Quality Assessment of Business Process Models based on Thresholds

Laura Sánchez-González[1], Félix García[1], Jan Mendling[2], Francisco Ruiz[1]

[1] Grupo Alarcos, Universidad de Castilla La Mancha, Paseo de la Universidad, nº4, 13071 Ciudad Real, España
{laura.sanchez | felix.garcia | francisco.ruizg}@uclm.es

[2] Humboldt-Universität zu Berlin, Unter den Linden 6, D-10099 Berlin, Germany, jan.mendling@wiwi.hu-berlin.de

Abstract. Process improvement is recognized as the main benefit of process modelling initiatives. Quality considerations are important when conducting a process modelling project. While the early stage of business process design might not be the most expensive ones, they tend to have the highest impact on the benefits and costs of the implemented business processes. In this context,quality assurance of the models has become a significant objective. In particular understandability and modifiability as particular quality characteristics are of special interest to facilitate evolution of business models given the highly dynamic environments in which business operates. These attributes can only be assessed a posteriori, so it is of central importance for quality management to identify significant predictors for them. A variety of structural metrics have recently been proposed, which are tailored to approximate these usage characteristics. The aim of this paper is to verify how understandable and modifiable BPMN models relate to these metrics by means of correlation and regression analyses. Based on the results we determine threshold values to distinguish different levels of process model quality. As such threshold values are missing in prior research, we expect to see strong implications of our approach on the design of modelling guidelines.

Keywords: Business process, measurement, correlation analysis, regression analysis, BPMN

1. Introduction

Organizations are increasingly concerned about business process improvement, since organizational excellence is recognized as a determination of business efficiency [1]. A business process is a complex entity, therefore improvement initiatives require a prior study of them at each of its lifecycle stages. The early phases of business process design might not be the most expensive ones, but they tend to have the highest impact on the benefits and costs of the implemented business processes [2]. However, process modeling on a large, company-wide scale require substantial efforts in terms of investments in tools, methodologies, training and the actual conduct of process modeling [3], resulting in several thousand models and involving a significant number of non-expert modellers. It is well known that poor quality of conceptual models can increase development efforts or results in a software system that does not satisfy user needs [4]. It is therefore vitally important to understand the factors of process model quality and to identify guidelines and mechanisms to guarantee a high level of quality from the outset. As Mylopoulos [5] suggests,''Conceptual modeling is the activity of formally describing some aspects of the physical and social world around us for the purposes of understanding and communication''. Therefore, understanding the process is a crucial task in any process analysis technique, and the process model itself should be intuitive and easy to comprehend [6].

An important step towards improved quality assurance is a precise assessment of quality. In this context, quality can be understood as "the totally of features and characteristics of a conceptual model that bear on its ability to satisfy stated or implied needs" [7]. We analyze quality from the perspective of understandability and modifiability, subcharacteristics of usability and maintainability, respectively [8]. Several initiatives about business process metrics were published [9]. Most of these metrics focus on structural aspects including size, complexity, coupling and cohesion. The significance of these metrics relies on a thorough empirical validation of their connection with quality attributes [10]. There are, to date, still rather few initiatives to investigate the connection between structural process model metrics and quality characteristics, so we detect a gap in this area which needs more empirical research.

In accordance with the previously identified issues, the purpose of this paper is to contribute to the maturity of measuring business process models. The aim of the empirical research presented herein is to validate the connections between an extensive set of metrics and the ease with which business process models can be understood (understandability) and modified (modifiability). This was achieved by adapting the measures defined in [11] to BPMN business process models [12]. The empirical data of six experiments which had been defined for previous works were used. A correlation analysis and a regression estimation were applied in order to test the connection between the metrics and both the understandability and modifiability of the models. After the selection of the more suitable metrics for understandability and modifiability, we extracted threshold values in order to evaluate the measurement results.

The remainder of the paper is as follows. In Section 2 we describe the theoretical background of our research and the set of metrics considered. Section 3 describes the series of experiments that were used, and presents the results (correlation and regression analysis and threshold values). Section 4 discusses the findings in the light of related work. Finally, Section 5 draws conclusions and presents topics for future research.

2. Theoretical Background

This section presents the background of our research. Section 2.1 discusses theories that are relevant when considering structural metrics for process models. Section 2.2 describes the set of process model metrics that we consider for this research.

2.1. Theoretical Considerations on Process Model Usability

The usability of process models can be approached from the perspective of the ISO 9126 standard on software engineering product quality [8]. This specification identifies several dimensions of usability and maintainability, of which understandability and modifiability are among the most important. The significance of these two dimensions relates to several observations.

The subject of understanding is well-suited to the role of a pillar in the quest for theories of process modelling quality. Insights from cognitive research on programming languages point to the fact that `design is redesign' [13]: a computer program is not written sequentially; a programmer typically works on different chunks of the problem in an opportunistic order. Therefore, the designer has to constantly reinterpret the current work context. There are some indications that process modelling involves this kind of re-inspection activities [14]. This fact points to understanding as an important quality factor. There are also indications that process models have to be constantly reworked and modified, and that a lack of maintenance procedures can have a detrimental effect on process modelling initiatives [15]. In other words, the process model should be constructed in such a way that it reveals its content in the best possible manner. Both understandability and modifiability can, therefore, be leveraged.

A set of different factors for process model understanding has been discussed in literature, including personal factors, modelling purpose, domain knowledge, and modelling notation [16]. Several works have identified structural parameters as significant factors in understanding [12-15]. The importance of structural aspects stems from cognitive considerations. Research into the cognitive dimensions framework defines flow charting languages as being abstraction hating [17]. This signifies that languages for process modelling do not provide a direct mechanism for grouping activities. Another characteristic is that there are so-called hidden dependencies in process models. This entails that attainable states and potential transitions have to be inferred by the reader of a process model. These points imply that even small changes to the structural level can make a process model much more difficult to understand. This raises the question of how structure can be effectively measured.

2.2. Structural Metrics for Process Models

There is a wide range of structural metrics for process models. Their advantage is that they can be objectively measured by considering the formal graph structure of a process model. These metrics are, therefore, also called internal attributes of a process model. In our discussion on usability and maintainability we are interested in how far these internal attributes can approximate understandability and modifiability. As these aspects cannot be directly measured for the process model at hand, they are referred to as external attributes. They need to be determined by empirical evaluation through, for example, the help of experiments. Figure 1 shows how this experimental data can then be used to correlate internal and external attributes. Once clear correlations have been identified, the data can be used to statistically estimate prediction models. Such a prediction model is typically derived through the use of regression analysis.

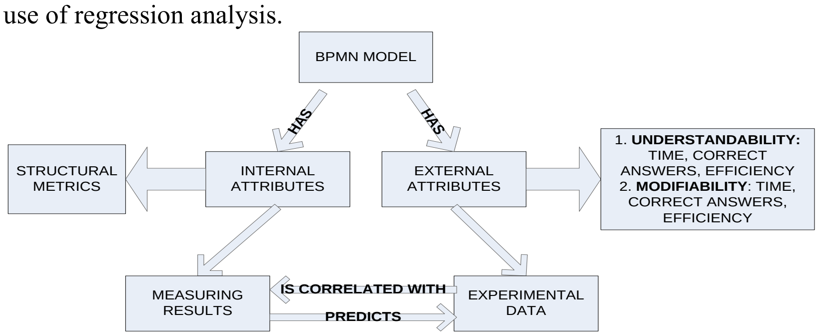

Fig. 1. Internal and external attributes of BPMN models

In this paper we consider a set of metrics defined in [9] for a series of experiments on process model understanding and modifiability. The hypothetical correlation with understandability and modifiability is annotated in brackets as (+) for positive correlation or (-) for negative correlation. The metrics include:

- Number of nodes (-): This variable is related to the number of activities and routing elements in a process model;
- Diameter (-): The length of the longest path from a start node to an end node in the process model;
- Density (-) relates to the ratio of the total number of arcs in a process model to the theoretically maximum number of arcs;
- The Coefficient of Connectivity (-) relates to the ratio of the total number of arcs in a process model to its total number of nodes;
- The Average Gateway Degree (-) expresses the average of the number of both incoming and outgoing arcs of the gateway nodes in the process model;
- The Maximum Gateway Degree (-) captures the maximum sum of incoming and outgoing arcs of these gateway nodes;
- Separability (+) is the ratio of the number of cut-vertices on the one hand, i.e. nodes that serve as bridges between otherwise disconnected components, to the total number of nodes in the process model on the other;
- Sequentiality (+) is the degree to which the model is constructed out of pure sequences of tasks.
- Depth (-) defines the maximum nesting of structured blocks in a process model;

- Gateway Mismatch (-) is the sum of gateway pairs that do not match with each other, e.g. when an AND-split is followed by an OR-join;
- Gateway Heterogeneity (-) is the extent to which different types of gateways are used in a process model;
- Cyclicity (-) relates the number of nodes in a cycle to the sum of all nodes;
- Concurrency(-) captures the maximum number of paths in a process model that may be concurrently activate due to AND-splits and OR-splits.

The series of experiments and their results are described in the following section.

3. Experimental Findings

In this section we describe the series of experiments used in this research, which were defined for previous works. Section 3.1 defines the research design. Among other aspects, we describe the subjects involved, the treatments and questions used, the variation of factors, and the response variables considered. Section 3.2 presents the results of the correlation analysis and Section 3.3 presents the results of the regression analysis. Section 3.4 discusses these results and their corresponding implications.

3.1. Research Design

This section describes the empirical analysis performed to test which structural metrics can be used as predictors of understandability and modifiability for BPMN models. Figure 2 shows the chronology of the experiments whose empirical data were used for the analysis. A total of six experiments were conducted: three (one experiment and two replicas) to evaluate understandability and three (one experiment and two replicas) to evaluate modifiability. Altogether, 127 students from four different universities took part in the experiments.

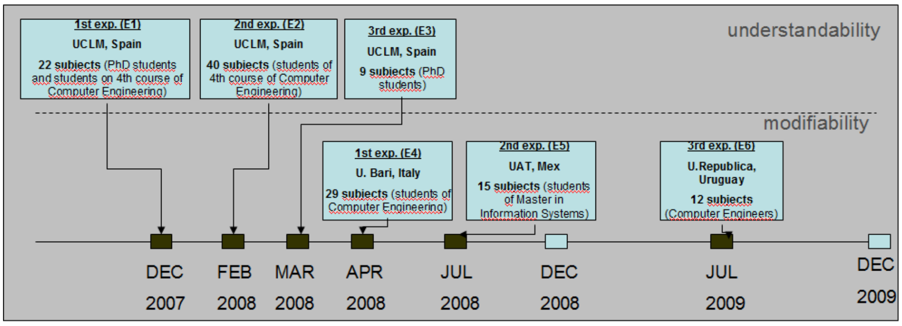

Fig. 2. Chronology of the family of experiments

The experimental material for the first three experiments consisted of 15 BPMN models with different structural complexity. Each model included a questionnaire related to its understandability. The experiments on modifiability included 12 BPMN models (selected from the 15 models concerning understandability) and each model was related to a particular modification task. A more detailed description of the design and the material used in the family of experiments can be found in [18].

It was possible to collect the following objective data for each model and each task: time of understandability or modifiability for each subject, number of correct answers in understandability or modifiability, and efficiency defined as the number of correct answers divided by time.

The first step in validating the error probability measures was to calculate their values in each of the 15 BPMN models designed for the family of experiments. The results are shown in Table 1:

Table 1. Mean and Standard Deviation of the sample models

Measures	Average	Standard deviation
Nº nodes	43.60	24.28
Diameter	12.20	5.185
Density	.038	.041
Coefficient of Connectivity	.944	.243
Average gateway degree	2.789	1.263
Maximum gateway degree	3.333	1.914
Separability	.384	.239
Sequentiality	.492	.271
Depth	1.733	1.279
Gateway mismatch	11.60	11.08
Gateway heterogeneity	-.689	.481
Cyclicity	.053	.124
Concurrency	.200	.414

Once the values had been obtained, the variability of the values was analyzed to ascertain whether the measures varied sufficiently to be considered in the study. Two measures were excluded as a result of this, namely Cyclicity and Concurrency, because the results they offered had very little variability (80% of the models had the same value for both measures, the mean value was near to 0, as was their standard deviation). The remaining measures were included in the correlation analysis.

The experimental data was accordingly used to test the following null hypotheses for the current empirical analysis, which are:

- For the experiments on understandability,
  H0,1: There is no correlation between structural metrics and understandability
- For the experiments on modifiability,
  H0,2: there is no correlation between structural metrics and modifiability

The following sub-sections show the results obtained for the correlation and regression analysis of the empirical data.

3.2. Correlation analysis

We first discuss the results for understandability and then turn to modifiability. **Understandability:** Understanding time is strongly correlated with most of the probability error measures (number of nodes, diameter, density, average gateway degree, depth, gateway mismatch, and gateway heterogeneity in all three experiments). There is no significant correlation with the connectivity coefficient, and the separability ratio was only correlated in the first experiment.

With regards to correct answers, size measures, number of nodes (-.704 with p-value of .003), diameter (-.699, .004), and gateway heterogeneity (.620, .014) have a significant and strong correlation. With regard to efficiency, we obtained evidence of the correlation of all the measures with the exception of separability.

The correlation analysis results indicate that there is a significant relationship between structural metrics and the time and efficiency of understandability. The results for correct answers are not as conclusive, since there is only a correlation of 3 of the 11 analyzed measures. In conclusion, measures with a significant correlation value (nº nodes, diameter, density, average gateway degree, maximum gateway degree, depth, gateway mismatch and gateway heterogeneity) can be traced back to particular BPMN elements, such as number of nodes (task, decision nodes, events, subprocesses, and data objects), decision nodes and sequence flow. We have therefore found evidence to reject the null hypothesis **H0,1.** The alternative hypothesis suggests that these BPMN elements affect the level of understandability of conceptual models in the following way:

- If there are more nodes, it is more difficult to understand models.
- If the path from a start node to the end is longer, it is more difficult to understand models.
- If there are more nodes connected to decision nodes, it is more difficult to understand models.
- If there is higher gateway heterogeneity, it is more difficult to understand models.

Modifiability: The correlation analysis results of the experiments concerning modifiability are described as follows. We observed a strong correlation between structural metrics and time and efficiency. For correct answers there is no significant connection in general, while there are significant results for diameter, but these are not conclusive since there is a positive relation in one case and a negative correlation in another. For efficiency we find significant correlations with average (.745, .005) and maximum gateway degree (.763, .004), depth (-.751, .005), gateway mismatch (-.812, .001) and gateway heterogeneity (.853, .000). We have therefore found some evidence to reject the null hypothesis **H0,2.** The usage of decision nodes in conceptual models apparently implies a significant reduction in efficiency in modifiability tasks. In short:

- If more nodes are connected to decision nodes, it is more difficult to modify the model.
- If there is higher gateway heterogeneity, it is more difficult to modify models.

3.3. Regression analysis

The correlation analysis presented above suggests that it is necessary to investigate the quantitative impact of structural metrics on the respective time, accuracy and efficiency dependent variables of both understandability and modifiability. This goal was achieved through the statistical estimation of a linear regression. The regression equations were obtained by performing a regression analysis with 80% of the experimental data (obtained from the family of experiments). The remaining 20% were used for the validation of the regression models.

a) Selection of models

Table 2 and Table 3 show the prediction models obtained for each experiment. All of the regression models obtained were significant with p-values below 0.05.

b) Validation of regression models

One of the threats to the validity of the findings of a study is that of not satisfying the statistical model assumptions. In the case of a linear regression model we must determine whether the observed data complies with the theoretical model. We verified the distribution of residuals, which is the difference between the predicted value with the regression equation and the actual value obtained in experiments. The residuals were analyzed for normality (Kolmogorov-Smirnov) and independence of the residuals (Durbin-Watson). The normality of the data is confirmed, since in all cases the p-value of Kolmogorov-Smirnov test is below 0.05. If the value of the second test (which typically ranges between 0 and 4) is 2, the residue is completely independent. Values between 1.5 and 2.5 are considered to be satisfactory. Values of residues for understandability and modifiability followed a normal distribution, with the exception of efficiency in E1. In the other cases, we can affirm the normality of the residuals obtained after regression analysis. For the verification of the independence of the residues we can verify compliance with the exception of the efficiency of E2 in understandability. As is true in most cases, we can state that the regression analysis is applicable to the data of the experiments.

c) Precision of models

The accuracy of the models was studied by using the Mean Magnitude Relative Error (MMRE) [19] and the prediction level Pred(25) and Pred(30) on the remaining 20% of the data, which were not used in the estimation of the regression equation. These levels indicate the percentage of model estimations that do not differ from the observed data by more than 25% and 30%. A model can therefore be considered to be accurate when it satisfies any of the following cases:

- *MMRE ≤ 0,25 or*
- *Pred (0,25) ≥ 0,75 or*
- *Pred (0,30) ≥ 0,70*

Understandability: The corresponding results are shown in Table 2 and Table 3. The best model for predicting the understandability time is obtained with the second replica E3, which has the lowest MMRE value of all the models. The best models with which to predict correct understandability answers originate from the first replication E2, and this also satisfies all the assumptions. For efficiency, no model was found that satisfied all the assumptions. The model with the lowest value of MMRE is obtained in the second replica E3. In general, the results further support the rejection of the null hypothesis **H0,1.**

Table 2. Prediction models of understandability

Undes-tand-ability	Exp	Prediction model	p-value	MMRE	p(0.25)	p(0,30)
Time	E1	T1 = 19.11 + 2 nºnodes + 3.2 gateway mismatch - 25.64 depth + 64.63 coeff. of connectivity -3.2 diameter	.000	.36	.12	.51
	E2	T2 = 95.91 + 1.51 nºnodes + 3.04 gateway mismatch- 17.35 depth - 55.98 sequentiality + 34.45 gateway heterogeneity	.000	.33	.47	.54
	E3	T3 = 47.04 + 2.46 nºnodes	.000	**.32**	**.51**	**.58**
Correct Answers	E1	CA1 = 3.125 - 0.004 nºnodes - 0.251 separability	.000	.21	.71	.71
	E2	CA2 = 3.17 - 0.005 nºnodes - 0.38 coeff. of connectivity + 0.17 depth - 0.015 gateway mismatch	.000	**.18**	**.79**	**.79**
	E3	No variable has been selected		---	---	---
Efficiency	E1	EF1 = 0.040 - 0.0004 nºnodes + 0.019 sequentiality + 0.014 density	.000	1.58	.17	.23
	E2	EF2 = -0.065 + 0.005 gateway mismatch + 0.114 sequentiality - 0.001 nºnodes	.000	4.14	.03	.03
	E3	EF3 = 0.042 - 0.0005 nºnodes + 0.026 sequentiality	.000	**0.84**	**.22**	**.25**

Table 3. Prediction models of modifiability

Modifiability	Exp	Prediction model	p-value	MMRE	p(0.25)	p(0,30)
Time	E4	E4 = 50.08 + 3.77 gateway mismatch + 422.95 density	.000	**.37**	**.31**	**.38**
	E5	E5 = 143.53 + 16.44 MaxGatewaysDegree	.010	.65	.45	.54
	E6	E6 = 175.97 + 3.88 gateway mismatch	.000	.54	.41	.50
Correct Answers	E4	CA4 = 1.85 - 3.569 density	.000	**.23**	**.82**	**.83**
	E5	CA5 = 0.62 + 0.684 sequentiality + 0.471 connectivity	.005	.28	.33	.51
	E6	No variable has been selected		---	---	---
Efficiency	E4	EF4 = 0.006 + 0.008 sequentiality	.000	**.62**	**.32**	**.42**
	E5	EF5 = 0.009 + 0.008 separability - 0.029 density	.030	.98	.45	.51
	E6	EF6 = 0.013 - 0.0002 gateway mismatch	.001	.72	.29	.37

Modifiability: We did not obtain any models which satisfy all of the assumptions for the prediction of modifiability time, but we have highlighted the prediction model obtained in E4 since it has the best values. However, the model to predict the number of correct answers may be considered to be a precise model as it satisfies all the assumptions. The best results for predicting efficiency of modifiability are also provided by E4, with the lowest value of MMRE. In general, we find some further support for rejecting the null hypothesis **H0,2.** The best indicators for modifiability are gateway mismatch, density and sequentiality ratio. Two of these metrics are related to decision nodes. Decision nodes apparently have a negative effect on time and the number of correct answers in modifiability tasks.

3.4. Discussion of Regression Results

The statistical analyses suggest rejecting the null hypotheses, since the structural metrics apparently seem to be closely connected with understandability and modifiability. There are certain metrics that may be considered to be the best owing to their significance in different experiments. For understandability these include Number of Nodes, Gateway Mismatch, Depth, Coefficient of Connectivity and Sequentiality. For modifiability Gateway Mismatch, Density and Sequentiality showed the best results. The regression analysis also provides us with some hints with regard to the interplay of different metrics. Some metrics are not therefore investigated in greater depth owing to their correlations with other metrics. For example, average gateways degree was found to correlate with depth (.810, p-value=.000) and gateway mismatch (.863, p-value=.000), signifying that information provided by these measures may be redundant. The contribution of this work is the evaluation of structural metrics by considering their relative importance in the regression analysis. We conclude that the understandability and modifiability of models is related to decision nodes and connections with others elements, which are represented in the selected measures. In the next section, we turn to threshold values. Thresholds are an important communication tool in order to state towards modellers when a process model might be considered to be of bad quality. We will focus on those metrics that are significant in the correlation and regression analysis.

3.5. Acceptable Risk Levels of Error Probability Metrics

After analyzing which measures are most useful to quantify understandability and modifiability, it is interesting to know what values of these measures indicate poor quality in models. That means, thresholds values of measures could be used as an alarm of detecting low-quality structures in conceptual models. Henderson-Sellers emphasizes the practical utility of thresholds by stating that “an alarm would occur whenever the value of a specific internal measure exceeded some predetermined value”[20]. The idea of extracting thresholds is to use them to identify unsound design structures, thus enabling engineers to gauge the threshold values to avoid obtaining hazardous structures [21]. The problem of determining appropriate threshold values is made even more difficult by many factors that may vary from experiment to experiment [22]. The identification of such threshold values, therefore, requires methods for quantitative risk assessment [23].

The statistical method used to extract threshold values is the method proposed by Bender [23]. It obtains thresholds values through a univariate logistic regression analysis. In this particular case, we use as a dependent variable the efficiency of understandability and modifiability. As a first step it is required to dichotomized the variable, signifying that it would be 1 when it was higher than the median and 0 when it was lower [24].

The method defines a "value of an acceptable risk level (VARL)". This value is given by a probability *p*0. This means that when measuring measures values below VARL, the risk of the model being non-understandable and non-modifiable is lower than *p*0. This value is calculated as follows:

$$VARL = p^{-1}(p0) = \frac{1}{beta}\left(\log\left(\frac{p0}{1-p0}\right) - alpha\right)$$

We consider these *p0* values to constitute different levels of understandability and modifiability, which is described as follows. :

- **Level 1:** there is a 10% of probability of considering the model efficient
- **Level 2:** there is a 30% of probability of considering the model efficient
- **Level 3**: there is a 50% of probability of considering the model efficient
- **Level 4:** there is a 70% of probability of considering the model efficient

For each experiment, we obtain different threshold values. They are stated in Table 4 and Table 5.

Table 4 Thresholds for error probability metrics related to understandability

level	Nº nodes			Gateway mismatch			Depth			Connectivity coefficient			Sequentiality		
	E1	E2	E3	E1	E2	E3	E1	E2	E3	E1	E2	E3	E1	E2	E3
1	63	67	65	27	30	29	4	4	4	1,7	1,7	1,6	0,1	0,1	0
2	49	50	50	16	17	16	2	2	2	1,1	1,1	1,1	0,36	0,37	0,32
3	38	37	37	7	6	6	2	1	1	0,6	0,6	0,6	0,58	0,58	0,64
4	32	29	30	2	0	0	1	1	1	0,4	0,4	0,4	0,7	0,7	0,84

Table 5 Thresholds for error probability metrics related to modifiability

level	Gateway mismatch			Density			Sequentiality		
	E4	E5	E6	E4	E5	E6	E4	E5	E6
1	31	75	32	0,2	0,5	1,1	0	0	0
2	18	31	18	0,1	0,2	0,36	0,3	0,05	0,2
3	7	0	6	0,004	0	0	0,5	0,8	0,6
4	1	0	0	0	0	0	0,6	1,2	0,8

The values described in Table 4 and Table 5 could be interpreted as follows: if number of nodes of a model is between 30 and 32, gateway mismatch is between 0 an 2, depth is 1, connectivity coefficient is 0,4 and sequentially is between 0,7 and 0,84 the probability of considering the model efficient in understandability tasks is about 70%, which means model has an acceptable level of quality. It is interesting to note that many of the threshold values are rather close to each other. This is a good indication that the thresholds can be considered to be rather stable.

Following the same steps, we extracted threshold values for the whole selected group of metrics, and organized them in different levels of understandability and modifiability. These levels classify business process models according to their quality (see Table 6). The values reported in the different rows are the median values drawn from the different experiments reported above.

The information contained in Table 6 can be interpreted as the following: if number of nodes is less or equal to 31, gateway mismatch is 1 or depth is 1, the model is considered as "very efficient" in understandability tasks, while if gateway is 1, density 0 or sequentiality is 0,86, the model is considered as "very efficient" in modifiability tasks. In the same way, if a model has more than 65 nodes, gateway mismatch is more than 29 or CFCxor is more than 30, the model is considered as very inefficient in understandability tasks and if gateway mismatch is about 46 or density is 0,6, the models is considered as very inefficient in modifiability tasks.

Table 6 Threshold values for conceptual model metrics

	1: very inefficient	2: rather inefficient	3: rather efficient	4: very efficient
		Understandability		
Nºnodes	65	50	37	31
GatewayMismatch	29	16	6	1
Depth	4	2	1	1
Coefficient of connectivity	1,7	1,1	0,6	0,4
Sequentiality	0,1	0,35	0,6	0,7
TNSF	72	49	34	20
TNE	20	12	7	2
TNG	17	10	5	0
NSFE	28	13	4	0
NMF	27	15	7	1
NSFG	40	22	11	0
CLP	7,5	4,23	2,2	0,2
NDOIN	31	44	4	0
NDOOUT	23	11	3	0
CFCxor	30	17	8	1
CFCor	9	4	1	0
CFCand	4	2	0	0
		Modifiability		
GatewayMismatch	46	22	4	1
Denstiy	0,6	0,22	0,0013	0
Sequentiality	0	0,18	0,6	0,86
NSFG	25	13	9	0
CLA	0,53	0,875	1,1	1,3
CFCxor	27	16	8	1
CFCor	9	4	1	0
CFCand	6	2,3	0	0

4. Related Metrics for Business Process Models

The interest in the measurement of business processes has grown in recent years. It is consequently possible to find a considerable amount of measurement proposals in literature. In previous works [25] we conducted a systematic review by following the Kitchenham and Charters protocol [26], as a result of which various relevant measurement proposals were selected, which could be grouped according to the lifecycle stage they have to be applied to. The most important stages are those of design and execution, and we therefore grouped the measures into "design measures" and "execution measures". Design measures are more numerous, specifically 80% of the proposals found. A summary of the proposed measures in selected publications (updated version of the systematic review until 2010) is shown in Table 4.

Table 4. Measures for Business Process Models

Source	Measurable Concept	Notation
Vanderfeesten et al [27], [28]	Coupling, cohesion, connectivity level	Petri net
Rolón et al. [29]	Understandability and modifiability	BPMN
Mendling [30]	Error probability	EPC
Cardoso [31] [32]	complexity	Graph
Jung [33]	Entropy	Petri net
Latva-koivisto [34]	complexity	Graph
Gruhn and Laue [35], [36]	complexity	UML, BPMN, EPC
Rozinat and van der Aalst [37]	compliance model-logs	Simulation Logs
Laue and Mendling [38]	Structuredness	EPC
Meimandi and Abdul Azim [39]	Activity complexity, control-flow complexity, data-flow complexity and resource complexity	BPEL
Bisgaard and van der Aalst [40]	Extended Control Flow Complexity, extended cyclomatic metric and structuredness	WF-net
Huan and Kumar [41]	Goodness of models respect generated logs in execution	Simulation logs

Some validated measures more directly related to this work are those of Cardoso [42] and Rolón [29]. Cardoso proposes a Control Flow Complexity metric (CFC). This measure takes into account the quantity and characteristics of the gateways that the business process presents, in order to provide a numerical indication of the complexity of the business process flow. This measure has been empirically validated through experiments, and a correlation analysis was carried out in [43], in which the specific measure was applied to BPMN models. On the other hand, Rolón [44] defined other measures that can be applied to BPMN models in order to quantify the understandability and modifiability of conceptual models. These measures have been validated through a correlation and regression analysis, which was published in [45]. We therefore extracted measures from this analysis, which are the most useful to measure understandability and modifiability (Table 5).

Table 5. Others validated understandability and modifiability measures

Measure	Description	U*	M*
	Measures of Rolón		
TNSF	Total Number of sequence flows	X	
TNE	Total Number of events	X	
TNG	Total Number of gateways	X	
NSFE	Number of sequence flows from events	X	
NMF	Number of message flows	X	
NSFG	Number of sequence flows from gateways	X	X
CLP	Connectivity level between participants	X	
NDOOut	Number of data objects which are outputs of activities	X	
NDOIn	number of data objects which are inputs of activities	X	
CLA	Connectivity level between activities		X
	Measures of Cardoso		
CFC	Control flow complexity. Sum over all gateways weighted by their potential combinations of states after the split	X	X

U*: Underttandability, M*: modifiability

A comparison of the correlation values of Cardoso and Rolón measures with respect to structural measures correlations presented in this work in each of the conducted experiments show that: CFC for understandability has a correlation value about 0.5, specifically CFC-efficiency (.503, .590, .515) and for modifiability it does not exceed 0.5: CFC-efficiency (-.412,-.126, -.252). Correlation values of Rolón measures are close to 0.6 for understandability, for example, between efficiency and NSFE (-.668, -.621, -.563) or CLA (-.676, -.635, -.600), and 0.4 for modifiability, TNG-efficiency (-.381, -.126, -.270) or NSFG-efficiency (-.413, -.130, -.250)). On the other hand, structural measures have correlation values for understandability around 0.8 as correlation values of efficiency and number of nodes are (-.835, -.796, -.943) or gateway mismatch are (-.761, -.768, -.737). Modifiability has also higher correlation values, for example (.814, .392, .273) for separability-efficiency or (-.573, -.655, -.751) for depth- efficiency. As a result, the validated structural measures seem to be better indicators of understandability and modifiability.

Our research on thresholds is informative to research on process modeling guidelines. Quality of conceptual process models is discussed by different frameworks such as SEQUAL or the Guidelines of Modeling [46, 47]. Many very operational guidelines on process modeling can be found in practitioner's books such as the one by Sharp and McDermott [48]. Up until now, we are only aware of the Seven Process Modeling Guidelines [49] as a guideline set that tries to define simple rules with a solid empirical foundation. This paper extends this stream of research by applying a threshold derivation approach from biometrics for the process model metrics. We deem this approach to be an important step towards translating statistical insights on correlations between metrics and quality attributes into operational design rules.

5. Conclusions and Future Work

In this paper we have investigated structural metrics and their connection with the quality of business process models, namely understandability and modifiability. We have analyzed performance measures including time, correct answers and efficiency from a family of experiments for correlations with an extensive set of structural process model metrics. Our findings demonstrate the potential of these metrics to serve as validated predictors of process model quality. This research contributes to the area of process model measurement and its still limited degree of empirical validation. Beyond that, we have adapted an approach for threshold derivation for process model quality assessment. The threshold approach can be regarded as an important step towards translating statistical insights into operational design rules.

This work has implications both for research and practice. The strength of the correlation of structural metrics with different quality aspects (up to 0.85 for gateway heterogeneity with modifiability) clearly shows the potential of these metrics to accurately capture aspects that are closely connected with actual usage. Moreover, it is possible to find threshold values for selected measures and set these values in different levels of quality related to understandability and modifiability for business process models. From a practical perspective, these structural metrics can provide valuable guidance for the design of process models, in particular for selecting semantically equivalent alternatives that differ structurally. A first attempt into this direction is made in [35].

In future research we aim to contribute to the further validation and actual applicability of process model metrics. First, there is a need for more cross validation of regression models. In particular, we will investigate in how far the regression models derived from this family of experiments provide good predictions on data that is currently collected in Berlin. Second, there is a need for more formal work on making metrics applicable in process modelling tools. Structural metrics provide very condensed information such that non-expert modellers will hardly be able to modify a model to improve the metrics. We see a huge potential for using behaviour-preserving change operations automatically for generating a model of higher quality. Techniques from graph-edit distance calculation will be a good starting point for this work.

Acknowledgments. This work was partially funded by projects INGENIO (Junta de Comunidades de Castilla-La Mancha, Consejería de Educación y Ciencia, PAC 08-0154-9262); ALTAMIRA (Junta de Comunidades de Castilla-La Mancha, Fondo Social Europeo, PII2I09-0106-2463), ESFINGE (Ministerio de Educación y Ciencia, Dirección General de Investigación/Fondos Europeos de Desarrollo Regional (FEDER), TIN2006-15175-C05-05) and PEGASO/MAGO (Ministerio de Ciencia e Innovación MICINN and Fondo Europeo de Desarrollo Regional FEDER, TIN2009-13718-C02-01).

References

1. Pfleeger, S.L., *Integrating Process and Measurement. In A. Melton (Ed.), Software Measurement.* International Thomson Computer Press, 1996: p. pg 53-74.
2. Rosemann, M., *Potential pitfalls of process modeling: part a.* Business process Management Journal, 2006. **12**(2): p. 249-254.
3. Indulska, M., P. Green, J. Recker, and M. Rosemann, *Business Process Modeling: Perceived Benefits.* Conceptual Modeling - ER 2009, 2009: p. 458-471.
4. Moody, D., *Theoretical and practical issues in evaluating the quality of conceptual models: current state and future directions.* Data and Knowledge Engineering, 2005. **55**: p. 243-276.
5. Mylopoulos, J., *Conceptual modeling and telos.* Conceptual Modeling, databases, and case: an integrated view of information systems development, 1992. **chap. 2.**: p. 49-68.
6. Dandekar, A., D.E. Perry, and L.G. Votta, *Studies in Process Simplification.* Proceedings of the Fourth International Conference on the Software Process, 1996: p. 27-35.
7. ISO/IEC, *ISO Standard 9000-2000: Quality Management Systems: Fundamentals and Vocabulary.* 2000.
8. ISO/IEC, *9126-1, Software engineering - product quality - Part 1: Quality Model.* 2001.
9. Sánchez González, L., F. García, F. Ruiz, and M. Piattini, *Measurement in Business Processes: a Systematic Review.* Business process Management Journal, 2010. **16**(1): p. 114-134.
10. Zelkowitz, M. and D. Wallace, *Esperimental models for validating technology.* IEEE Computer, Computing practices, 1998.
11. Mendling, J., *Metrics for Process Models: Empirical Foundations of Verification, Error Prediction, and Guidelines for Correctness*. 2008: Springer Publishing Company, Incorporated.
12. OMG. *Business Process Modeling Notation (BPMN), Final Adopted Specification*. 2006; Available from: http://www.omg.org/bpm.
13. Gilmore, D. and T. Green, *Comprehension and Recall of miniature programs.* International Joual of Man-Machine Studies archive, 1984. **21**(1): p. 31-48.
14. Rittgen, P., *Negotiating Models.* CAiSE 2007, 2007: p. 561-573.
15. Rosemann, M., *Potential pitfalls of process modeling: part b.* Business process Management Journal, 2006. **12**(3): p. 377-384.
16. Mendling, J., H.A. Reijers, and J. Cardoso, *What makes process models understandable?* Business process Mangement 2007: p. 48--63.
17. Green, T.R.G. and M. Petre, *Usability analysis of visual programming environments: a cognitive dimensions framework.* J. Visual Languages and Computing, 1996. **7**: p. 131-174.
18. ExperimentsURL, *http://alarcos.inf-cr.uclm.es/bpmnexperiments/*. 2009.

19. Foss, T., E. Stensrud, B. Kitchenham, and I. Myrtveit, *A Simulation Study of the Model Evaluation Criterion MMRE.* IEEE Transactions on Software Engineering, 2003. **29**: p. 985-995.
20. Henderson-Sellers, B., *Object-Oriented Metrics: Measures of Complexity.* Prentice-Hall, 1996.
21. Shatnawi, R., W. li, J. Swain, and T. Newman, *Finding Software Metrics Threshold values using ROC Curves.* Sofware Maintenance and Evolution: Research and Practice, 2009.
22. Churchill, G.A. and R.W. Doerge, *Empirical Threshold Values for Quantitative Trait Mapping.* Genetics Society of America, 1995. **138**: p. 963-971.
23. Bender, R., *Quantitative Risk Assessment in Epidemiological Studies Investigatin Threshold Effects.* Biometrical Journal, 1999. **41**(3): p. 305-319.
24. Royston, P., G.A. Douglas, and W. Sauerbrei, *Dichotomizing continuous predictors in multiple regression: a bad idea.* Statistics in Medicine, Wiley InterScience, 2005. **25**: p. 127-141.
25. Sánchez, L., F. García, F. Ruiz, and M. piattini, *Measurement in Business Processes: a Systematic Review.* Business process Management Journal, in press.
26. Kitchenham, B. and S. Charters, *Guidelines for performing systematic literature reviews in software engineering*, T. report, Editor. 2007, Keele University and University of Durham.
27. Vanderfeesten, I., H.A. Reijers, and W.M.P. van der Aalst, *Evaluating Workflow Process Designs using Cohesion and Coupling Metrics.* Computer in Industry, 2008.
28. Vanderfeesten, I., H.A. Reijers, J. Mendling, W.M.P. van der Aalst, and J. Cardoso, *On a Quest for Good Process models: the Cross Conectivity Metric.* International Conference on Advanced Information Systems Engineering, 2008.
29. Rolón, E., F. García, and F. Ruiz, *Evaluation Measures for Business Process Models.* Simposioum in Applied Computing SAC06, 2006.
30. Mendling, J., *Testing Density as a complexity Metric for EPCs*, in *Technical Report JM-2006-11-15*. 2006.
31. Cardoso, J., *How to Measure the Control-Flow Complexity of Web Processes and Workflows*, in *Workflow Handbook 2005*. 2005.
32. Cardoso, J., *Business Process Quality Metrics: Log-based Complexity of Workflow Patterns.* On the Move to Meaningful Internet Systems, 2007: p. 427--434.
33. Jung, J.Y., *Measuring entropy in business process models.* International Conference on Innovative Computing, Information and Control, 2008. **0**: p. 246-252.
34. Latva-Koivisto, A.M., *Finding a Complexity Measure for Business Process Models.* Individual Research Projects in Applied Mathematics, 2001.
35. Gruhn, V. and R. Laue, *Complexity Metrics for business Process Models.* International Conference on Business Information Systems, 2006.

36. Gruhn, V.a.L., R., *Adopting the Cognitive Complexity Measure for Business Process Models.* Cognitive Informatics, 2006. ICCI 2006. 5th IEEE International Conference on, 2006. **1**: p. 236--241.
37. Rozinat, A. and W.M.P. van der Aalst, *Conformance checking of processes based on monitoring real behavior.* Information Systems, 2008. **33**: p. 64--95.
38. Laue, R. and J. Mendling, *Structuredness and its Significance for Correctness of Process Models.* Information Systems and E-Business Management, 2009.
39. Meimandi Parizi, R. and A.A.A. Ghani, *An Ensemble of Complexity Metrics for BPEL Web Processes.* Ninth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, 2008: p. 753-758.
40. Bisgaard Lassen, K. and W. Van der Aalst, *Complexity Metrics for Workflow Nets.* Information and Software Technology, 2008: p. 610-626.
41. Huan, Z. and A. Kumar, *New quality metrics for evaluating process models.* Business Process Intelligence workshop, 2008.
42. Cardoso, J., *Process control-flow complexity metric: An empirical validation.* SCC '06: Proceedings of the IEEE International Conference on Services Computing, 2006: p. 167--173.
43. Rolón, E., J. Cardoso, F. García, F. Ruiz, and M. piattini, *Analysis and Validation of Control-Flow Complexity Measures with BPMN Process Models.* The 10th Workshop on Business Process Modeling, Development, and Support, 2009.
44. Rolón, E., F. Ruiz, F. García, and M. Piattini, *Applying Software Process Metrics in Business Process.* Procesos y Métricas, Asociación Española de Métricas del Software, 2006. **3**(2).
45. Rolon, E., L. Sanchez, F. Garcia, F. Ruiz, M. Piattini, D. Caivano, and G. Visaggio, *Prediction Models for BPMN Usability and Maintainability.* BPMN 2009 - 1st International Workshop on BPMN, 2009: p. 383-390.
46. Krogstie, J., G. Sindre, and H. Jorgensen, *Process Models representing Knowledge foraction: a revised Qualtiy Framework.* European Journal of Information Systems, 2006. **15**(1): p. 91-102.
47. Becker, J., M. Rosemann, and C. Uthmann, *Guidelines of Business Process Modeling In: W. van der Aaslt, J. Desel and A. Oberweis, Editors, Business Process Management Models, Techniques and Empirical Studies.* Springer, 2000: p. 30-49.
48. Sharp, A. and P. McDermott, *Workflow Modeling: Tools for Process Improvement and Application Development.* Artech House Publischers, 2001.
49. Mendling, J., H.A. Reijers, and W. Van der Aalst, *Seven Process Modeling Guidelines (7PMG).* Information and Software Technology, 2010. **52**(2): p. 127-136.