From Process Mining to Process Design: a Simulation Model to Reduce Conformance Risk

Piera Centobelli, Giuseppe Converso, Mosè Gallo, Teresa Murino, and Liberatina Carmela Santillo

Abstract— **An operator's mistakes are frequently attributed to inexperience, which leads to a superficial management of these kinds of 'accidental' errors. The weakness of this traditional approach is that it isolates the wrong actions from the general systematic context, overlooking potential learning from error analysis. Modern approaches push organisations to prevent the occurrence of errors and, if necessary, to start recovery operations. The methodology proposed in this paper is aimed precisely at this second type of approach. There is a growing need for systems that show errors or deviations from the desired state of the system, leading operators towards the proper execution of tasks and reducing 'conformance risk'. This paper presents a methodology and a simulation model of 'Conformance Risk Aware Design' in order to support decision-makers in modelling business processes.**

Index Terms—**BPM, BPMN, Business Process Modelling, Compliance Checking, Simulation, System Dynamics**

List of acronyms	
BPM	Business Process Management
CI	Conformance Index
SD	System Dynamics
BPMN	Business Process Modeling Notation
PMC	Process Model Conformance
PMCI	Process Model Conformance Index

I. Introduction

THE enterprise environment, increasingly complex and connected to agents outside the organisation, imposes a rapid evolution of management techniques. A new practice as therefore been established to manage organisations as dynamic entities crossed horizontally by processes. Business process management (BPM) techniques have led to many discussions and investigations related to process management, inducing the emergence of various techniques and notations for the graphical modelling of business processes. Process mapping has become very important, especially for the design phase of a new model. Modern organisations must respond to changing market needs and submit to the normative and ethical constraints imposed by government authorities. Such constraints must be translated into procedures during the mapping phase in order to later guide operator behaviour. High-quality mapping improves the processes themselves, establishing procedures that comply with the above-mentioned restrictions and regulations. The computerisation of organisational processes also provides tools for assessing mapping quality. Every process produces, in each instance, a log file that can be used to reconstruct the real procedures and operations performed by operators as well as to highlight deviations and errors from the process mapped during the planning stage. Such gaps represent an unacceptable loss for organisations. These losses manifest their effects in different ways, and organisations need tools for managing and limiting the risks resulting from deviations from the best solutions. These tools should include practical solutions and mapping tools to prevent such deviation. In this respect, then, deviation from a process can be understood as a failure of the system during the process design phase. The remainder of this paper is organised into five sections. In the second section, a literature review concerns the process mining concepts of compliance checking and non-conformance risk. The third section focuses on the methodological approach developed here for monitoring system deviations. The fourth section shows and analyses the simulation results, and, finally, in the fifth section, conclusions and future developments of this research are presented.

Manuscript received October 28, 2014; revised May 29, 2015.

Piera Centobelli is with the Department of Chemical, Materials and Industrial Production Engineering, University of Naples Federico II, Piazzale Tecchio 80, 80125, Naples, Italy (phone: +39-081-7682388; e-mail: piera.centobelli@unina.it).

Giuseppe Converso, Mosè Gallo, Teresa Murino and Liberatina Carmela Santillo are with the Department of Chemical, Materials and Industrial Production Engineering, University of Naples Federico II, Piazzale Tecchio 80, 80125, Naples, Italy (e-mail: giuseppe.converso@unina.it; mose.gallo@unina.it; murino@unina.it; santillo@unina.it).

II. Relevant literature

Process mining is a BPM technique that allows business processes to be analysed based on event logs. The basic idea is to extract knowledge from event logs recorded by an information system [1]. Process mining is quite a recent technique that merges both the modelling and the analysis of business processes by using computational intelligence [2]. Starting from log files extracted by an informatics system an organisation has already implemented, process mining mainly aims to extract knowledge from these log files, from which it is possible to collect information about past processes and to deduce, monitor, and improve processes in a wide variety of application domains [3].

Alves de Mederios et al. [4] and Van der Aalst [5] state that the purpose of business process mining is to obtain a process model formulation starting from event logs. The events recorded by computer systems are used to extract information about business activities and their causal relationships. There are three basic types of techniques for business process mining: process discovery, conformance

(Advance online publication: 10 July 2015)

checking (or conformance analysis), and extension.

• Discovery techniques take an event log as the input to produce a process model. There are many examples in the literature of the automatic construction of models using Petri Nets [6],[7].

• Conformance checking techniques take an event log and a process model as inputs [8]. Their output comprises diagnostic information that shows discrepancies between the process model and the system's behaviour as recorded in the event log. If discrepancies between the model and log are found, the model needs to be improved in order to better capture reality.

• Once the quality of a model has been established, the technique of extension is used to improve the model with new paradigms and new features to be monitored, such as cost control, time, or resources.

Conformance checking is one of the most interesting aspects of process mining. In the design phase, an organisation makes a trade-off between compliance needs and an estimation of process costs. A highly prescribed process provides a large number of control points and defined activities, which leads the process itself to be slower, increasing the direct and indirect costs of the process and, beyond that, increasing average execution time.

The proposal developed in this paper is to use a risk assessment approach to create a simulation model that performs conformance checking during the design phase, reducing non-conformance risk and related costs. The simulation logic, used in this paper, is the System Dynamics, chosen for its efficiency, intuitive application, and it's sensitivity analysis capacity, applicable in different context as shown by [9].

In our proposal, the role of compliance checking in the design phase is to foresee possible deviations from the ideal process. This new vision optimises the overall cost of process implementation, thereby reducing organisational costs and losses depending on the costs of process control versus the cost of deviations from ideal execution.

Through a systematic collection and evaluation of the relevant literature, an analysis of current approaches to business process mining and conformance risk is presented in this section, showing the contribution of a simulation model based on the methodological approach proposed in this paper. As a concept, conformance risk is relatively new, appearing in the literature since the beginning of the 21st century. A search from 1996 to 2014 (limited to journal articles in English) found 132 articles on the topic. We identified 20 articles with a focus on non-conformance risk or compliance checking analysis. Several articles analyse conformance risk in the application field of data security [10]-[13]; these cannot be compared directly with our paper topic.

The topics of process mining and conformance are closely related to software development and implementation. Da Cruz and Ruiz analyse process mining as applied to software development [14]. Their case study of a software maintenance company allows them to examine the factors involved in the operation of a knowledge discovery process. Yeung analyses the concept of choreography conformance as a fundamental requirement for the implementation of collaborative processes, developing a new formal approach to study the integration of web services and conformance verification that involves model checking to verify choreography conformance [15].

Legal requirements and market pressure for operational excellence are the most important drivers of the development of process compliance measurement [16]. The authors review different approaches for compliance measures proposed in recent years, observing how all of them are grounded in state-based techniques. Consequently, they propose different compliance measures based on causal behavioural profiles. Mendoza et al. provide a systematic and integrated approach to analyse the phases of specification, design, and verification using a model checking technique [17]. Their main aim is to describe and validate a formal compositional verification approach to verifying business processes.

Business processes affect a variety of data sources such as databases and audit trails, and in recent years, process mining techniques have been adopted in different commercial BPM systems. Tran et al. develop a model-driven and view-based approach to solving problems related to compliance concerns that uses domain-specific language to allow technical and non-technical experts to operate according to their expertise and knowledge [18]. Ellis et al. discuss model-log conformance, providing a definition of model fidelity [19]. They analyse how to manipulate the residual error factor of this model, demonstrating an important aspect of generation algorithms using normalised, probabilistic languages.

Process conformance has four dimensions, namely (1) fitness, (2) precision, (3) generalisation, and (4) structure, according to Munoz-Gama and Carmona [20]. These authors develop a metric to evaluate the dimension of precision; the proposed approach is implemented on an open-source process mining platform. In [21] Van der Aalst analyses how process models based on actual behaviour differ from process models made in advance. In fact, conformance checking techniques show important deviations between models and reality. The author lists seven problems related to process modelling, based on 100 process mining projects. Caron et al. propose a rule-based, process mining approach for the timely investigation of process event data [22]. This two-dimensional, business rule taxonomy can be used as a source for the development of rule-based compliance checking.

Service-oriented architecture facilitates the online monitoring of system behaviours during process execution. In [23] Rodríguez et al. propose an approach to managing, assessing, and improving compliance in service-oriented architectures. They develop a tool to design compliant, service-based processes that create awareness of a company's compliance state and help analyse why and where compliance violations have occurred.

Wang et al. develop and apply to an important Chinese port specialising in bulk cargo a methodology based on process mining techniques for knowledge acquisition [24]. The methodology consists of five phases: extraction, pre-processing, explorative analysis, performance analysis, and process conformance analysis. By using this method, the authors provide a support tool for process improvement in logistics.

The increasing availability of event data makes process analysis very complex. Many recent studies focus on complexity reduction through the decomposition (splitting) of process mining problems into smaller problems using the notion of passages or trace clustering to group similar traces [25],[26]. Jagadeesh Chandra Bose and Van der Aalst propose an approach for trace alignment that reduces event log complexity, exploring the logs in simpler ways [27]. Trace alignment applications allow the detection of deviations between anomalous and normative traces. Van der Aalst elaborates an approach to decomposing process mining problems [28]. This approach can be combined with different conformance checking techniques and existing process discovery. The main aim is to divide process mining problems into many smaller problems, the results of which can be combined to solve the original problem. Van der Aalst and Verbeek propose decomposing process mining problems by using the notion of passages, realising process discovery and conformance checking for the overall problem by aggregating the results of each technique for each passage [29].

Usually, discrepancies detected after the execution of process instances derive from either an incorrectly designed model (a declarative process) or irregular process execution. Borrego and Barba develop a constraint-based approach to the conformance checking of declarative processes [30]. The recorded instances were classified as compliant or non-compliant in order to determine which parts of the model were poorly designed. In sum, the review of the literature presented in this section shows no studies involving a process mining approach applied to a method of conformance checking through simulation tools preliminary to the design of production processes.

III. From the Methodological Approach to the Combined Simulation Model

The proposed research framework arises from the intention to overcome the risk of non-compliant process execution (also called the 'deviation problem'), in particular as related to service processes or, more generally, to processes involving intangible assets for which quality is not measurable. Given this aim, the paper presents a new procedural path for solving the process deviation problem during the design phase. This path consists of four steps:

1. Risk assessment related to the deviation of the i-th node (activity) from the mapped procedure, which gives back, as result, the 'critical nodes' of the process;
2. Introduction of a conformity assessment pattern in order to identify the criticality of each upstream node and to report the procedures to be followed for the online management of each node;
3. Formulation of an indicator (named the conformance index, or CI) that takes into account a reliability model with reference to possible deviations of the real system from the mapped process; and
4. Implementation of a simulation using combined tools to calculate the effects due to the direct action of and the feedback connected to the introduction of the patterns.

From a methodological point of view, the work is anchored on two major pillars: (i) the analysis of process mapping languages through process mining techniques and the subsequent evaluation of their implementation, which leads to the 'deviation problem' definition and its possible solution on the semantic level of process mapping languages and (ii) the use of a simulation model to measure the quality of process design that is implemented for each case as a solution to the deviation problem. The first pillar, which is based on a traditional view of risk analysis, manages the concept of non-conforming process execution and determines two innovative design outputs, namely a semantic solution and a measure index. By contrast, the second pillar is the innovation of a scientific approach to verify the effectiveness and efficiency of a dynamically proposed process design solution. This requires two simulations suitably combined with each other, as shown in Section IV.

The proposed simulation model has four fundamental features:

- It forecasts the costs of deviation vs. anti-deviation solutions;
- It considers human behaviour;
- It is composed by different computational tools based on dynamic analysis; and
- It returns considerations about process robustness for the production of services.

The scientific approach adopted in the model's definition is structured as follows:

1. Identify the objectives and logic of the simulation: fuzzy logic and system dynamics (SD);
2. Identify the system variables;
3. Define and specify the system variables and the relationships among them;
4. Define the model design: interaction among the variables, fuzzification, and the stock and flow diagram;
5. Define the scenarios;
6. Run the simulation; and
7. Analyse the results.

A. Risk assessment, conformity pattern, and Conformance Index definition

The main aspect of any future evaluation is to identify in a systematic and detailed way all risks related to an individual business process. This step can be conducted by using different methodologies. Thanks to the results of the risk identification phase, it is possible to quantitatively evaluate risk of activities. Typically, for this phase, it is possible to identify two factors affecting risk quantification: (1) the probability of the event linked to the specific considered risk will occur and (2) its impact on business performance.

The overall assessment of individual risk is obtained by the product of the probability of occurrence and the impact assessment; risk is then summarised in an index of criticality. In this way, after establishing an acceptable risk level, it is possible to obtain numerical values in order to determine which activities require preventive interventions. Process mapping techniques identify activities as nodes. All risk deviation management policies should start from the proper evaluation of the conformance risk of all mapped elements. Each mapped node represents a point of the process that could split from the ideal condition (e.g., due to faulty decision-making or the delayed execution of operations), which could affect both the process and the overall organisation. The methodological approach developed in this paper assigns to each node of the map a risk value (R_i) equal to the product of the deviation's probability (Pr_i), obtained from log files of similar processes and similar nodes, and a measure of the effects on the organisation (Imp_i) due to a deviation from the desired condition. The output is a ranking and clustering of nodes into two categories: critical and not critical. For this reason, several Pr_i - Imp_i charts are introduced. The functions reported in these charts represent the thresholds for an acceptable level of risk.

This methodological approach introduces a new element into the design phase of the process, called a 'conformance controller'. The conformance controller manages the risk the process deviates from the desired condition. The control loop is shown in Figure 1. This controller is ideally included in the BPM objects library and, in particular, among the control elements.

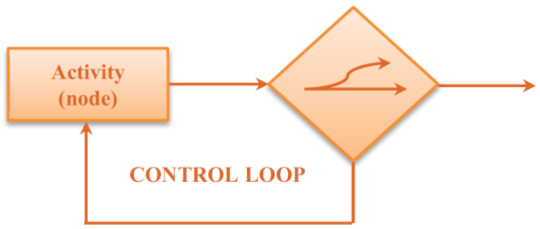

Fig. 1. Conformance controller loop.

The controller is independent of the mapped process. Properly sized, this indicator supports the design phase of the mapping process in several applications. The conformity assessment pattern uses the list of 'critical' nodes as input in order to identify the critical upstream node. The introduction of this new conformance checking concept, as well as the conformance controller element, makes the mapping process lean. A single graphical symbol has been introduced to show a complex concept that should be represented with several syntactic elements.

The quality of the mapped process can be measured through the CI. This index measures if the choices made during the design stage regarding the critical activities are well built. The index is calculated as follows:

$$CI = \frac{R_{d_{max}}}{R_{ist}} = \frac{\sum R_i}{\sum c_i * R_{i_c} + \sum D_i} \qquad (1)$$

where:

$R_{d_{max}}$ is the deviation risk when no activities have been restricted or, more precisely, the sum of the conformance risks assessed to individual activities;

R_{ist} is equal to the losses imputable to the specific instance of the process, given by the sum of the costs of implementing the control system associated with the single instance - expressed as a percentage of the R_i value of the node to which the control refers - and the value of deviations incurred in the specific instance;

R_i i-th conformance-risk assessment;

c_i control unit cost; and

R_{i_c} conformance risk assessment of the controlled nodes.

CI values higher than one indicate a positive situation, while values lower than one represent poor management of the business process map design.

CI values higher than one indicate a positive situation, while values lower than one represent poor management of the business process map design.

IV. The Combined Simulation Model

The proposed simulation tool allows users to evaluate the inclusion of 'conformance patterns' in the planning process, under certain conditions of operating costs and process times, simulating the influence of human action on the execution of the designed process. To this end, the simulation model supports each methodology of decision-making during the design stage of a service (or, more generally, of intangible assets), shifting the time axis of the data mining process forward in time in order to forecast the best trade-off between operating costs in the production of intangible assets and non-conformance risks.

This paper provides a synergy of fuzzy logic inference and a dynamic simulation by using system dynamics. Fuzzy logic is necessary for the model because non-conformance risk is related, as a phenomenon, to human behaviour. Human behaviour is modelled by using a random variable, the central value of which is calculated through fuzzy logic, which allows the assignment of a numerical value to qualitative considerations ('computing with words').

The simulation model measures the deviations in the process between the mapped and the implemented versions, one of the fundamental analyses to perform during process execution. In this sense, the mapping process is used as a tool to prevent slippages in execution, allowing the design of appropriate solutions for this phenomenon during the early analysing and mining phases. This objective is achieved through a 'binding' mapping, which provides a set of control points, mapped (using the syntax of the BPMN) as diamonds downstream of the critical activities. This consideration, obviously, requires the preliminary risk assessment associated with the activity of the process, calculated according to the technique described above.

This model must be simulated under different scenarios with different values of the process variables set in an appropriate manner such that the simulation is able to reproduce the behaviour of the system related to a particular type of mapping and, thus, to a specific arrangement of mapped control points. Their inclusion induces contrasting effects on the system: the more control points are included, the greater the mapping's robustness and the greater its graphical complexity, but consequently this latter effect reduces mapping's robustness.

The dynamics of the system evolve as the task mapped is performed according to a comparison between the 'robustness' of the mapped activities—a numerical value obtained by using fuzzy logic tools—and 'noise', which is modelled probabilistically, following the recent literature on that subject. The final output will be an evaluation and comparison of the advantages in terms of the predictability of process execution, implying the costs saved along with the costs associated with economic losses and process lead times, for the specific mapping as implemented under various design assumptions.

The simulation tool was developed by using a combination of the Matlab® Fuzzy Inference System and Powersim Studio software to model the relationship between the characteristic parameters of the simulated design system. The model developed is shown in Figure 2.

For a full understanding of the model, it is appropriate to illustrate its variables. These are not divided into input and output variables because the logic through which the model is developed, system dynamics, elaborates on the entire system performance; thus, all variables can be used, regardless of their purpose.

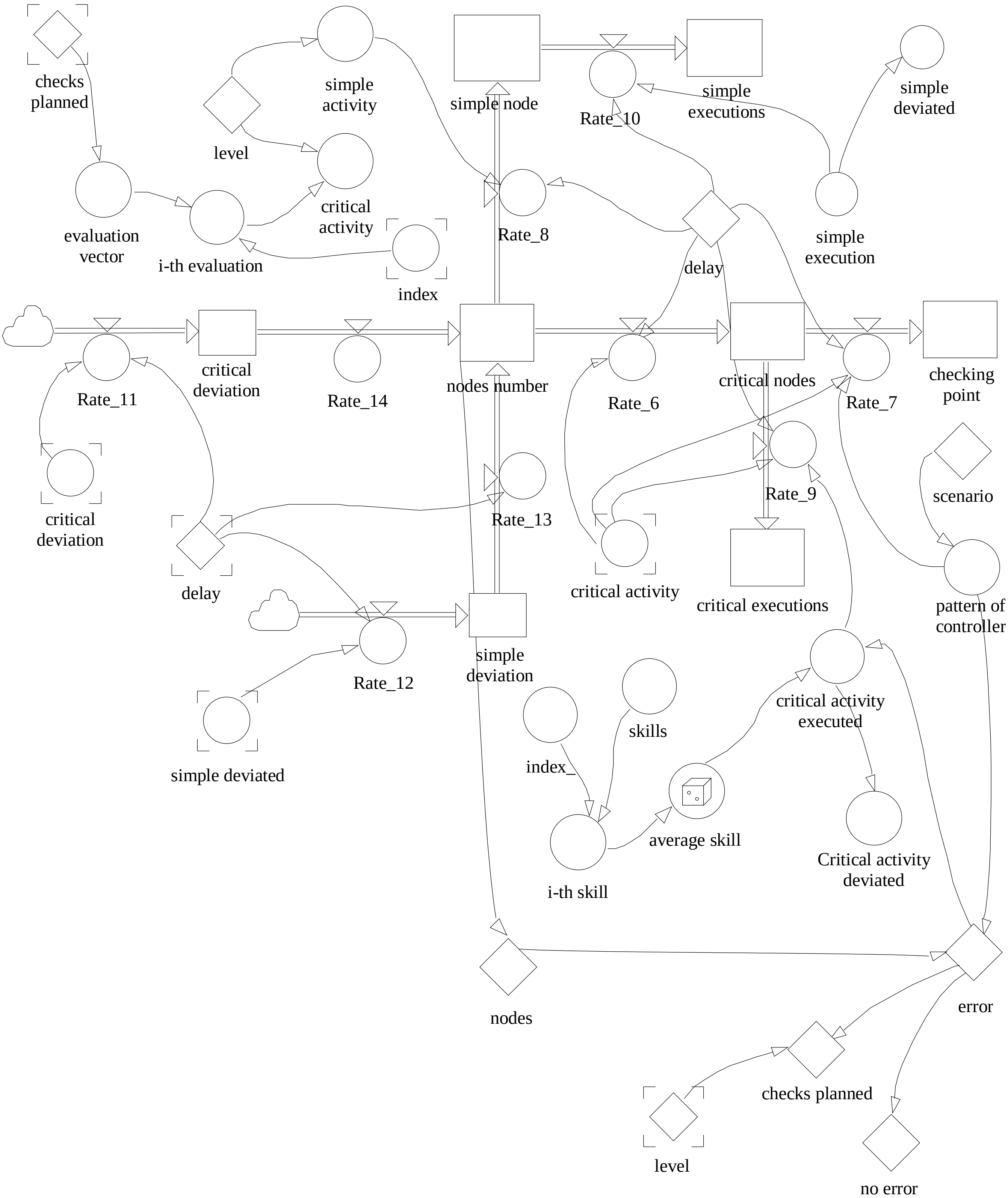

Fig. 2. Simulation model of process mapping

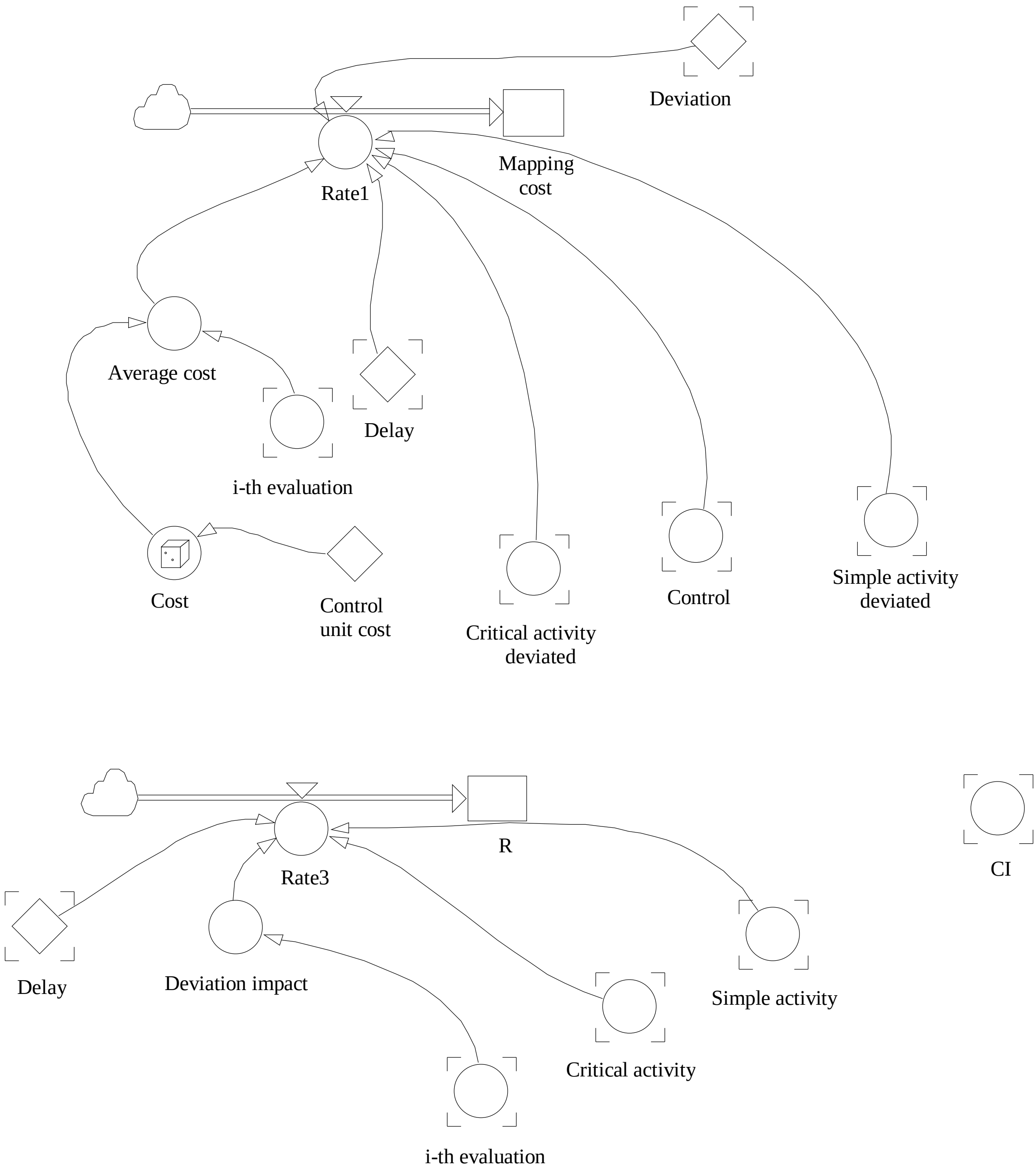

Fig. 3. Simulation model for CI identification.

For this reason, it is preferable instead to highlight the key variables. The most important variables are:

- Maximum level of acceptable conformance risk;
- Control configuration: conformance pattern vs. conformance controller;
- Cost index: average ratio between the control cost of a single node and its deviation cost;
- Lead time: ratio between the average lead time of the i-th control and the total lead time for the overall process (during simulation scenarios, this variable was increased by steps of 0.01 in order to stress the system performance to the limit conditions).

For each time step, the system analyses the occurrence of a new mapped node, assigns to it an R_i value, and identifies it as a 'critical node' if R_i is higher than the maximum level of acceptable risk conformance defined during the design stage.

Each control increases the total number of nodes to be processed, meaning that the error probability is represented by the value of a specific variable, called 'error'. This value should be compared with another represented by a variable called 'skill', which is examined in more detail in the following section. Finally, the system updates the CI value. The corresponding model for determining the CI value is outlined in Figure 3.

A. *The 'error' variable approach*

The 'error' variable was modelled by referring to recent research on the quality of business process modelling during the design phase (Mendeling et al., 2007). In particular, the evaluation of statistical probability errors will evolve (in terms, that is, of model quality) depending on the number of nodes included in a model. From a conformance checking perspective, these probabilities can be used as the average probability of deviation from the mapping. The correlation coefficient calculated on data dispersion allows us to find a linear equation that well approximates the trend y = 0.0067 + 0.1667 * x, as shown in Figure 4.

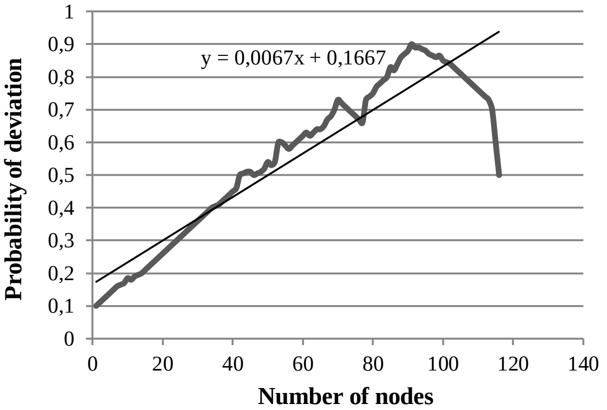

Fig. 4. Distribution of error probability [31].

B. *The 'skill' variable*

The inclusion of control points in the mapping aims to prevent deviation from requirements. The existence of control points downstream of 'critical' nodes should push the system to not deviate from what is mapped. Fuzzy logic was used to simulate, with a standard deviation assigned by the system, the average numeric value of mapping robustness, allowing a randomly extracted value to be compared at each step with the probabilistic variable 'error' (through the use of a verbal feedback regarding the mapping, typical of this logic). The level of perceived criticality and distance from the nearest control point influence the elements of each single mapped node.

After the evaluation of this numerical value through fuzzy logic, the results are exported to the simulation software. The methodology for this involves two key operations: fuzzification and defuzzification. Fuzzification first involves building a three-column matrix containing historical values referring to the probability of the successful execution of a single node at its risk values and distances from controls. Then, the shape and number of member functions for each input variable and the shape (linear or constant) of the output are set. Starting, therefore, from a numerical estimate of data from the 'train FIS', the system selects the estimation method (hybrid to a more accurate estimate), fault tolerance, and number of epochs for adapting the system to the data provided (in general, more than 10). In fuzzification, the system creates functions and fuzzy rules using a mixed technique of 'least squares' and 'backpropagation'. It is possible, finally, to check data quality by using a fixed test section, checking data generated by the model with those imported from the workspace. Defuzzification involves providing the software with a simple line of code, skill = evalfis ([lr, d], "name rescue"), where the values in brackets are those of the node.

C. *Other variable used in the simulation model*

In this section, we discuss some aspects of the simulation model not previously covered. In particular, according to the notation and nomenclature used by the SD approach, we list the main elements of the model, along with their relationships.

Critical Deviations: it is a level variable (which in system dynamics approach is an entity whose value is modified by the linked flow variables) related to risky assets diverted. The input to this variable is represented by the possible generation of, at most, a critical node at each step. In this level variable, then, a number of critical nodes will accumulate equal to the number of times for which the level of risk is higher than the constant limit. The output from the level concerns the correct execution of risk whenever the variable 'skill' exceeds or equals the variable 'error'. At the end of the simulation, this level variable accumulates only those critical nodes for which there has been a failure to perform and which therefore represent a real problem for the organization.

Critical Execution: it is a level variable, characterized by the property to be logically dual than the previous variable, 'critical deviation'.

Simple Node: This level variable constitutes a dual element compared with that of the critical nodes. Whenever the risk value is below the limit, this variable increases by 1; otherwise, this variable for the time step takes the value 0. Of course, at each time step, the sum of the number of critical and simple nodes can be a maximum of 1.

Simple Deviations: Similar to the level variable concerning risky assets diverted, this variable counts the number of times for which the generation of a node's input stream does not match a simple node's execution itself.

Simple Executions: A level variable with dual logic than 'simple deviation': the first is complementary of the second.

Control Nodes: This level variable includes a control point, a graphic controller, or the whole pattern of conformance (if we are mapping in BPMN) at some point in the designed mapping process. According to the scenario, the inclusion of control nodes occurs downstream of several nodes in different ways (conformance or conformance controller pattern).

Diverted Control: Since the controls are also subject to the robustness of the system and its characteristic 'noise', the 'Diverted Control' level variable represents the accumulation of controls that may be skipped by the system.

Costs: The variable 'Costs' accumulates the input costs associated with the controls generated in the same scenario.

Deviation Impact: This variable allows the calculation of the PMC index, which measures the impact of the deviation of the i-th activity, if this occurs. This variable is set at each step equal to the value of the corresponding element of the vector of evaluation.

R: A variable that accumulates all the costs related to a given deviation from a mapped process, including the cost of the implicit element of a 'conformance controller' downstream of each critical activity.

PMCI: The variable output by our model, that is, an indicator of mapping quality, from a conformance point of view and calculated at each step of the process.

Lead Time Target: A variable representing the value (in terms of time) that accumulates over process execution. A lead time value is assigned to each activity; in order to lead time values in percentages for each activity they have been normalized according to the LTT.

Lead Time: The expected lead time after including the control pattern.

Control Time Percentage: A constant representing the average execution time of the included control pattern.

Time Mean: An auxiliary variable that serves to randomise the time value of the control pattern.

Surprise%: Percentage indicator of the delay induced by the inclusion of the pattern. Equal to (Lead Time Target - Lead Time) / Lead Time.

Before implementing and running a simulation tool, it is necessary to define a series of controllable quantities. The configuration of these variables determines the scenario boundaries within which the results obtained from the simulation can be evaluated. These variables are described as follows:

Scenario: A simple variable modelled with a constant, discrete construct that allows moving from one scenario to another depending on whether the input value is 1 or 2. This simulation element determines the 'acceptable level of risk', the 'cost of implementation of the i-th control', and, finally, the 'percentage duration of the i-th control with respect to lead time'.

N: The number of mapped nodes, a level variable whose value is decided by the user, but the part of the control patterns increases in a manner dependent on the value of the constant 'scenario', and not the nodes performed in line with the procedure.

Level (level of tolerated risk): Just like the previous item, the level variable of tolerated risk is modelled with a constant. This level variable (decided by the modeller) remains constant for all the activities of the process.

Time Control Percentage: The inclusion of control points and monitoring within the mapped process affects one of the basic process design parameters, Lead Time, which in the field of production facilities is a competitive lever of primary importance. Precisely for this reason, it is mandatory to consider the relative increase in time resulting from the inclusion of any single control point. This parameter can be expressed as a statistical Gaussian variable, because the various control points (belonging to the same process) are similar in terms of the percentage of required time compared with total Lead Time. Obviously, as this percentage increases, the tendency to map control points will decline.

Unit Cost: This variable represents a single control cost; its value is between 0 and 1, representative of the ratio between the i-th control cost and the cost of the i-th deviation. Similar to the control time, we chose to model this element as an auxiliary variable with a statistical nature. Given all the controls of a specific process, it is reasonable to believe that they have an economic cost following a Gaussian distribution with a variance as small as possible around the average value.

D. Model validation

In order to validate the model, experiments were conducted in a real environment with a trial process comprising 25 nodes at several operators, at different times, and under general conditions. The collected data were averaged in order to derive the process deviations, the Lead Time, and the (standard) production costs of a service (or intangible asset). Subsequently, 100 runs of the same process were simulated by means of our stock and flow diagram, in any operating condition, in order to replicate the likely initial implementation conditions of a designed process, that is, before the first execution of the process mining procedure and, thus, before conformance checking.

The levels of the i-th control cost and time were increased with a step of 0.05 in order to assess the limit operating conditions. The performance of the PMCI, as the cost of the i-th control changes, was then compared with the trends for the two alternative conformance pattern mappings (pattern conformance vs. conformance controller). The comparison of the PMCI values calculated in simulated conditions and then observed in real conditions downstream of our experimental process mining procedure, shows similar results, with a difference of less than 4.5%. Although, no application has yet been made to a real case, the experimentally obtained validation is encouraging and certainly sufficient to promote, for the future, more detailed investigation and to warrant a careful analysis of the simulation results reported below.

V. Simulation Results

In this discussion, 25 'activities' are defined. This choice is neither random nor arbitrary, meeting the guidelines for the design of 'robust processes' suggested by [31] that invite designers to map processes with a number of elements fewer than 50 nodes, a number beyond which it is even recommended to use process under-modelling, because non-compliance enforcement would then be too high. Considering the included elements planned for checking, it is reasonable, therefore, to limit to 25 the number of elementary tasks included in the simulation.

The logical process that leads to any consideration of the simulation begins with the creation of a 25 × 2 matrix, where the number of rows is equal to the number of mapped nodes and the second column represents the risk level of the i-th node evaluated during risk assessment.

This paper simulates three scenarios, corresponding to three perceptions concerning the acceptable risk of deviation from the mapped procedure. The first scenario considers, as critical, all nodes for which the deviation risk exceeds 4; the system simulates process execution when providing 'conformance checker' mapping elements downstream of each node exceeding that threshold value of 4. This is the average scenario.

The second simulated scenario considers, as critical, all nodes for which the deviation risk exceeds 2; this is the severe scenario. The third simulation scenario considers, as critical, all nodes for which the deviation risk exceeds 6; this is the worst-case scenario. Again, downstream of each critical node, the model provides a conformance checker. Figure 5 shows the results obtained for each scenario.

In the first scenario, there are 17 critical nodes, or 68% of the total nodes mapped in the process. In the second scenario, there are 11 critical nodes, or 44% of the total nodes. In the third scenario, there are only three critical nodes, 17% of the total nodes. In this case, the designed process is more suitable as regards deviation risk.

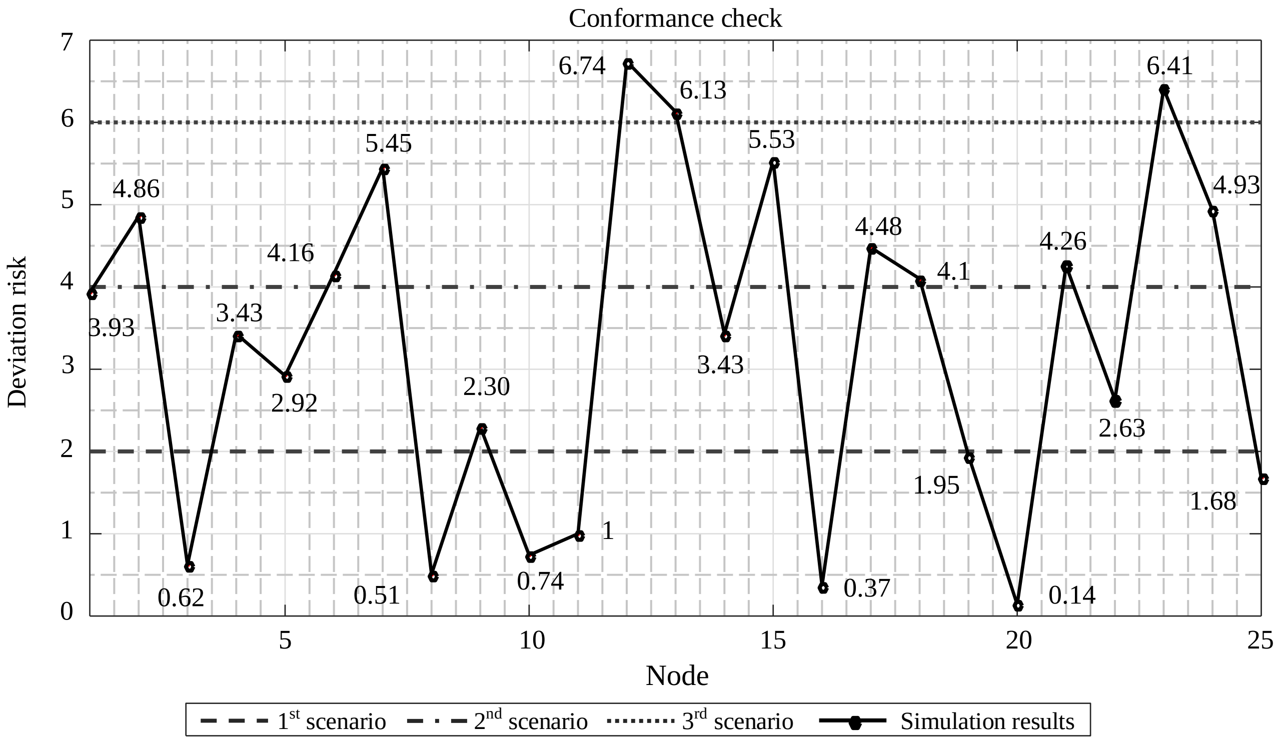

Fig. 5. Comparison of deviation risk for each scenario

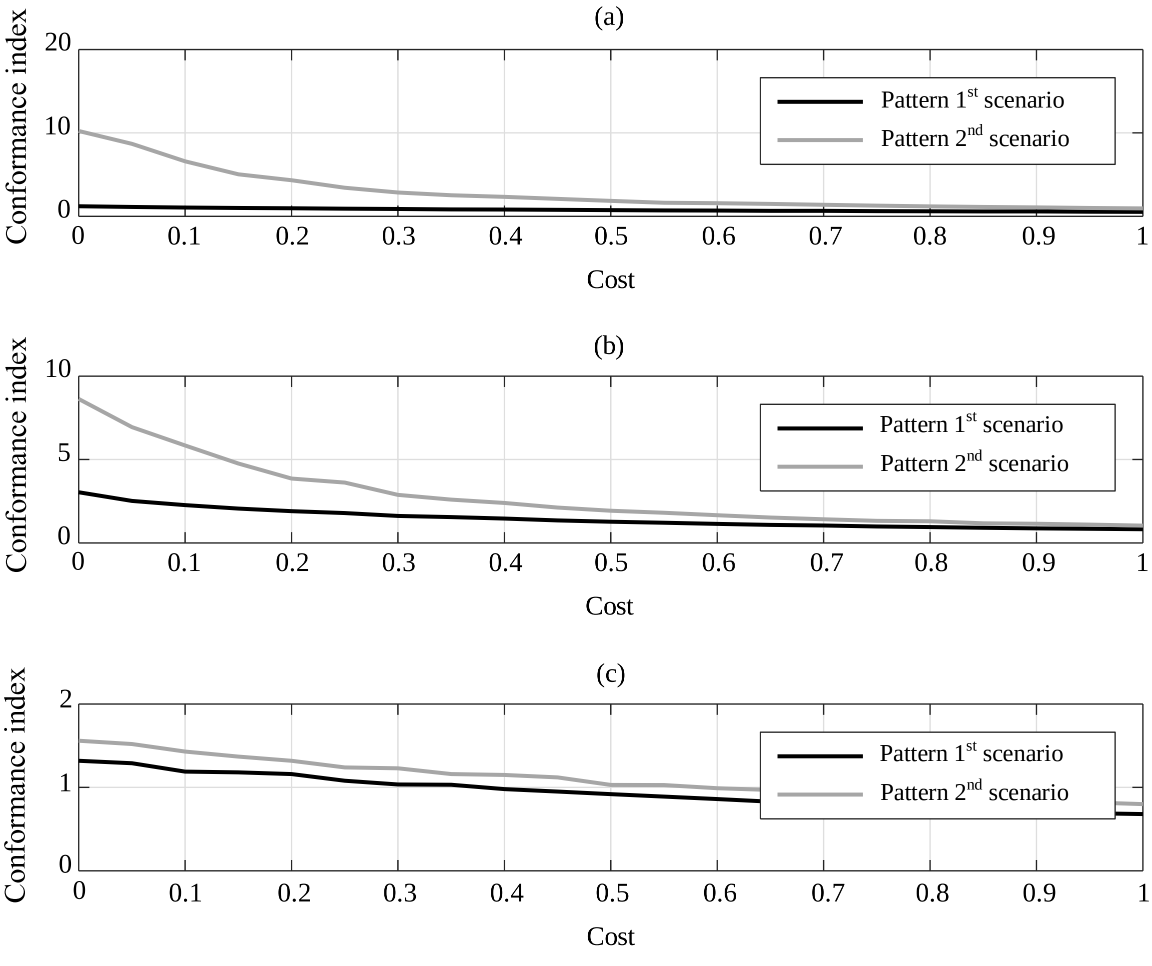

Fig. 6. (a) Graphical representation of the CI according to growing Unit Cost in the first scenario; (b) Graphical representation of the CI according to growing Unit Cost in the second scenario; (c) Graphical representation of the CI according to growing Unit Cost in the third scenario.

The three simulated cases were tested under different operating conditions; in particular, the three design choices were subjected to different levels of cost control and time required to carry out individual controls. From Figures 6a, 6b, 6c it can be deduced that the inclusion of an implicit element for controlling conformance (which, recall, is the semantic solution) in place of the pattern currently used in BPMN allows a considerable increase in PMCI for all three considered scenarios.

PMCI, moreover, is much higher with lower levels of acceptable risk and, therefore, with more patterns included. Lower levels of acceptable risk contribute to a dramatic increase in mapping complexity and, consequently, in the number of deviations. More specifically, the average percentage increase in PMCI is 72% in the case of management at level 4.23% in the case of management at level 2, and only 16% in the case of management at level 6.

In terms of design decisions supporting the proposed process and related scenarios, it is clear that designers must make their own choices about the management of conformance risk based on the average unit cost of the insertion of control patterns. For unit cost values less than 0.25, in fact, they will have to choose to control only nodes with an R_i value greater than 2, while for unit cost values higher than 0.25, a control is appropriate only downstream of nodes with an R_i value greater than 4.

Another interesting result consists of the lower efficiency of conformance risk $R_i \geq 6$ at any conceivable unit cost level, particularly for unit cost configurations higher than 0.5. This latter result is important since it substantially contrasts with the expectation that control patterns would be limited to a smaller number. Having hypothesised as unacceptable, in fact, a deviation percentage of lead time in excess of 10% (as generally used), it is easy to understand, as the boundary conditions encourage the management of risk level above 6 (more severe). Finally, Fig. 6, 7, and 8 show how the proposed solution must surely imply a lengthening of the average process lead time. The three curves relate the average percentage differences with respect to the lead time of each scenario, showing how the solution with fewer controls (risk level 6) is able to have a better lead time than the others, but with unavoidable negative consequences on the average value of PMCI.

VI. Conclusions

Business process modelling can be applied to all organisations. Process mining can be used to diagnose the actual processes. This is valuable because in many organisations most stakeholders lack a correct, objective, and accurate view on important operational processes. Process mining can subsequently be used to improve such processes. We highlighted a need for decision-makers to improve mapping quality in order to manage the risk of deviations from designed procedures. In order to satisfy this need we defined a solution consisting in the introduction of a new pattern from the BPMN modelling semantic.

The proposed pattern, in turn, has to demonstrate its efficacy in reducing risk deviation of any business process.

By combining Fuzzy Logic with System Dynamics simulation paradigm we designed a model to show the remarkable success in the simulated scenarios, particularly in the more prescriptive ones, which best represent the current reality of business. The methodology here showed remarkable success in the simulated scenarios, particularly in the more prescriptive ones, which best represent the current reality of business.

The proposed combined simulation model results can be used to identify possible inefficiencies, deviations, and risks. The relevant literature highlights the incidence of these issues with a particular focus on service business processes. Nevertheless the conformance problem shows to be a relevant issue also in manufacturing systems. Hence, we expect a wide applicability of this combined simulation approach.

This last consideration opens a new issue and a new challenge for process mapping. In the future, we are going to implement the BPM methodology, the combined simulation model, and the ‘conformance controller’ pattern in the design of actual processes.

This leads to a new issue and a new challenge for process mapping. In the future, we expect to test the system, the methodology, the application, and the ‘conformance controller’ in the design of new, real processes. This task can be quite challenging when the entity measured is so uncertain, however estimated, critically affecting decision-making.

References

[1] A. Tiwari, C.J. Turner and B. Majeed, “A review of business process mining: State of the art and future trends”, *Business Process Management Journal*, vol. 14, no. 1, pp. 5-22, 2008.

[2] W.M.P. Van der Aalst et al., “Process Mining Manifesto”, in *Lecture Notes in Business Information Processing: Business Process Management Workshops*, pp. 169-194, 2012.

[3] A.D. Ionita, M. Florea and L. Jelea, “4+1 Views for a business cooperation framework based on SOA”, *IAENG International Journal of Computer Science*, vol. 36, no. 4, pp. 332-343, 2009.

[4] A.K. Alves de Medeiros, B.F. van Dongen, W.M.P. van der Aalst and A.J.M.M. Weijters, “Process mining: Extending the a-algorithm to mine short loops”, BETA Working Paper Series, WP 113, Eindhoven University of Technology, Eindhoven, 2004.

[5] W.M.P. van der Aalst, “Trends in business process analysis”, in *Proceedings of the 9th International Conference on Enterprise Information Systems (ICEIS)* 2007, Madeira, Institute for Systems and Technologies of Information, Control and Communication, INSTICC, Portugal, p. 12-22.

[6] J. Carmona, “Projection approaches to process mining using region-based techniques”, *Data Mining and Knowledge Discovery*, vol. 24, no. 1, pp. 218-246, 2012.

[7] J.J.-P.Z. De Souza Tavares and T.A. Saraiva, “Elementary Petri net inside RFID distributed database (PNRD)”, *International Journal of Production Research*, vol. 48, no. 9, pp. 2563-2582, 2010.

[8] V.S.M. Lam, “A precise execution semantics for BPMN”, *IAENG International Journal of Computer Science*, vol. 39, no. 1, pp. 20-33, 2012.

[9] M. Keyanpour and A. Mahmoudi, “A hybrid method for solving optimal control problems”, *IAENG International Journal of Applied Mathematics*, vol. 42, no. 2, pp. 80-86, 2013.

[10] H. Jalali and A. Baraani, “Genetic-based anomaly detection in logs of process aware systems”, *World Academy of Science, Engineering and Technology*, 64, pp. 304-309, 2010.

[11] Z. Jaroucheh, X. Liu and S. Smith, “Recognize contextual situation in pervasive environments using process mining techniques”, *Journal of Ambient Intelligence and Humanized Computing*, vol. 2, no. 1, pp. 53-69, 2011.

[12] N. Ahmed and R. Matulevičius, “Securing business processes using security risk-oriented patterns”, *Computer Standards and Interfaces*, vol. 36, no. 4, pp. 723-733, 2014.

[13] G. Müller and R. Accorsi, “Why are business processes not secure?” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8260 LNCS, pp. 240-254, 2013.

[14] J.I.B. Da Cruz and D.D. Ruiz, “Conformance analysis on software development: An experience with process mining”, *International Journal of Business Process Integration and Management*, vol. 5, no. 2, pp. 109-120, 2011.
[15] W.L. Yeung, “A formal and visual modeling approach to choreography based web services composition and conformance verification”, *Expert Systems with Applications*, vol. 38, no. 10, pp. 12772-12785, 2011.
[16] M. Weidlich, A. Polyvyanyy, N. Desai, J. Mendling and M. Weske, “Process compliance analysis based on behavioural profiles”, *Information Systems*, vol. 36, no. 7, pp. 1009-1025, 2011.
[17] L.E. Mendoza, M.I. Capel and M.A. Pérez, “Conceptual framework for business processes compositional verification”, *Information and Software Technology*, vol. 54, no. 2, pp. 149-161, 2012.
[18] H. Tran, U. Zdun, T. Holmes, E. Oberortner, E. Mulo and S. Dustdar, “Compliance in service-oriented architectures: A model-driven and view-based approach”, *Information and Software Technology*, vol. 54, no. 6, pp. 531-552, 2012.
[19] C.A. Ellis, K. Kim, A. Rembert and J. Wainer, “Investigations on Stochastic Information Control Nets”, *Information Sciences*, vol. 194, pp. 120-137, 2012.
[20] J. Munoz-Gama and J. Carmona, “A general framework for precision checking”, *International Journal of Innovative Computing, Information and Control*, vol. 8, no. 7 B, pp. 5317-5339, 2012.
[21] W.M.P. van der Aalst, “What makes a good process model?: Lessons learned from process mining”, *Software and Systems Modeling*, vol. 11, no. 4, pp. 557-569, 2012.
[22] F. Caron, J. Vanthienen and B. Baesens, “Comprehensive rule-based compliance checking and risk management with process mining”, *Decision Support Systems*, vol. 54, no. 3, pp. 1357-1369, 2013.
[23] C. Rodríguez, D. Schleicher, F. Daniel, F. Casati, F. Leymann, and S. Wagner, ”SOA-enabled compliance management: Instrumenting, assessing, and analyzing service-based business processes”, *Service Oriented Computing and Applications*, vol. 7, no. 4, pp. 275-292, 2013.
[24] Y. Wang, F. Caron, J. Vanthienen, L. Huang and Y. Guo, “Acquiring logistics process intelligence: Methodology and an application for a Chinese bulk port”, *Expert Systems with Applications*, vol. 41, no. 1, pp. 195-209, 2014.
[25] J. De Weerdt, S. Vanden Broucke, J. Vanthienen and B. Baesens, “Active trace clustering for improved process discovery”, *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 12, art. no. 6507222, pp. 2708-2720, 2013.
[26] M. Song, H. Yang, S.H. Siadat and M. Pechenizkiy, “A comparative study of dimensionality reduction techniques to enhance trace clustering performances”, *Expert Systems with Applications*, vol. 40, no. 9, pp. 3722-3737, 2013.
[27] R.P. Jagadeesh Chandra Bose and W.M.P. van Der Aalst, “Process diagnostics using trace alignment: Opportunities, issues, and challenges”, *Information Systems*, vol. 37, no. 2, pp. 117-141, 2012.
[28] W.M.P. van Der Aalst, “Decomposing Petri nets for process mining: A generic approach”, Distributed and Parallel Databases, 31 (4), pp. 471-507, 2013.
[29] W.M.P. Van Der Aalst and H.M.W. Verbeek, ”Process discovery and conformance checking using passages”, *Fundamenta Informaticae*, vol. 131, no. 1, pp. 103-138, 2014.
[30] D. Borrego and I. Barba, “Conformance checking and diagnosis for declarative business process models in data-aware scenarios”, *Expert Systems with Applications*, vol. 41, no. 11, pp. 5340-5352, 2014.
[31] J. Mendling, G. Neumann and W.M.P. Van Der Aalst, “Understanding the occurrence of errors in process models based on metrics”, in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 4803 LNCS (PART 1), pp. 113-130, 2007.