here, the use case of process forecasting lies in predicting how processes need to be designed to comply with future process demands. The proposed types of process forecasting – visualized in Fig. 1 – are explained below via illustrative examples.

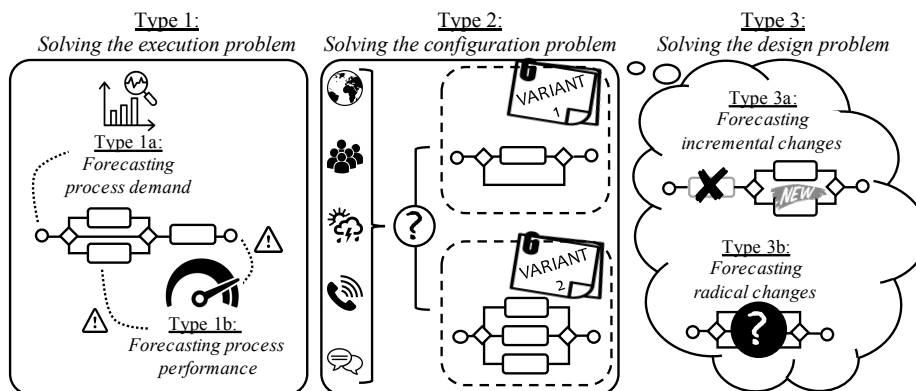


Fig. 1. Visualization of the three types of process forecasting.

Type 1: Solving the execution problem. This type of process forecasting is concerned with anticipating process-related issues of day-to-day operations. This involves two

forecasting tasks: forecasting process demand, i.e., the number and type of instances arriving in a future time period (Type 1a) (see e.g., [14]) as well as forecasting the expected performance of running and future process instances (Type 1b). The latter involves the prediction of performance indicators (e.g., cycle time) (see e.g., [15]), the process outcome (e.g., the probability of violating business constraints) (see e.g., [8]) as well as the sequence of activities (see e.g., [16]). When focusing on running instances, this may also include forecasting the next activities to be executed. An illustrative example of solving the execution problem can be inferred from [17], where the authors describe a retailer deploying forecasting techniques in order to prevent its stores from running out of stock. In the first step, the retailer may predict how many and which products will be sold in each store to anticipate the most cost-effective point in time when replenishments are needed (Type 1a). This forecast might be derived by taking into account seasonality aspects (e.g., higher demand for certain products shortly before Christmas) or consumer trends. After having ordered the replenishments, the retailer might be interested in the probability of delayed deliveries (Type 1b). As shown in [17], taking contextual information such as future weather conditions and their impact on transport routes into account, the retailer may detect delays before they occur. This enables proactive rescheduling of transport routes to prevent the upcoming delays.

Type 2: Solving the configuration problem. This type of process forecasting problem exists when a concrete model from some configurable process model needs to be created (see use case “Configure Configurable Model” in [6]), i.e., when an organization needs to select between alternative process model variants based on contextual variables (e.g., time, location, weather). In this regard, process forecasting can help anticipate the process model variant that is needed at a certain future point in time. This enables organizations to better prepare the execution of a process model variant and, thus, reduces process latency. For example, Rosemann and Recker [18] describe an insurance company that has designed process variants for lodging insurance claims based on different levels of severity of storms during the Australian storm season. As soon as a storm occurs, its severity is evaluated and the execution of the corresponding variant of the process model is triggered. In this case, process latency could be reduced by taking forecasts of the severity of storms into account to predict the process variant that is needed and proactively initiate targeted measures.

Type 3: Solving the design problem. The design of a business process is driven by the requirements assigned to the respective process [19]. Here, the use case of process forecasting lies in anticipating the changes in process models that will occur or be demanded in the future. Thereby, in line with the common view in business process improvement literature [20], we distinguish between incremental and radical changes. Whereas incremental changes are adaptations of existing process models (Type 3a), radical changes address the creation of entire new models (Type 3b). An example setting can be derived from [21]. This work describes a bank that aims to decide on the channels that should be offered to customers to conduct their banking activities in the future. As analyzed in [21], the customer use of a certain channel depends on customers’ intrinsic attributes (e.g., attitude towards technology or age of customers). Knowing this and building on information about future changes in customer characteristics, process forecasting could discover which channels should or should not be offered in the future.

For instance, at a certain future point in time, the fraction of customers using telephone banking might be forecasted to decrease considerably, resulting in the recommendation to shut down the telephone service (Type 3a). This would enable the bank to timely initiate associated actions such as cancelling contracts with external service providers and to plan the re-allocation of resources. Going one step further, process forecasting could also predict entire new process models for conducting banking activities (Type 3b). For instance, accounting for the rapidly expanding usage of virtual voice assistants, forecasting techniques could be able to predict the point in time when money transfers via virtual voice assistants are desired, affordable, and viable. Additionally, by learning from related process model designs (e.g., from other industries having implemented voice-based interactions into their process models), an algorithm could output the bank’s future process model supporting money transfers via virtual voice assistants.

4 Definition of Process Forecasting

The use cases outlined in Section 3 showed that process forecasting types differ in their objective, input, output, and time horizon. However, all types pursue the same overarching goal, namely to boost organizational preparedness for future business processes. As a result of the conducted literature review and the identified types of process forecasting, we define process forecasting as an *umbrella concept for BPM methods and techniques that aim to predict future business process demands, performance, and designs*. For the purpose of our research, we deliberately refrain from developing a specific process forecasting method. Rather, we structure the field of action by proposing a framework of relevant input and output parameters. The specific set of input variables used to derive forecasts primarily depends on the applied forecasting method [12]. The method to be used, in turn, might be constrained by data availability [13]. As the literature review showed, process forecasts can generally be achieved by combining or extrapolating data from past and running process executions (see e.g., [8]). More sophisticated are context-aware process forecasts that take the future process environment into account (see e.g., [10]). A schematic view of the input and output parameters for applying process forecasting methods is shown in Fig. 2.

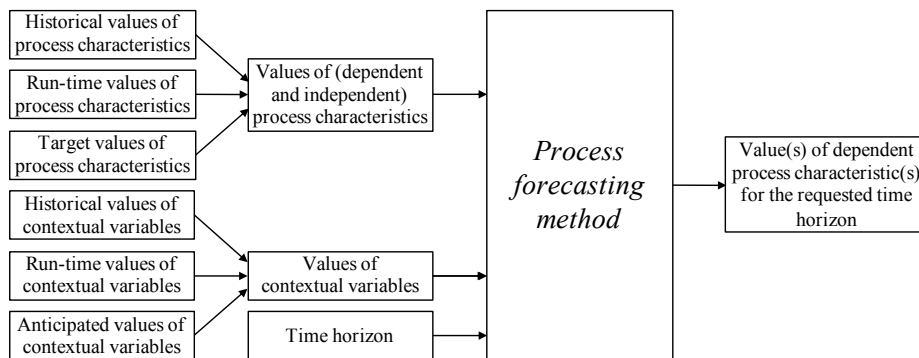


Fig. 2. Schematic view of relevant input and output parameters of process forecasting.

As can be inferred from Fig. 2, a process forecasting method aims to predict the values of one or more dependent process characteristics for a given time horizon based on available historical, run-time, and target values of (independent and dependent) process characteristics as well as historical, run-time, and anticipated values of contextual variables. Thereof, the time horizon is a mandatory input, whereas the other input variables depend on the problem to be solved. Next, we define all the proposed components.

Process characteristics include both *process performance indicators* (PPIs) and the typical *core elements* of business processes. PPIs are measures of the critical success factors of business processes such as cycle time or cost. PPIs can be defined over a single process instance or a group of instances (e.g., instances occurring within a period of time) [22]. The core elements of a business process are essential to its execution and understanding [23]. Referring to de Leoni et al. [24], we differentiate core elements into *data* (e.g., data required to execute a process), *resources* (e.g., the resource performing or supporting a particular activity), *time* (e.g., the duration of an activity), and the *control flow*, i.e., the executed activities and their temporal and logical relationship. The control flow of a process can, for example, manifest itself in the form of an *event log*, i.e., a collection of sequences of observed and recorded events, a *simulation model*, i.e., a conceptual model of a collection of processes with a finite imitation of its operations, or an ordinary *process model*, i.e., a conceptual model of a collection of processes [25]. Process characteristics can be classified as dependent or independent, where the former is the process characteristic to be forecasted and the latter is any other process characteristic taken into account to derive the forecast. The selection of dependent process characteristics primarily hinges on the forecasting problem to be solved and ranges from PPIs, through involved resources, to event logs and process models.

Contextual variables describe the environment in which a business process operates. With regard to the classification presented by Rosemann et al. [23], we distinguish internal, external, and environmental context. Internal context involves the internal environment of an organization having an indirect impact on a business process (e.g., policies, resource capacity, and corporate strategy). External context captures factors beyond an organization's control sphere but within its business network (e.g., characteristics of suppliers and customers, industry-specific factors such as trends driving the demand for an industry's service, and regulations). Environmental context is the environment beyond the business network in which an organization is embedded (e.g., weather, seasonality, and political system).

Time horizon is the period of time for which a forecast is produced. By transferring the classification suggested in general forecasting literature [12], we categorize time horizons into short-, medium- and long-range horizons. In the context of process forecasting, based on the time horizons found in the related literature, we define short-range forecasts such that they cover the prediction of process characteristics of running instances (e.g., forecasting the remaining duration of a running process execution). Medium-range forecasts are based on weekly or monthly time spans from now (e.g., how many employees will be required next Monday to serve arriving customers?). Finally, long-range forecasts cover a (multi-)annual time span (e.g., forecasting the future process model in one year from now).

5 Process Forecasting Reference Process

Below, we present the process forecasting reference process, describing the basic steps to be followed in process forecasting initiatives. Fig. 3 provides an overview of the proposed six steps. In the remainder of this section, we describe each step in detail.

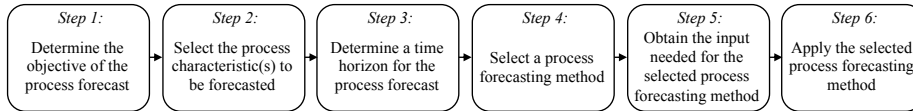


Fig. 3. Process forecasting reference process.

Step 1: Determine the objective of the process forecast. In the first step, the objective of the forecast needs to be defined. This involves questions like “Why is the forecast useful?”, “In which way is the forecast used?”, and “Who needs the forecasting results?”. The overarching goal of process forecasting is to detect and proactively manage process-related issues and opportunities. As can be inferred from the identified types of process forecasting, the instantiations of this objective are manifold. They range from achieving an adequate resource or materials planning and ensuring flawless process executions (solving the execution problem) to reducing process latency (solving the configuration problem), or discovering how future process models will look like (solving the design problem). The defined objective further determines the level of accuracy and the level of aggregation required in the forecast, i.e., whether the forecasting task focuses on predicting single process instances, an aggregation of multiple process instances occurring in a certain time interval, or on predicting complete process models. In general, it is agreed in literature that forecasting for groups (of process instances) is more accurate than for individuals [12]. Moreover, when focusing on a single process instance, one should be aware that this can result in losing sight of the totality of ongoing process executions. For instance, taking actions to prevent a single instance from failing may have negative and unanticipated effects on other running process instances.

Step 2: Select the process characteristic(s) to be forecasted. After determining why to forecast, the next step is to define what to forecast. The selection of the dependent process characteristic(s) depends on the objective of the forecast. For instance, an organization that aims to solve the design problem, e.g., to discover how a process needs to be designed to satisfy future customer needs, should select the future process model as prediction target. In contrast, for an organization interested in an estimation of how a process will perform in the future (solving the execution problem), the prediction of one or several PPIs might be expedient. For instance, in [15], the prediction of cycle time is emphasized, as this enables organizations to provide customers with waiting times guarantees. These were shown to increase customer satisfaction, if they are met [15]. When selecting a dependent process characteristic, analysts should consider that the explanatory power of a forecast can be increased by choosing leading indicators, i.e., the key factors that are known to influence unwanted changes of a certain higher-level process characteristic (lagging indicator), as the prediction target [24]. Leading indicators can be defined on different levels. For instance, human resources involved in a process can be defined as leading indicator for the lagging indicator cycle

time (e.g., due to different levels of experience of the employees) [24]. Cycle time, in turn, is a leading indicator for the lagging indicator process cash flow. Based on the executed literature review, we obtained the following relationship: the lower the level of a leading indicator, the higher the forecast's explanatory power. Thus, adhering to the example sketched above, forecasting the employees that will be assigned to work on the process of interest (instead of forecasting the cycle time), may reveal additional information, as it enables analysts to not only detect how but also why a process will perform in a certain manner. More of these cause-effect relationships between process characteristics were discovered as part of the literature review and are shown in Fig. 5.

Step 3: Determine a time horizon for the process forecast. As a third step, the length of time on which a forecast is based and how far into the future the forecast is generated, needs to be determined. Thereby, it should be taken into account that, in general, the accuracy of forecasts decreases as the time horizon extends [12]. However, short-range forecasts provide decision makers only with little time to act. The earlier an issue is detected, the more can be done to proactively solve it [2]. For example, when solving the execution problem, let us assume that a group of instances of a process is forecasted to be delayed. Detecting this within a short-range forecast, i.e., when the instances are already running, enables decision makers to shift resources working on other processes to the forecasted process. This might eliminate the delay of the forecasted process, but in turn lead to an increased cycle time of the other processes running simultaneously. In contrast, a medium-range forecast might have provided decision-makers with enough time to make resources available without affecting the performance of other processes. When solving the design problem (e.g., implementing a new process), only long-range forecasts may be expedient, as strategical decisions of this sort generally are subject to long lead-times. Further, regarding forecasts that rely on historical data, the decision on a time horizon should be taken in consideration of the time period for which historical data is available. It can be expected that a short observation period leads to less accurate long-range forecasts than a multi-annual one [12].

Step 4: Select a process forecasting method. In the next step, the method for process forecasting needs to be selected. Depending on the characteristics of the problem at hand (e.g., objective of the forecast, type of dependent process characteristic, and data availability), the applicability and suitability of distinct forecasting methods should be evaluated. The chosen forecasting method then determines the set of potential input variables and the way they are processed [12]. As can be inferred from the diversity of methods proposed in the relevant literature, process forecasting is not limited to specific methods, i.e., statistical and judgmental forecasts as well as a combination can be applied equally well. Whereas statistical methods make use of historical data and, thus, underlie the assumption that observed dependencies will continue in the future, judgmental methods such as panel approaches or Delphi studies are based upon opinions of experts [12]. Expert judgments are particularly helpful when historical data is unavailable or unable to "explain" the future properly. This may, for instance, be the case when predicting sales for an entire new product or forecasting radical changes of process models (cf. Type 3b, Section 3). Further approaches such as planning algorithms (see e.g., [19]) or cognitive computing (see e.g., [26]) are also conceivable for certain forecasting problems. The methods most commonly used in the literature related to process

forecasting – especially in the field of predictive process monitoring – are based on machine learning (increasingly deep learning), constraint satisfaction, and quality-of-service aggregation [27]. Thus, the majority of existing works focuses on statistical forecasts, aiming to learn from past as well as present dependencies between process characteristics and to transpose this knowledge into the future (see e.g., [16, 24]).

Step 5: Obtain the input needed for the selected process forecasting method.

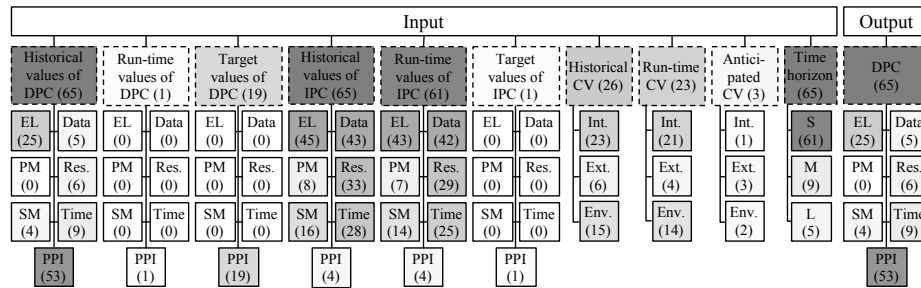
After having selected a method, in the next step, the required input needs to be collected. In most cases, forecasts are based on large amounts of data. Against the backdrop of the recent uptake of new methods such as deep learning, analysts are enabled to draw on structured as well as unstructured data (e.g., in the form of images, voice, and videos). Further, the rising availability of micro-grained data about historical and ongoing business process executions, particularly in the form of process logs, pushes the boundaries of data that can be taken into account for a process forecasting task. As pictured in Fig. 2, both *process characteristics* and *contextual variables* should be collected and used as input for a process forecast. The exploitation of the relationship between historical values of the dependent process characteristic and all other *process characteristics* is motivated in [24]. As an example, de Leoni et al. [24] mention the possible dependency between involved resources and customer satisfaction, i.e., certain resources involved in a process execution may lead to a lower customer satisfaction. Besides historical and run-time values, target values of *process characteristics* in the form of future process requirements (e.g., service level agreements that have to be met) may also be necessary to forecast certain prediction targets. Further, the literature suggests taking *contextual variables* into account. Such data is agreed to have a high explanatory power on process behavior [10]. For example in [24], the consideration of the context variable *weather* as input to forecast the process characteristic *activity duration* is emphasized, assuming that certain resources, which in turn have an impact on the activity duration, are more efficient when the weather is good. Here, besides considering past relationships between context and process executions, it is also conceivable to take future context data (e.g., the weather forecast for next week) into account. Linking this information with the dependency patterns learned in the past may increase the forecast’s accuracy. Further, depending on the type of the process forecast, certain data may be mandatory. For instance, the configuration problem can only be solved, if the set of process model variants as well as their extrinsic trigger points are available.

Step 6: Apply the selected process forecasting method. Having implemented all the previous steps, as a final step, the process forecasting method is applied.

6 Results of the Structured Literature Review

Below, we describe the results of the structured literature review. These built the theoretical backbone of the process forecasting definition and reference process. Following the research approach described in Section 2, we classified the methods found in the literature with respect to their input and output. Fig. 4 shows the components (dashed borderline) of the proposed definition of process forecasting. These components are divided into sub-components (full borderline) based on their conceptualizations

described in Section 4. The numbers in brackets indicate the number of methods using the (sub-)components, with the shadings of gray indicating whether a (sub-)component is used in many (dark gray) or few (light gray) methods.¹



Abbreviations: dependent process characteristic(s) (DPC), independent process characteristic(s) (IPC), contextual variables (CV), event log (EL), process model (PM), simulation model (SM), resource (Res.), process performance indicator(s) (PPI), internal context (Int.), external context (Ext.), environmental context (Env.), short-range (S), medium-range (M), long-range (L).

Fig. 4. Heat map of the (sub-)components of process forecasting.

The results show that all of the framework's components are addressed at least once in the literature, whereby none of the existing techniques exploits all components. Regarding the sub-components, the analysis reveals that forecasting a process model has not been addressed and run-time values of the dependent process characteristic(s) as well as target values of independent process characteristic(s) have only been considered by one existing technique each. This is not surprising, as taking run-time values of the dependent process characteristic(s) into account becomes relevant for medium- and long-range time horizons. The majority of the identified and analyzed techniques, however, focus on a short-range time horizon, i.e., on the prediction of running instances. The results disclose that most of the proposed techniques are data-driven. Among the data-driven forecasting methods, a large number uses historical and run-time values from event logs, data attributes assigned to activities or processes, and resources involved in a process execution as independent process characteristics. Concerning the selection of the dependent process characteristic(s), PPIs, particularly the cycle time, have received most attention, whereas the prediction of activity-specific attributes (e.g., involved resources) is fragmentarily addressed. More than a third of the analyzed approaches exploit historical and run-time contextual data to potentially lever the accuracy of the forecasts, among which the lion's share considers the internal context. In contrast, anticipated contextual data has been used rarely.

Most existing techniques focus on the prediction of PPIs. However, the literature emphasizes that the explanatory power of forecasts can be increased by selecting leading instead of lagging indicators as prediction targets. This insight was also integrated into the proposed reference process (cf. Step 2, Section 5). In the analyzed literature, numerous causalities between process characteristics are mentioned. Fig. 5 schematizes the knowledge scattered across the literature by means of an acyclic directed graph.

¹ The individual classification of the methods and further details on the results of the literature review are available at [researchgate.net/publication/323691573_Process_Forecasting](https://www.researchgate.net/publication/323691573_Process_Forecasting).

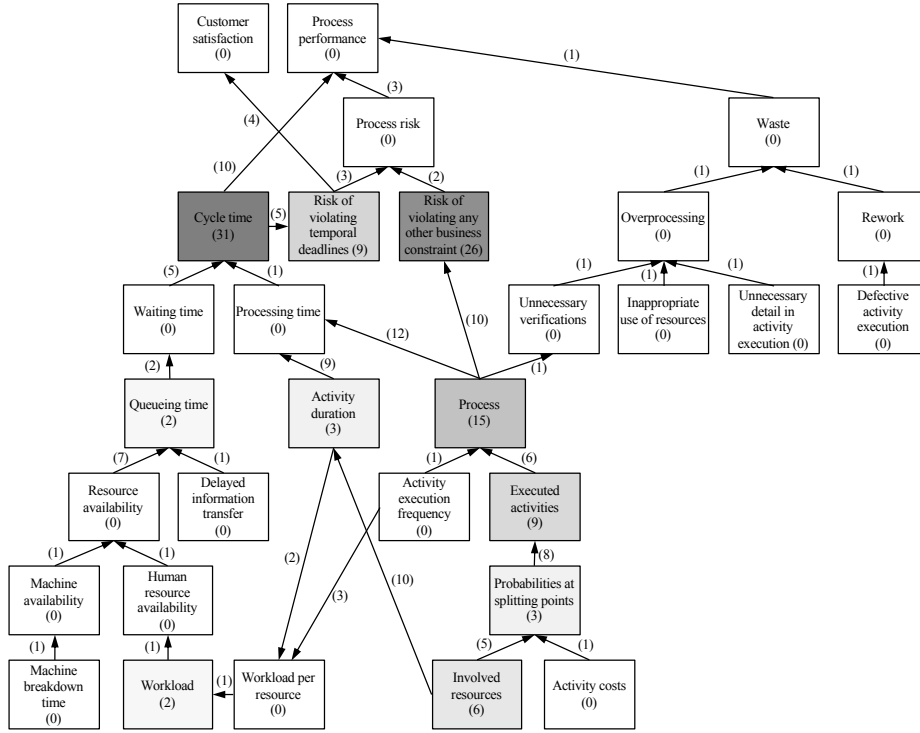


Fig. 5. Cause-effect relationships between process characteristics.

The graph illustrates which process characteristics are addressed (this involves either using the process characteristic as a prediction target or highlighting a process characteristic as leading indicator for another process characteristic) and which causal relationships are proposed. In the figure, nodes are process characteristics and edges describe the cause-effect relationships between process characteristics (e.g., “A → B” describes “A contributes to changes in B.”). In addition, numbers in the labels of nodes and edges indicate the number of papers selecting a certain process characteristic as prediction target and confirming a certain cause-effect relationship. As in Fig. 4, the shadings of nodes indicate whether a process characteristic is frequently (dark gray) or rarely forecasted (light gray). A number of zero in the label of a node indicates that this process characteristic was not explicitly forecasted in the analyzed techniques. However, it is included in the figure, as it was mentioned in at least one cause-effect relationship. Given a forecasting objective, the graph can guide an analyst through the selection of the dependent process characteristic (cf. Step 2, Section 5) or the identification of influential independent process characteristics needed to make an accurate forecast (cf. Step 5, Section 5). The further analysts get down the causality chain, the closer they will get to the leading indicators of future process behavior. For instance, let an organization be interested in the cycle time of future process executions. Separately predicting the processing and waiting times of future process executions may further localize the problem. An even higher explanatory power of the forecast can be achieved

by predicting process characteristics located at the end of a causality chain, e.g., the breakdown time of a machine, as this may reveal that the long waiting times result from the poor condition of a machine involved in future process executions. This can enable the company to proactively take targeted actions and eliminate issues before they occur.

7 Related Work

Due to its interdisciplinary nature, process forecasting is linked to many disciplines beyond BPM (e.g., operations management, demand forecasts, pervasive computing). To position process forecasting as BPM sub-area, we compare it with other BPM sub-areas, namely process flexibility, declarative process modeling, emergent workflow, business process intelligence, process mining, predictive process monitoring, process simulation, and process planning. The positioning is shown in Fig. 6. For the sake of transparency, we deliberately abstract from overlaps between other sub-areas.

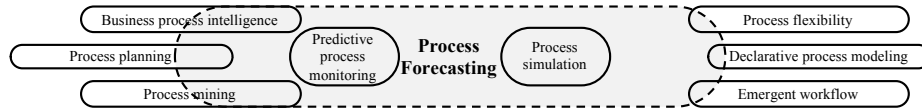


Fig. 6. Positioning of process forecasting against BPM-related sub-areas.

Process flexibility enables adapting processes to internal or external triggers without completely replacing them [18]. Flexibility approaches usually do not take a forecasting perspective. Rather, they are concerned with how to quickly react to changes. However, linking extant flexibility types [1] to our process forecasting types reveals that process flexibility is an enabler for process forecasting, because organizations can only benefit from process forecasting, if processes can be adapted flexibly. For instance, the strategy *flexibility by deviation*, i.e., allowing for short-term deviations, enables solving the execution problem. *Flexibility by design*, based on many model variants, builds the foundation for solving the configuration problem. *Flexibility by change* supports solving the design problem in the short-term, as it enables changing process models at run-time. The fourth strategy, *flexibility by under-specification*, allows for the formulation of incomplete process models at design time and the addition of process model fragments at run-time. As such, it helps solve the execution and design problem.

Next, *declarative process modeling* implements flexibility by design and under-specification in a non-procedural way. Thus, it also serves as a tool for realizing process forecasting and enables solving the execution, configuration, and design problem. The basic idea is to model processes via constraints that must be satisfied by every process instance, instead of rigorously defining the control flow [1]. Consequently, more options to proactively manage process-related issues and opportunities are created.

Similarly, *emergent workflow* follows the idea to design or adapt process models at run-time, if needed [28]. This creates the ability to flexibly react to changes or issues that are anticipated by means of process forecasting.

Business process intelligence (BPI) refers to the application of techniques that support the collection, analysis, and presentation of business process information with the aim to enable better decision-making [29]. In contrast to process forecasting, however,

most BPI techniques target retrospective analyses, i.e., process instances are analyzed after their termination providing insights into how processes are executed. Among these techniques, an application area that receives a lot of attention is process mining.

Process mining aims to extract knowledge from process logs [30]. As such, it sets the basis for process forecasting methods that rely on process data. As opposed to its predominant focus on understanding past process behavior, in recent years, proactive process mining approaches emerged that support real-time and future-oriented analyses. The majority deals with the performance prediction of running instances and, thus, relates to the execution problem of process forecasting (cf. Type 1b, Section 3). This research field is commonly referred to as predictive process monitoring.

Predictive process monitoring is concerned with predicting how running process instances will unfold up to their completion [16]. In essence, existing approaches use past process execution data, partial traces of the monitored process instances, and partly also contextual data to predict the remaining duration (see e.g., [15]), process-related risks, i.e., the likelihood or severity of a process fault [9], the outcome (e.g., whether or not a running instance will violate a compliance rule [8]), or the future path of a process instance [16]. With the aim to combine the approaches for different but related prediction tasks, a general framework to predict process characteristics of running instances was proposed in [24]. Differently from the above-mentioned works, process forecasting goes beyond the prediction of individual running instances. Predicting process characteristics of a group of running and/or future instances arriving in a distinct time interval is hardly addressed in literature [14].

Further, *process simulation* evaluates the impact of design decision on business processes prior to implementation [31]. Thus, process simulation is a part of process forecasting, as it is concerned with understanding dependencies between process characteristics. Once understood, the discovered dependencies help derive process forecasts.

Finally, *process planning* deals with the automated construction of new process models. It facilitates the design of process models that comply with future process demands [19]. Consequently, process planning is linked to solving the design problem of process forecasting, as it tackles the same issue, i.e., to reduce process latency by enabling a timely preparation of future process models. Thus, process planning can be seen as a specific instantiation of a process forecasting method that uses future process demands as input, namely in the form of target values of independent process characteristics, and forecasts the design of a process model that is conform with these demands.

As summarized above, there are several heterogeneous approaches pursuing the same overarching objective, namely to proactively manage process-related issues and opportunities. In particular, the methods focusing on the prediction of process characteristics differ in terms of the time horizon of the prediction, the input variables taken into account, the process characteristics to be forecasted, their level of aggregation, and the techniques to be used. Process forecasting, as an umbrella concept, consolidates and extends these existing future-oriented BPM methods and reveals that there is a considerable need for future research.

8 Conclusion

In this paper, we proposed the concept of process forecasting. With the aim to stimulate a community-wide discussion about the uptake of proactive BPM, we presented various use cases that motivate the need for process forecasting. We also provided a definition and a reference process as well as a positioning against related BPM sub-areas. Process forecasting is an umbrella concept for future-oriented BPM methods and techniques that tackle process-related issues before they occur, proactively seize process-related opportunities, and reduce process latency. The proposed definition and the reference process were built on the results of a structured literature review.

This paper revealed that the range of existing future-oriented BPM methods and techniques is not sufficient. Whereas the first type of process forecasting, i.e., solving the execution problem, has been addressed by research fields such as predictive process monitoring, forecasting techniques that solve the configuration or design problem hardly exist. This strengthens the need for future research. Further, the limitations of this paper also stimulate future research. As research on process forecasting is still in its early stages, our rationale for this paper was to create a first overall understanding of the need for and scope of process forecasting. With a focus on the interdisciplinary nature of the topic, we acknowledge that the scope of the conducted literature review should be broadened. In future work, we intend to enrich process forecasting by exploiting extant methods from other disciplines. Another avenue for future research is to provide methodological guidance for choosing a suitable forecasting technique within a process forecasting initiative. In addition, we plan to conduct case studies validating the applicability of our concept, to set up a research agenda for process forecasting, and to devise a proactive BPM life cycle. However, we trust that this paper is a solid starting point for discussing and exploring the under-researched potentials of proactive BPM.

References

1. Schonenberg, H., Mans, R., Russell, N., Mulyar, N., Van Der Aalst, W.: Process flexibility: A survey of contemporary approaches. *LNBIP*. 10, 16–30 (2008).
2. Krumeich, J., Werth, D., Loos, P.: Enhancing Organizational Performance through Event-based Process Predictions. *Am. Conf. Inf. Syst.* (2015).
3. Verenich, I., Dumas, M., La Rosa, M., Maggi, F.M., Di Francescomarino, C.: Minimizing Overprocessing Waste in Business Processes via Predictive Activity Ordering. *Int. Conf. Adv. Inf. Syst. Eng.* (2016).
4. Leveling, J., Edelbrock, M., Otto, B.: Big data analytics for supply chain management. In: *IEEE Int. Conf. on Industrial Engineering and Engineering Management*. (2014).
5. Selcuk, S.: Predictive maintenance, its implementation and latest trends. *J. Eng. Manuf.* 231, 1670–1679 (2017).
6. van der Aalst, W.M.P.: *Business Process Management - A Comprehensive Survey*. *ISRN Softw. Eng.* (2013).
7. Vom Brocke, J., Simons, A., Niehaves, B., Reimer, K., Cleven, A.: Reconstructing the Giant: On the Importance Of Rigour In Documenting The Literature Search Process. In: *ECIS*. pp. 2206–2217 (2009).
8. Maggi, F.M., Di Francescomarino, C., Dumas, M., Ghidini, C.: Predictive monitoring of business processes. *LNCS*. 8484, 457–472 (2014).
9. Conforti, R., De Leoni, M., La Rosa, M., Van Der Aalst, W.M.P., Ter Hofstede, A.H.M.: A recommendation system for predicting risks across multiple business process instances. *DSS*

- 69, 1–19 (2015).
10. Folino, F., Guarascio, M., Pontieri, L.: Discovering Context-Aware Models for Predicting Business Process Performances. In: OTM 2012. pp. 287–304 (2012).
 11. Clarivate Analytics: Web of Science Databases, <https://clarivate.com/products/web-of-science/databases/>. Accessed 17 March 2018.
 12. Heizer, J., Render, B., Munson, C.: Principles of Operations Management: Sustainability and Supply Chain Management. Pearson Education (2016).
 13. Reid, R.D., Sanders, N.R.: Operations Management: An Integrated Approach. Wiley (2010).
 14. Folino, F., Guarascio, M., Pontieri, L.: A Prediction Framework for Proactively Monitoring Aggregate Process-Performance Indicators. In: IEEE Int. Enterprise Distributed Object Computing Conf. pp. 128–133 (2015).
 15. Rogge-Solti, A., Weske, M.: Prediction of business process durations using non-Markovian stochastic Petri nets. *Inf. Syst.* 54, 1–14 (2015).
 16. Di Francescomarino, C., Ghidini, C., Maggi, F.M., Petrucci, G., Yeshchenko, A.: An Eye into the Future: Leveraging A-priori Knowledge in Predictive Business Process Monitoring. In: BPM 2017. pp. 252–268 (2017).
 17. Metzger, A., Franklin, R., Engel, Y.: Predictive monitoring of heterogeneous service-oriented business networks: The transport and logistics case. In: SRII Global Conf. (2012).
 18. Rosemann, M., Recker, J.C.: Context-aware Process Design: Exploring the Extrinsic Drivers for Process Flexibility. In: CAiSE 2006. pp. 149–158 (2006).
 19. Heinrich, B., Klier, M., Zimmermann, S.: Automated planning of process models: Design of a novel approach to construct exclusive choices. *DSS* 78, 1–14 (2015).
 20. Childe, S.J., Maull, R.S., Bennett, J.: Frameworks for Understanding Business Process Re-engineering. *Int. J. Oper. Prod. Manag.* 14, 22–34 (1994).
 21. Sousa, R., Amorim, M., Rabinovich, E., Sodero, A.C.: Customer Use of Virtual Channels in Multichannel Services: Does Type of Activity Matter? *Decis. Sci.* 46, 623–657 (2015).
 22. del-Río Ortega, A., Resinas, M., Ruiz-Cortés, A.: Defining Process Performance Indicators: An Ontological Approach. In: OTM 2010. pp. 555–572 (2010).
 23. Rosemann, M., Recker, J.C., Flender, C.: Contextualization of Business Processes. *Int. J. Bus. Process Integr. Manag.* 3, 47–60 (2008).
 24. De Leoni, M., Van Der Aalst, W.M.P., Dees, M.: A general process mining framework for correlating, predicting and clustering dynamic behavior based on event logs. *Inf. Syst.* 56, 235–257 (2016).
 25. Polyvyanyy, A., Ouyang, C., Barros, A., van der Aalst, W.M.P.: Process querying: Enabling business intelligence through query-based process analytics. *DSS* 100, 41–56 (2017).
 26. Röglinger, M., Seyfried, J., Stelzl, S., Zur Muehlen, M.: Cognitive Computing: What’s in for Business Process Management? An Exploration of Use Case Ideas. In: BPM Workshops (2017).
 27. Metzger, A., Leitner, P., Ivanovic, D., Schmieders, E., Franklin, R., Carro, M., Dustdar, S., Pohl, K.: Comparing and Combining Predictive Business Process Monitoring Techniques. *IEEE Trans. Syst. Man, Cybern. Syst.* 45, 276–290 (2015).
 28. Jorgensen, H., Carlsen, S.: Emergent Workflow: Planning and Performance of Process Instances. In: Workflow Management (1999).
 29. Castellanos, M., Alves De Medeiros, A.K., Mendling, J., Weber, B., Weijters, A.J.M.M.: Business Process Intelligence. In: Handbook of research on business process modeling. pp. 456–480 (2009).
 30. van der Aalst, W.M.P.: Process Mining: Discovery, Conformance and Enhancement of Business Processes. (2011).
 31. Jansen-Vullers, M., Netjes, M.: Business Process Simulation - A Tool Survey. In: Workshop and Tutorial on Practical Use of Coloured Petri Nets and the CPN Tools (2006).