Proceedings of the Federated Conference on
Computer Science and Information Systems pp. 919–922
ISBN 978-83-60810-51-4

Proposal of Square Metrics for Measuring Business Process Model Complexity

Krzysztof Kluza, Grzegorz J. Nalepa
AGH University of Science and Technology
al. A. Mickiewicza 30, 30-059 Krakow, Poland
E-mail: {kluza,gjn}@agh.edu.pl

***Abstract*—Business Process (BP) metrics are used for controlling the quality and improving models. We give an overview of the existing metrics for describing various aspects of BP models. We propose simple yet practical square metrics for describing complexity of a BP model. These metrics are easy to interpret and provide some basic information about the structural complexity of the model. The proposed metrics are to be used with models built with Business Process Model and Notation (BPMN). It is currently the most widespread language used for BP modeling.**

***Index Terms*—BPMN, Business Processes, Business Process Measurement, Quality Metrics, Complexity Metrics**

I. Introduction

BUSINESS Process (BP) models constitute a graphical representation of processes in an organization. The Business Process Model and Notation (BPMN) [1] visual language helps in collaboration between developers, architects and business analysts. Although there are many new tools in this area, they do not make BPMN models more comprehensible. Thus, some *measuring features describing model complexity* are desired.

To be a base for communication, models should be easy to understand and to maintain. Although visualizing can help in evaluation task [2], it would be useful to have *measures that can provide information about understandability and maintainability* of a BP model [3]. BP metrics can help to control, estimate and improve processes during design [4].

This paper gives an overview of selected existing metrics for describing various aspects of BP models. As an original contribution, we propose a new, simple, yet practical metric for BPMN models, which adapts concepts of some commonly known bibliometric indicators for the BP model purposes.

The rest of this paper is organized as follows: Section II presents BPMN model and its elements. Section III describes the existing approaches to measuring various aspects of BP models. In Section IV the new square metrics for BP are proposed. The paper is summarized in Section V.

II. BPMN model and elements

A Business Process [5] is a collection of related tasks that produce a specific service or product (serve a particular goal) for a particular customer. BPMN is the most widespread notation for modeling BP. It uses a set of predefined graphical elements to depict a process and how it is performed.

Although BPMN 2.0 defines three models to cover various aspects of processes, in most cases, using only the Process Model is sufficient. Four basic categories of elements used to model such processes are: flow objects (*activities*, *gateways*, and *events*), connecting objects (*sequence flows*, *message flows*, and *associations*), swimlanes, and artifacts.

III. Metrics for Business Processes State of the Art

One of the most influential papers in the business process modeling field by Mendling et al. [6] concerns guidelines for modelers, which should be taken into account when modeling business processes. They formulated seven guidelines and prioritized them with the help of industry experts [6]:

1) Model as structured as possible.
2) Decompose a model with more than 50 elements.
3) Use as few elements in the model as possible.
4) Use verb-object activity labels.
5) Minimize the routing paths per element.
6) Use one start and one end event.
7) Avoid OR routing elements.

However, the existing tools do not require to comply with such requirements, so a user has to adhere to these guidelines itself.

In the past decades, a lot of research has been conducted on the measurement of software programs [7]. Metrics have been used for many purposes, such as predicting errors [8], supporting refactoring or estimating costs of software [9] or measuring software functional size [10]. Khlif et al. noticed the similarity between object-oriented software and business process model [11]. Other authors [12], [13] argue that workflow processes are quite similar to software programs in such respects as: focusing on information processing, dynamic execution that follows a static structure and having a similar compositional structure. Table I shows some basic concepts which are similar in both the BPMN and OO approaches. Some of the software metrics have been adapted to analyze and study business processes characteristics [14], [11], [15].

According to Conte et al. [16], the quality of software design is related to five design principles: coupling, cohesion, complexity, modularity and size [15]:

- *Coupling* is measured by the number of interconnections among modules.

The paper is supported by the *BIMLOQ* Project funded from 2010–2012 resources for science as a research project.

Table I
SIMILARITIES BETWEEN BPMN AND OO CORE CONCEPTS [11], [15]

Object oriented software	**BPMN models**
Classes/Packages	Sub Processes, Processes
Methods	Tasks
Method invocations	Control flow or message flow incoming to a task
Variables/Constants	Data objects
Comment lines	Annotations

- *Cohesion* is a measure of the relationships of the elements within a module.
- *Complexity* measures the number and size of control constructs.
- *Modularity* measures if the system's components may be separated and recombined by logical partitioning.
- *Size* measures the overall dimension of software.

In [14], Cardoso et al. introduced several simple complexity metrics adapted from software engineering. Based on the simple metric, which counts the number of Lines of Code (**LOC**) of a program, they proposed three metrics:

- **NOA** – Number of activities in a process.
- **NOAC** – Number of activities and control-flow elements.
- **NOAJS** – Number of activities, joins, and splits.

To evaluate the difficulty of producing business process, Cardoso introduced Control-Flow Complexity (**CFC**) metric, which borrows some ideas from McCabe's cyclomatic complexity [17]. This metric uses the number of states induced by control-flow elements in a process. Moreover, for estimating process length, volume, and difficulty, Cardoso et al. introduced the notion of Halstead-based Process Complexity (**HPC**) measures [14]. This composite measure of complexity comprises a set of primitive measures:

- $n1$ – the number of unique activities, splits and joins, and other control-flow elements,
- $n2$ – the number of unique data variables that are manipulated by the process and its activities,
- $N1$ and $N2$ can be easily derived directly from $n1$ and $n2$ as total numbers of the control flow and data elements.

Thus, the HPC measures can be calculated as follows:

- **process length**: $N = n1 * log_2(n1) + n2 * log_2(n2)$
- **process volume**: $V = (N1 + N2) * log_2(n1 + n2)$
- **process difficulty**: $D = (n1/2) * (N2/n2)$

Although the HPC measures for processes can have several advantages (do not require in-depth analysis of process structures and can predict rate of errors) [14], it is not defined how the primitive measures should be adapted for BPMN.

Another metric adapted by Cardoso et al. is the Information Flow Metric. They proposed a metric Interface Complexity (**IC**) of an activity IC = Length*(NoI*NoO)^2. In the formula, Length of the activity can be calculated using traditional SE metrics such as LOC and the number of inputs (**NoI**) and outputs (**NoO**) of the activity follow directly from the model.

A. M. Latva-Koivisto [18] proposed the Coefficient of Network Complexity (**CNC**) metric for business processes. CNC is a widely used metric in network analysis and was proposed to measure the degree of complexity of a critical pass network. Cardoso et al. [14] precised the formula:
CNC = number of arcs / (no. of activities, joins, and splits).

Rolon et al. defined in [19] a large set of simple metrics for the evaluation of conceptual models. Their metrics are grouped into two main categories: Base and Derived Measures. The base measures are defined as a number of each kind of BPMN elements that a business process model is composed of, e.g. **NT** (Number of Tasks) – the total number of tasks in a process model. Base on such measures they proposed several derived measures which uses some simple measurement function that describes the proportions among different elements of the model, e.g. **TNA/NSFA** (Total Number of Activities/Number of Sequence Flows between Activities) – describes connectivity level between activities in a process.

Selected simple metrics defined by Sánchez-González et al. [20] are presented in Table II. Each of these metrics is easy to understand and describes one aspect of a process model.

In [3], Gruhn and Laue took into consideration other properties of processes to define some new metrics. They proposed analyzed nesting properties – Maximum Nesting Depth (**MaxND**) and Mean Nesting Depth (**MeanND**): The nesting depth of an action is the number of decisions in the control flow that are necessary to perform this action. They noticed that a greater nesting depth implies greater complexity and both nesting depth metrics have a strong influence on other structure-related complexity metrics.

They also formulated Cognitive Complexity (**CC**) metric which uses the cognitive weight for business process models. Their metric is based on the research by Shao and Wang upon the metric which measures the effort required for comprehending a piece of software. The Cognitive Functional Size (**CFS**) metric for software measurement uses predefined (based on empirical studies) cognitive weights for basic control structures. The cognitive weight of a control structure is a measure for the difficulty to understand this control structure. Hence, such a metric measures indirectly difficulty to understand a model (thus complexity).

In [12] Reijers and Vanderfeesten defined cohesion and coupling metrics for the design of activities in a workflow design, based on an information processing perspective on workflow processes. Their metrics concerned the relations between activities and information elements, as well as connections between activities in a process. Then in [13], they evaluated their workflow process designs using these metrics. Unfortunately, they do not use the BPMN notation for their solution and the presented metrics are very complex.

Lassen and van der Aalst proposed three metrics for workflow nets [21]:

- Extended Cardoso Metric (**ECaM**) – a Petri net version of metric that generalizes and improves the original CFC metric proposed by Cardoso. It focuses on the syntax of the model and ignore the complexity of the behavior.
- Extended Cyclomatic Metric (**ECyM**) – directly adapted from McCabe Cyclomatic. It focuses on the resulting behavior and ignore the complexity of the model itself.

Table II
SIMPLE METRICS DEFINED IN [20]

Metric	Description
Number of nodes	Number of activities and routing elements in a model.
Diameter	Length of the longest path from a start node to an end node.
Density	Ratio of the total number of arcs to the maximum number of arcs.
Separability	Ratio of the number of cut-vertices on the one hand to the total number of nodes in the process model on the other.
Sequentiality	Degree to which the model is constructed out of pure sequences of tasks.
Depth	Maximum nesting of structured blocks in a process model.
Gateway Heterogeneity	Number of different types of gateways used in a model.
Cyclicity	Number of nodes in a cycle to the sum of all nodes.
Concurrency	Maximum number of paths in a process model that may be concurrently activate due to AND-splits and OR-splits.

- Structuredness Metric (**SM**) – which stems from the observation that workflows are often structured in terms of design patterns.

ECaM and ECyM metrics are just simple modifications of the CFC metric. SM, in turn, constitutes a new metric, which recognizes different kinds of structures and scores them, in contrast to ECaM and ECyM that focus on a single aspect, behavior or syntax, and do not consider the interaction between the different elements. Although SM can be easily computed by a machine, it is not very intuitive for a designer. Moreover, it concerns Petri net, thus can not be directly used for BPMN.

In several survey papers [22], [23], [24], [25] some overviews of the current state of the art in various areas of business process metrics are presented. It can be concluded that this field of research is still growing and there is a lot of potential for further development of business process metrics.

In a systematic review of measurement in business processes [4], Gonzales et al. distinguished several measurable concepts for business process models, such as: complexity, understandability, quality, entropy, density, cohesion and coupling. However, most research in the area of business process measurement focus on complexity, because it is connected with understandability of the process.

IV. PROPOSAL OF SIMPLE SQUARE METRICS: DSM AND PSM

Although simple metrics, such as NOA or Diameter, can be easily calculated, they do not take into account the variety of structures used by the model. Several metrics that take into account model structure, e.g. CNC or TNA/NSFA, are either complex to calculate or not intuitive for a designer. Moreover, some of presented metrics do not concern the BPMN notation, thus they have to be adapted for this purpose (e.g. HPC).

We propose using a measure taking into account both types of process elements and their number. The idea originates in such concepts as h-index [26] or Durfee Square. Based on the distribution of the types of process elements, Durfee Square Metric (**DSM**) equals d if there are d types of elements which occur at least d times in the model (each), and the other types of elements occur no more than d times (each).

To give a more accurate representation of the distribution shape, we propose using Perfect Square Metrics (**PSM**) based on the g-index [27] as well. Thus, PSM can be defined as follows: given a set of element types ranked in decreasing order of the number of their instances, the PSM is the (unique) largest number such that the top p types occur (together) at least p^2 times.

Both DSM and PSM are intended to measure simultaneously the variety and the number of process elements. There are several advantages of the proposed metrics. They are very intuitive for a designer (a natural number that is easy to interpret) and not very complex to calculate. Moreover, they can be easily used to measure any process model, particularly for BPMN models.

Let us consider an exemplary process and calculate the presented metrics for its model. Figure 1 presents a BPMN process model from the domain of earthquake response (presented originally in [28]). Table III presents the selected metrics calculated for this exemplary model.

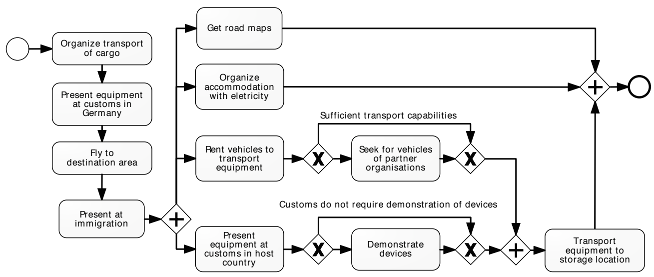

Figure 1. An exemplary process model for our metrics evaluation [28]

Table III
SELECTED METRICS CALCULATED FOR THE EXEMPLARY PROCESS MODEL

Metric	Result for the example
NOA(C/JS)	NOA=11, NOAC=NOAJS=20
CFC	5
HPC	N=V=58.6, D=7.5
CNC	1.2
TNA/NSFA	0.58
Diameter	13
DSM and PSM	DSM=3, PSM=4

To *evaluate* the proposed metrics, a number of BP cases have been evaluated. The cases represented different levels of complexity w.r.t. BPMN artifacts used. The results are being evaluated. An experimental tool that calculates these metrics has been implemented. After the preliminary discussion with the BP designers, it can be observed, that proposed measures can be helpful in describing both the complexity of the models, and their maintainability. The main use scenario is to help the designer to adapt the designed model and control its complexity by using the calculated measures during the design.

V. Conclusions and future works

This paper focuses on a problem of analyzing the business process characteristics and measuring its quality. Currently, a lot of research is carried out in the area of business processes measurement. Most of the initiatives in this area are adapted from the software engineering field and have not been empirically validated yet.

We provided an overview of the current state of research in this area, in order to identify existing research gaps and future research directions. This fits to our current research directions covering integration of BPMN models with business rules [29], [30], and their execution [31].

Complex metrics struggle with human understandability. Although they measure different aspects of processes and often provide valuable data, they are difficult to interpret. We believe that the DSM and PSM metrics can be a simple alternative to already known metrics which are easy to explain to users.

As the metrics are expressed as easy to interpret natural numbers, they are very intuitive. Using them one can obtain information about the structural complexity of the model of business processes. It is important to mention that the proposed metrics are not difficult to calculate. The research presented in this paper is a proposal for further studies. Our future work will focus on practical evaluation of the proposed metrics and their adjustment to diverse BP cases.

References

[1] OMG, "Business Process Model and Notation (BPMN): Version 2.0 specification," Object Management Group, Tech. Rep. formal/2011-01-03, January 2011.

[2] J. Baumeister and M. Freiberg, "Knowledge visualization for evaluation tasks," *Knowledge and Information Systems*, pp. 1–30, 2010.

[3] R. Laue and V. Gruhn, "Complexity metrics for business process models," in *Business Information Systems, 9th International Conference on Business Information Systems, BIS 2006, May 31 - June 2, 2006, Klagenfurt, Austria*, H. C. M. Witold Abramowicz, Ed., 2006, pp. 1–12.

[4] L. Sánchez-González, F. García, F. R. González, and M. P. Velthuis, "Measurement in business processes: a systematic review," *Business Process Management Journal*, vol. 16, no. 1, pp. 114–134, 2010.

[5] S. A. White and D. Miers, *BPMN Modeling and Reference Guide: Understanding and Using BPMN*. Lighthouse Point, Florida, USA: Future Strategies Inc., 2008.

[6] J. Mendling, H. A. Reijers, and W. M. P. van der Aalst, "Seven process modeling guidelines (7pmg)," *Information & Software Technology*, vol. 52, no. 2, pp. 127–136, Feb 2010.

[7] A. Ligęza, "Intelligent data and knowledge analysis and verification; towards a taxonomy of specific problems," in *Validation and verification of knowledge based systems: theory, tools and practice*, A. Vermesan and F. Coenen, Eds. Kluwer Academic Publishers, 1999, pp. 313–325.

[8] H. Wang, T. M. Khoshgoftaar, J. V. Hulse, and K. Gao, "Metric selection for software defect prediction," *International Journal of Software Engineering and Knowledge Engineering*, vol. 21, no. 2, pp. 237–257, 2011.

[9] R. Grady, "Successfully applying software metrics," *Computer*, vol. 27, no. 9, pp. 18–25, 1994.

[10] C. Monsalve, A. Abran, and A. April, "Measuring software functional size from business process models," *International Journal of Software Engineering and Knowledge Engineering*, vol. 21, no. 3, pp. 311–338, 2011.

[11] W. Khlif, N. Zaaboub, and H. Ben-Abdallah, "Coupling metrics for business process modeling," *International Journal of Computers*, vol. 4, no. 4, 2010.

[12] H. Reijers and I. Vanderfeesten, "Cohesion and coupling metrics for workflow process design," in *Business Process Management*, ser. Lecture Notes in Computer Science, J. Desel, B. Pernici, and M. Weske, Eds. Springer Berlin / Heidelberg, 2004, vol. 3080, pp. 290–305.

[13] I. Vanderfeesten, H. A. Reijers, and W. M. P. van der Aalst, "Evaluating workflow process designs using cohesion and coupling metrics," *Computers in Industry*, vol. 59, no. 5, pp. 420–437, May 2008.

[14] J. Cardoso, J. Mendling, G. Neumann, and H. A. Reijers, "A discourse on complexity of process models," in *Proceedings of the 2006 international conference on Business Process Management Workshops, Vienna, Austria*, ser. BPM'06, S. D. e. a. J. Eder, Ed. Berlin, Heidelberg: Springer-Verlag, 2006, pp. 117–128.

[15] I. Vanderfeesten, J. Cardoso, J. Mendling, H. Reijers, and W. van der Aalst, "Quality metrics for business process models," in *BPM and Workflow Handbook 2007*, L. Fischer, Ed. Lighthouse Point, Florida, USA: Future Strategies Inc., 2007, pp. 179–190.

[16] S. D. Conte, H. E. Dunsmore, and V. Y. Shen, *Software engineering metrics and models*. Redwood City, CA, USA: Benjamin-Cummings Publishing Co., Inc., 1986.

[17] J. Cardoso, "How to measure the control-flow complexity of web processes and workflows," in *Workflow Handbook 2005*, L. Fischer, Ed. Lighthouse Point, Florida, USA: Future Strategies Inc., 2005, pp. 199–212.

[18] A. M. Latva-Koivisto, "Finding a complexity for business process models," Helsinki University of Technology, Tech. Rep., Feb 2001.

[19] E. R. Aguilar, F. Ruiz, F. García, and M. Piattini, "Applying software metrics to evaluate business process models," *CLEI Electronic Journal*, vol. 9, no. 1, June 2006.

[20] L. Sánchez-González, F. García, J. Mendling, F. Ruiz, and M. Piattini, "Prediction of business process model quality based on structural metrics," in *Proceedings of the 29th international conference on Conceptual modeling, Vancouver, Canada*, ser. ER'10. Berlin, Heidelberg: Springer-Verlag, 2010, pp. 458–463.

[21] K. B. Lassen and W. M. P. van der Aalst, "Complexity metrics for workflow nets," *Information and Software Technology*, vol. 51, no. 3, pp. 610–625, Mar 2009.

[22] K. Thammarak, "Survey complexity metrics for reusable business process," in *Proceedings from 1st National Conference on Applied Computer Technology and Information System (ACTIS) 2010*. Bangkok Suvarnabhumi College, 2010, pp. 18–22.

[23] G. Muketha, G. A.A.A., S. M.H., and R. Atan, "A survey of business process complexity metrics," *Information Technology Journal*, vol. 9, no. 7, pp. 1336–1344, 2010.

[24] M. Becker and R. Laue, "A comparative survey of business process similarity measures," *Computers in Industry*, vol. 63, no. 2, pp. 148–167, Feb 2012.

[25] R. Dijkman, M. Dumas, B. van Dongen, R. Käärik, and J. Mendling, "Similarity of business process models: Metrics and evaluation," *Information Systems*, vol. 36, no. 2, pp. 498–516, Apr 2011.

[26] J. E. Hirsch, "An index to quantify an individual's scientific research output," *PNAS*, vol. 102, no. 46, pp. 16 569–16 572, 2005.

[27] L. Egghe, "Theory and practise of the g-index," *Scientometrics*, vol. 69, no. 1, pp. 131–152, 2006.

[28] M. Weidlich, S. Zugal, J. Pinggera, D. Fahland, B. Weber, H. Reijers, and J. Mendling, "The impact of change task type on maintainability of process models," in *In Proceedings from the 1st Workshop on Empirical Research in Process-Oriented Information Systems (ER-POIS 2010), Hammamet, Tunesia, 07-08 June 2010*, 2010, pp. 43–54.

[29] G. J. Nalepa, "Proposal of business process and rules modeling with the XTT method," in *Symbolic and numeric algorithms for scientific computing, 2007. SYNASC Ninth international symposium. September 26–29*, V. Negru and et al., Eds., IEEE Computer Society. Los Alamitos, California ; Washington ; Tokyo: IEEE, CPS Conference Publishing Service, september 2007, pp. 500–506.

[30] K. Kluza, T. Maślanka, G. J. Nalepa, and A. Ligęza, "Proposal of representing BPMN diagrams with XTT2-based business rules," in *Intelligent Distributed Computing V. Proceedings of the 5th International Symposium on IDC 2011, Delft, the Netherlands*, ser. SCI, F. M. B. et al., Ed. Springer-Verlag, 2011, vol. 382, pp. 243–248.

[31] K. Kluza, K. Kaczor, and G. J. Nalepa, "Enriching business processes with rules using the Oryx BPMN editor," in *Artificial Intelligence and Soft Computing: 11th International Conference, ICAISC 2012: Zakopane, Poland, April 29–May 3, 2012*, ser. LNAI, L. R. et al., Ed., vol. 7268. Springer, 2012, pp. 573–581.