

Evaluation of Experiences from Applying the PREDIQT Method in an Industrial Case Study

Aida Omerovic^{1,2}, Bjørnar Solhaug¹ and Ketil Stølen^{1,2}

¹SINTEF ICT, P.O. Box 124, 0314 Oslo, Norway

²University of Oslo, Department of Informatics, P.O. Box 1080, 0316 Oslo, Norway

Email: {aida.omerovic, bjornar.solhaug, ketil.stolen}@sintef.no

Abstract—We have developed a method called PREDIQT for model-based prediction of impacts of architectural design changes on system quality. A recent case study indicated feasibility of the PREDIQT method when applied on a real-life industrial system. This paper reports on the experiences from applying the PREDIQT method in a second and more recent case study – on an industrial ICT system from another domain and with a number of different system characteristics, compared with the previous case study. The analysis is performed in a fully realistic setting. The system analyzed is a critical and complex expert system used for management and support of numerous working processes. The system is subject to frequent changes of varying type and extent. The objective of the case study has been to perform an additional and more structured evaluation of the PREDIQT method and assess its performance with respect to a set of success criteria. The evaluation argues for feasibility and usefulness of the PREDIQT-based analysis. Moreover, the study has provided useful insights into the weaknesses of the method and suggested directions for future research and improvements.

Index Terms—Quality prediction, System architectural design, Change impact analysis, Modeling, Simulation.

I. INTRODUCTION

When adapting a system to new usage patterns, processes or technologies, it is necessary to foresee the implications that the architectural design changes have on system quality. Predictability with respect to non-functional requirements is one of the necessary conditions for the trustworthiness of a system. Examination of quality outcomes through implementation of the different architecture design alternatives is often unfeasible. A model-based approach is then an alternative. We have developed a method called PREDIQT with the aim to facilitate model-based prediction of impacts of architectural design changes on system quality. Examples of quality characteristics include availability, scalability, security and reliability.

A recent case study [14] indicated feasibility of the PREDIQT method when applied on a real-life industrial system. The promising empirical results and experiences from the previous case study encouraged further and more structured evaluation of the PREDIQT method. This paper addresses experiences from applying PREDIQT on another real-life industrial system from a different domain and with different system characteristics (lifetime, purpose, technology the system is implemented on, number of users and kind of users), compared to the previous case study.

The target system analyzed serves as a semantic model and a repository for representation of the system owner's core

working processes and rules, and as a knowledge database. It is a business-critical and complex expert system used for management and support of numerous working processes, involving hundreds of professional users every day. The system is subject to frequent architectural design changes of varying type and extent. The system owner, who was also the client commissioning the analysis, required full confidentiality with respect to the kind of system targeted, the models obtained, the personnel involved and the name of the organization. This paper reports solely on the experiences obtained by the participants of the real-life case, describes the process undergone, the evaluation results, the observations and the properties of the artifacts. The reported experiences and results have provided valuable insight into the strengths and weaknesses of the method.

The case study was conducted in the year 2010. The first overall two phases of the PREDIQT method were conducted in their entirety, while the last phase was partially covered. In addition, the method is assessed through a thought experiment-based evaluation of predictions and a postmortem review. All prediction models were developed during the analysis and the entire target system (within the predefined scope) was analyzed. The analysis was performed in the form of five workshops and six intermediate meetings in a fully realistic setting in terms of the scope, the objectives, the process, the prediction models and the participants.

The rest of the paper is structured as follows. We briefly present the PREDIQT method in Section II. The research method is summarized in Section III. The process undergone during the PREDIQT-based analysis is presented in Section IV. Results of evaluation and a postmortem review are summarized in Section V. Section VI provides an evaluation of the experiences and results, with respect to five pre-defined success criteria, before concluding in Section VII.

A full technical report [16] from the case study is available and includes: details regarding the research method and success criteria; an outline of setup and data collection during the PREDIQT-based analysis; a presentation of the outcomes of the process in terms of artifacts, evaluation results and observations; the design of the evaluation template used in relation to the postmortem review, a detailed presentation of the feedback received through the evaluation template; a discussion of the threats to validity and reliability; and a summary of the related work.

II. OVERVIEW OF THE PREDIQT METHOD

The PREDIQT method produces and applies a multi-layer model structure, called prediction models, which represent system relevant quality concepts (through “Quality Models”), architectural design (through “Design Models”), and the dependencies between architectural design and quality (through “Dependency Views”). The Design Models are used to specify the target system and the changes whose effects on quality are to be predicted. The Quality Models are used to formalize the quality notions and define their interpretations. The values and the dependencies modeled through the Dependency Views (DVs) are based on the definitions provided by the Quality Models. The DVs express the interplay between the system architectural design and the quality characteristics. Once a change is specified on the Design Models, the affected parts of the DVs are identified, and the effects of the change on the quality values are automatically propagated at the appropriate parts of the DV. This section briefly outlines the PREDIQT method in terms of the process and the artifacts. For further details on PREDIQT, see [14, 15, 17].

The process of the PREDIQT method consists of three overall phases. Each phase is decomposed into sub-phases, as illustrated by Figure 1. Based on the initial input, the stakeholders involved deduce a high level characterization of the target system, its scope and the objectives of the prediction analysis, by formulating the system boundaries, system context (including the usage profile), system lifetime and the extent (nature and rate) of design changes expected. Quality Models are created in the form of a tree, by defining quality notions with respect to the target system. The Quality Models represent a taxonomy with interpretations and formal definitions of system quality notions. The total quality of the system is decomposed into characteristics, sub-characteristics and quality indicators. The Design Models represent the architectural design of the system.

For each quality characteristic defined in the Quality Model, a quality characteristic specific DV is deduced from the Design Models and the Quality Models of the system under analysis. This is done by modeling the dependencies of the architectural design with respect to the quality characteristic that the DV is dedicated to, in the form of multiple weighted and directed trees. A DV comprises two notions of parameters:

- 1) EI: Estimated degree of Impact between two nodes, and
- 2) QCF: degree of Quality Characteristic Fulfillment.

Each arc pointing from the node being influenced is annotated by a quantitative value of EI, and each node is annotated by a quantitative value of QCF.

Figure 2 shows an excerpt of an example DV with fictitious values. In the case of the *Encryption* node of Figure 2, the QCF value expresses the goodness of encryption with respect to the quality characteristic in question, e.g., security. A quality characteristic is defined by the underlying system specific Quality Models, which may for example be based on the ISO 9126 product quality standard [1]. A QCF value on a DV expresses to what degree the node (representing system

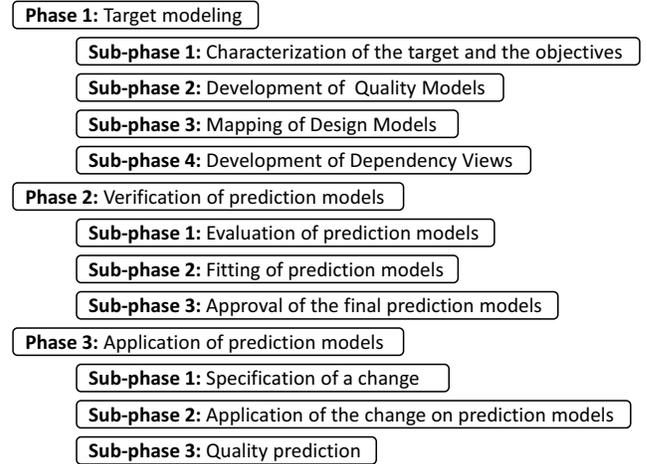


Fig. 1. A simplified overview of the process of the PREDIQT method

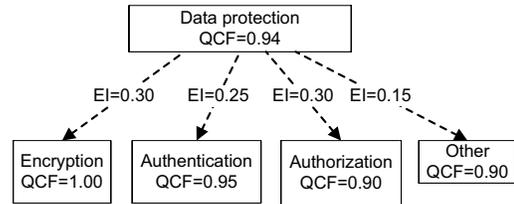


Fig. 2. Excerpt of an example DV with fictitious values

part, concern or similar) is realized so that it, within its own domain, fulfills the quality characteristic. The QCF value is based on the formal definition of the quality characteristic (for the system under analysis), provided by the Quality Models. The EI value on an arc expresses the degree of impact of a child node (which the arc is directed to) on the parent node, or to what degree the parent node depends on the child node, with respect to the quality characteristic under consideration.

“Initial” or “prior” estimation of a DV involves providing QCF values to all leaf nodes, and EI values to all arcs. Input to the DV parameters may come in different forms (e.g., from domain expert judgments, experience factories, measurements, monitoring, logs, etc.), during the different phases of the PREDIQT method. The DV parameters are assigned by providing the estimates on the arcs and the leaf nodes, and propagating them according to the general DV propagation algorithm. Consider for example the *Data protection* node on Figure 2 (denoting: DP: Data protection, E: Encryption, AT: Authentication, AAT: Authorization, and O:Other):

$$QCF_{(DP)} = QCF_{(E)} \cdot EI_{(DP \rightarrow E)} + QCF_{(AT)} \cdot EI_{(DP \rightarrow AT)} + QCF_{(AAT)} \cdot EI_{(DP \rightarrow AAT)} + QCF_{(O)} \cdot EI_{(DP \rightarrow O)} \quad \text{Eq. 1}$$

The DV-based approach constrains the QCF of each node to range between 0 and 1, representing minimal and maximal characteristic fulfillment (within the domain of what is represented by the node), respectively. This constraint is ensured through the formal definition of the quality characteristic rating (provided in the Quality Models). The sum of EIs, each between 0 (no impact) and 1 (maximum impact), assigned

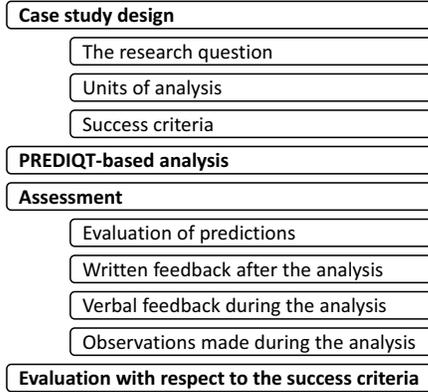


Fig. 3. Main stages of the research method

to the arcs pointing to the immediate children must be 1 (for model completeness purpose). Moreover, all nodes having a common parent have to be orthogonal (independent). The dependent nodes are placed at different levels when structuring the tree, thus ensuring that the needed relations are shown at the same time as the tree structure is preserved.

The general DV propagation algorithm, exemplified by Eq. 1, is legitimate since each quality characteristic DV is complete, the EIs are normalized and the nodes having a common parent are orthogonal due to the structure. A DV is complete if each node which is decomposed, has children nodes which are independent and which together fully represent the relevant impacts on the parent node, with respect to the quality characteristic that the DV is dedicated to.

The rationale for the orthogonality is that the resulting DV structure is tree-formed and easy for the domain experts to relate to. This significantly simplifies the parametrization and limits the number of estimates required, since the number of interactions between the nodes is minimized. Although the orthogonality requirement puts additional demands on the DV structuring, it has shown to represent a significant advantage during the estimation.

The “Verification of prediction models” is an iterative phase that aims to validate the prediction models (with respect to the structure and the individual parameters), before they are applied. A measurement plan with the necessary statistical power is developed, describing what should be evaluated, when and how. Both system-as-is and change effects should be covered by the measurement plan. Model fitting is conducted in order to adjust the DV structure and the parameters, to the evaluation results. The objective of the “Approval of the final prediction models” sub-phase is to evaluate the prediction models as a whole and validate that they are complete, correct and mutually consistent after the fitting. If the deviation between the model and the new measurements is above the acceptable threshold after the fitting, the target modeling is re-initiated.

During the “Application of prediction models” phase, a specified change is applied to the Design Models and the DVs, and its effects on the quality characteristics at the various abstraction levels are simulated on the respective DVs.

The “Application of the change on prediction models” phase involves applying the specified architectural design change on the prediction models. When an architectural design change is applied on the Design Models, it is according to the definitions in the Quality Model, reflected to the relevant parts of the DV. Thereafter, the DV provides propagation paths and quantitative predictions of the new quality characteristic values, by propagating the change throughout the rest of each one of the modified DVs, based on the general DV propagation algorithm. We have earlier developed tool support [14] based on MS Excel for simulation and sensitivity analysis related to the DVs.

III. RESEARCH METHOD

The research method is motivated by the guidelines for case study research provided by Yin [18]. A deductive approach is undertaken, where the already defined PREDIQT method is exposed to an empirical trial in the form of a case study.

The main stages of the research method are depicted by Figure 3. The case study design included characterization of research question, the units of analysis and the success criteria, as the main outcomes. The success criteria (specified and used in Section VI) are deduced based on the needs of the three main stakeholder groups: the customers, the domain experts and the analyst. The details of the case study design and deduction of the success criteria are presented in [16].

The PREDIQT-based analysis was performed by following the pre-defined process of the PREDIQT method. However, instead of performing predictions of effects of future changes during the last workshop (as specified by the PREDIQT process), we chose to demonstrate how prediction models can be applied by simulating the effects of reversal of a very large already implemented change. As such, the model application phase is not fully covered, but only demonstrated. The affected Design Model and DV elements were identified and their modified parameter values estimated by the domain experts. Thereafter, the simulation on the DVs was made by the analyst.

Additionally, in order to evaluate the predictions obtained, a thought experiment regarding the effect of the change on the root nodes of the respective DVs, was performed by the domain experts. Thus, this was a part of the method assessment. The overall assessment measures included: written feedback from the analysis participants (affiliated with the customer organization) provided (through an evaluation template) upon completion of the analysis and the above mentioned thought experiment-based evaluation, verbal feedback during the analysis from the analysis participants, and observations made by the analyst during the analysis. Based on the results of the PREDIQT-based analysis and the assessment, an evaluation with respect to the success criteria was provided.

IV. OVERVIEW OF THE PROCESS UNDERGONE DURING THE PREDIQT-BASED ANALYSIS

This section focuses on the process of the PREDIQT-based analysis (see Figure 3). We chronologically outline the relevant

events and meetings in terms of their contents, participation, preparation and the time spent.

Two preliminary meetings were held between the customer representatives and the analyst. The preliminary meetings were spent for motivating the analysis and identifying the challenges which the analysis should address. Thereafter, the analysis was organized in the form of five workshops and six working sessions in between some of the workshops. The workshops gathered both the domain experts and the customer (managers), and aimed to report on the current results and reach a milestone which the management should be involved in. The intermediate working sessions gathered the domain experts and the analyst to work tightly together on a particular task as a prerequisite for the forthcoming workshop. Table I outlines the process of the analysis. The first column specifies the type of the meeting (*PM*: preliminary meeting, *W*: workshop and *S*: working session) followed by the sequence number with respect to the kind of meeting. Column two specifies the date of the meeting. The third column lists the participants (note that all managers and domain experts are affiliated with the customer organization, while the analyst and the secretary belong to an external research group). The fourth column describes the contents and achievements of the meeting. The fifth column specifies the preparation activities for the meeting in question. The last column shows the approximate time spent (in terms of man-hours) during the meeting and in preparing for it. *T* denotes the total number of hours spent by all participants of the meeting (including the analyst), while *A* denotes the number of hours spent by the analyst only. The time spent on reporting and dissemination of the results after completion of meeting W5, is not included in the last column.

The case study was conducted in a realistic setting, with the objective of fully testing the feasibility of the method and providing added value for the customer. The target system is developed in-house, is rather complex and is used by numerous surrounding systems. The system represents an important asset for the customer organization. The changes on the system are implemented collectively approximately two times a year, while the individual changes are considered and designed frequently. The extent and number of changes are increasing. There is a requirement on the time to market of certain types of changes. The system and the associated semantic model are complex and it is therefore very hard to test all effects of changes (i.e. the cost of testing becomes an increasing problem). Alternative or complementing methods for testing are therefore desirable. For instance, prediction of change impacts can potentially be used to tune testing.

V. ASSESSMENT

This section reports on the assessment part of the research method, depicted by Figure 3. Evaluation of the predictions based on a thought experiment is presented first. Secondly, the written feedback (postmortem review) provided by the analysis participants from the customer organization upon completion of the above mentioned evaluation, is summarized. The third subsection reports on the verbal feedback provided, during

the study, by the analysis participants from the customer organization. Lastly, the experiences and observations made by the analyst during the case study, are summarized.

A. Evaluation of predictions

During the last part of the W5 meeting (that is, upon completion of the PREDIQT-based analysis), a thought experiment was performed by asking the domain experts to estimate the new root node QCF values on the respective DVs, due to a specified change (given the current and the new QCF values of the leaf nodes affected, as well as the current QCF value of the root node). The change specified was a major, already implemented architectural design change, which added a new functionality to the system. The evaluation assumed reversal of the change. The change affected up to three leaf nodes on each DV. The purpose of the thought experiment was to test usefulness of the predictions obtained from the models. That is, we assume that the domain experts have thorough knowledge of the system, and that their root node estimates reflect the reality of how the quality characteristics are affected by the change. Then, the simulated root node value is compared to the thought experiment provided one.

The thought experiment showed the following relationship between the simulated root node QCF values and their corresponding estimates (provided by the domain experts), regarding the respective above presented simulations on:

- the first one of the two DVs dedicated to *Maintainability*: no deviation between estimated (by the domain experts) and simulated (by PREDIQT)
- the second one of the two DVs dedicated to *Maintainability*: estimated is 4.5% higher than simulated
- the first one of the two DVs dedicated to *Usability with respect to the contents*: estimated is 3% higher than simulated
- the second one of the two DVs dedicated to *Usability with respect to the contents*: estimated is 7.7% higher than simulated

B. Written feedback after the analysis

The summary provided here is based on contents analysis of the answers of five respondents. The answers have been provided on a pre-defined evaluation template [16]. The answers have been abstracted and categorized in order to reduce the volume of raw text and reveal possible similarities and contrasts. Table II summarizes the background of the respondents.

The main strengths pointed out are: “The PREDIQT method is useful and it suits well the problem addressed”(R2), “It was a way to in a systematic manner divide the problem in smaller parts, and then aggregate the quality level for the whole model”(R3), and “Modeling concept – propagation of assessments”(R4). A weakness repeatedly pointed out is the missing formal mapping of the parameter estimates to the model, i.e. the parameter estimates may be too sensitive to the context and the interpretation (R1, R3, R4, R5).

TABLE I
OUTLINE OF THE PROCESS OF THE PREDIQT-BASED ANALYSIS

Meeting	Date	Participants	Contents	Preparation	Hours
PM1	March 25 2010	Two managers. The analyst.	Customer's presentation of the needs and challenges regarding quality, particularly security and interoperability of the systems. A brief presentation of the PREDIQT method and its possible application in the case study. Planning of the forthcoming meeting with the domain experts and the overall customer representatives.	Clarified formalities regarding communication channels and information exchange.	T:5 A:3
PM2	May 11 2010	Four managers. Three domain experts. The analyst.	Characterization (by the customer organization representatives) of the system architecture and main challenges that the case study may focus on. A presentation of the PREDIQT method and its possible application to the context.	The analyst received the input requested: system and enterprise architecture documentation, requirements specification, system design documentation, service level agreement and operational environment specification.	T:10 A:3
W1	June 15 2010	Three managers. Three domain experts. The analyst. The secretary.	The customer organization representatives characterized the target and the scope of the analysis: defined the target, defined the operational profile (current variability and expected changes in usage pattern, number of users, number of requests and amount of data), defined the expected lifetime of the system, specified type and extent of the expected changes, and characterized the main quality characteristics of the system.	The documentation studied by the analyst and clarifications or additional information needs communicated with the customer.	T:15 A:8
S1	June 17 2010	Two domain experts. The analyst.	Given to the analyst by the domain experts: a demo of the target system, a presentation of the functional properties of the system, specification of typical faults and failures due to changes of the system, and an overview of the testing procedures. Clarifications of the written input.	The analyst specified questions and additional information needs to the domain experts.	T:10 A:5
W2	Aug. 17 2010	Two domain experts. Three managers. The analyst. The secretary.	The analyst presented initial Quality Models (compliant with ISO 9126 [1]) and Design Models. Model revision in the group.	The analyst requested and received further documentation regarding system design. Development of system Quality Models and Design Models, by the analyst.	T:30 A:20
S2	Sept. 6 2010	Three domain experts. The analyst.	The analyst presented the updated Quality Models and Design Models. Selected use scenarios and change cases were undergone in the group, in order to check if the current models support their specification. Revision of all quality and Design Models in the group.	Updates (based on the discussion from W2 meeting) of system Quality Models and Design Models, by the analyst.	T:15 A:7
W3	Sept. 9 2010	Two domain experts. Three managers. The analyst. The secretary.	The analyst presented the current version of all prediction models. Revision of the Quality Models. Revision of the Design Models. Characterization of the types of potential architectural design changes. Preliminary approval of the available prediction models (Quality Models and Design Models).	Updates (based on the discussion from S2 meeting) of system Quality Models and Design Models, by the analyst.	T:20 A:10
S3	Sept. 28 2010	Four domain experts. The analyst.	The analyst presented the approach regarding the DV structure development (assumptions, rules, DV syntax and DV semantics) and an early draft of a DV, for the domain experts. Development of the DV structures in the group.	Development of an initial draft of a DV structure (by the analyst), for triggering the discussion and exemplification.	T:20 A:10
S4	Sept. 29 2010	Four domain experts. The analyst.	The analyst presented the approach regarding the (DV) parameter estimation (how to deduce the values, how to use the Quality Models, syntax and semantics of QCFs and EIs [14]), for the domain experts. Further development of the DV structures and DV parameter estimation in the group.	Documentation of the DV structure in the tool (MS Excel sheet customized for DVs in PREDIQT analysis). The analyst received documentation on typical system changes.	T:20 A:10
S5	Oct. 11 2010	Four domain experts. The analyst.	Further DV parameter estimation.	Documentation of the updated DVs in the tool.	T:15 A:5
W4	Oct. 20 2010	Three domain experts. One manager. The analyst. The secretary.	Validation of the DVs based on a thought experiment addressing randomly selected parts of the DVs. Model fitting of the DVs.	The analyst prepared a thought experiment setup based on the changes that the system has undergone.	T:20 A:8
S6	Oct. 22 2010	Two domain experts. The analyst. The secretary.	Continued validation of the DVs based on a thought experiment of addressing randomly selected parts of the DVs. Model fitting of the DVs. Final approval of the prediction models. Specification of changes which are to be simulated in the demo of meeting W5.		T:15 A:2
W5	Nov. 3 2010	Three domain experts. One manager. The analyst. The secretary.	A summary of the results provided by the analyst: overview of the process undergone, and a presentation of the final prediction models. A demo of application of the prediction models: change specification, application of the change on the prediction models and quality prediction in terms of propagation paths and the modified QCF values.	The analyst prepared a simulation demo.	T:20 A:8

TABLE II
BACKGROUND OF THE RESPONDENTS

	Respondent R1	Respondent R2	Respondent R3	Respondent R4	Respondent R5
Position	Senior Researcher	Chief Specialist	Software Architect	Senior Principal Engineer	Work Process Developer
Education (degree)	MSc	MSc	MSc equivalent	MSc	MSc
Years of professional experience	15	20	27	33	20
Role in the case study	Coordinator	Manager	Expert	Expert	Expert

All five respondents agreed that the models facilitate communication, knowledge exchange and understanding of the target system, its architecture and its quality characteristics. R1 argues that “the workshops force people to communicate and harmonize into one model; the system is clarified and parts of the architecture are disclosed and discussed; the most important part is assigning estimates on quality characteristics, which forces people to make statements”. R2 argues that “the method provides a good model of the system, which can be communicated around; when a multi-disciplinary group manages to make a model of a complex problem and communicate around it, you have achieved a good result; when you additionally can make predictions based on the model, the result is even better.”

R1 points out that the effort needed for conducting the analysis is reasonable from a typical management consulting perspective, but in an engineering context, more effort should be directed towards specific parts.

Regarding the future use of the method, R1 expresses the intention to use the models developed in the future, for purpose of architecture development and dependability analysis. R2 and R3 express the wish to use the method in future projects, given that financing can be provided. R4 intends to use the prediction models if they can be tailored to specific use cases, while R5 writes: “I believe the model can be used to understand and predict the result/risk in different changes”.

R1 expresses that the PREDIQT method “has already served the purpose in creating understanding and analysis. If incorporated with more tool support, I think it can be utilized in practice”. R2 expresses that PREDIQT is very much better than no method, but it is unknown what it takes for it to be perfect. R3 and R4 express that the benefit from the method and quality of the predictions depend on the modeling skills and granularity of the models. R5 points out the challenge of interpreting the predictions due to the lack of documentation of the assumptions made during the parameter estimation.

Regarding challenges with usage of the method, R2 expresses two main issues: “access to competent resources to make the models and interpretation of the predictions and the corresponding uncertainty which requires competence”. R3 points out three challenges: “be sure that you have modeled the most important aspects; models need to be verified; define the values in a consistent way”. R4 sees the uncertainty challenge in the fact that the changes are marginal and therefore give small effects on the numbers, while R5 relates uncertainty to the insufficiently formal interpretation of the parameter values due to the assumptions made during their estimation.

Regarding the main benefit of the method, R2 expresses

that PREDIQT “reduces uncertainty at change, but does not eliminate it; but it does systematize the uncertainty and reduce it sufficiently so that the method absolutely is valuable”. R3 sees the discussion of the quality characteristics and agreement upon the most important ones, as the main benefit.

The improvements suggested include simpler tool support, stricter workshops, increased traceability between the models, reuse of the Design Models based on other notations, and in-advance preparation of the experts.

C. Verbal feedback during the analysis

The verbal feedback includes the responses and comments from the analysis team, given during the different meetings – mainly by the end of the analysis. These are summarized as follows:

The quality values (or their relative distance) should be mapped to monetary values or a similar measure of cost/gain in order to facilitate a cost-benefit analysis and ease interpretation of the DV parameters.

The granularity of the changes is given by the granularity of the models. That is, minor changes may have very negligible impact on the models, unless the models are fine grained. A remedy is to deliberately increase the detail level of certain parts of the models. Still, although the parameters in such cases are almost unchanged, the prediction models help understand the propagation paths.

The process of developing and verifying the models facilitates discussions, system understanding and knowledge exchange among the participants.

The analyst should be aware of the possible bias or interests of the participants, particularly when the parameters are based on domain expert judgments.

Certain parameters require a holistic approach (e.g. business perspective) or a special background (e.g. end-user). Some parameters may be uncertain due to lack of representation of such competence in the domain expert panel.

Better documentation of the semantics and contextual information regarding the DV nodes, is needed. This would ease the use of DVs and particularly parameter estimation when some time has passed after the DV structure is developed.

Active participation of the domain experts in the model development contributes not only to the model quality, but also to the experts’ understanding of the models, and ability to use and maintain the models after the analysis.

The time spent on development of the prediction models is much longer than the time spent on the model verification. This has shown to be beneficiary, since model development was founded on numerous documents which the domain experts

could interpret and relate to the quality notions. Doing this early in the process and consistently on all parts of the models while discussing the models in the group, is preferred to verifying certain parts of the models. Ideally, one should do both, but when the resources are limited, the choice we made was preferred (due to higher model quality early in the process, as well as more extensive brainstorming and discussions in the group) provided that the verification is satisfactory.

The estimates are much more informative when considered and interpreted relative to each other, than individually. When one estimate is unambiguous in terms of the interpretation of the value and the assumptions made during its estimation, values of the others (on the same DV) may be compared to the well known one, in order to be interpreted.

D. Observations made during the analysis

Some of the main experiences and observations made by the analyst are presented in the sequel.

One of the main challenges for the analyst during the development of the Design Models was acquiring an understanding of the expert terminology used in the system. The documentation received and the S1 meeting rectified this.

Regardless of how well the analyst understands the target system and its quality characteristics, it is crucial that the analyst does not develop the prediction models alone. The model development and verification trigger many useful discussions among the domain experts, and help reveal inconsistencies and misunderstandings. In addition, the prediction models are intended to be used and maintained by the domain experts, who need to be able to relate to the models and the tools they are developed in. The optimal approach is that the analyst presents an initial version of the models, which are discussed, corrected and further developed in the group. Errors or missing parts in the initial models are often an advantage, as they trigger the discussions in the group.

It is important to dedicate sufficient resources to characterization of the target, provision of the input and formation of a common understanding of the Quality Models. These are prerequisites for avoiding elementary discussions and ambiguities during the rest of the analysis.

The analyst has to be aware of the inconsistencies of the terminology used in documents and the verbal communication among the domain experts, as well as between the overall stakeholders. Any such inconsistencies should be clarified, preferably through the Quality Models or the Design Models.

The PREDIQT method has to be sufficiently understood by all parties, and it is important to use a notation for the prediction models that all analysis participants can relate to.

The time taken to estimate the parameters of the DVs is at least twice as long as the time needed to develop the structure of the DVs. It is necessary to explain that the DV structure is developed with respect to both Design Models and Quality Models, since dependencies are modeled with respect to the respective quality characteristic that the DV is dedicated to. Availability and common understanding of the Quality Models during parameter estimation is crucial.

The structure of the DVs may need to be adjusted during the DV parameter estimation. For this, tool support more flexible than what our MS Excel sheets currently offer, is needed.

When developing the DVs, certain assumptions and choices are made. Traces to the specific Design Model elements may exist, and only certain indicators from the Quality Models may be used in estimation. The current tool support is insufficient for efficiently documenting these aspects “on the run” during the meetings.

Since a PREDIQT-based analysis requires considerable effort from the customer organization, it is essential to ensure commitment of the management and allocate the resources needed.

It is important to make the right balance between the representativeness of the domain expert panel and the effectiveness of the analysis, when choosing the size of the analysis group. Although a larger group is likely to increase statistical significance of the data, time consumption on the discussions may rapidly grow with the number of the participants. Therefore, one should ensure that a fraction of the domain expert panel is present at all meetings and provides continuity, while some turnover of the overall participants depending on the goal of the meeting may be beneficiary. The turnover however necessitates updates of the participants on both the PREDIQT method and on the current status/results of the analysis. There is clearly a trade-off between the resource consumption and the model quality.

The meetings should be as tightly scheduled as possible, provided that the necessary preparations are feasible. The rationale is to prevent the need to updates on recent results.

Approximately half a year has been a reasonable time allocation for this case study. In a commercial analysis, a tighter course during a shorter period of time could be achieved, if the participants can prioritize the analysis even more among their overall tasks and if the tool support is improved.

VI. EVALUATION WITH RESPECT TO THE SUCCESS CRITERIA

In this section we evaluate the performance of the PREDIQT method in this case study, with respect to the success criteria (SC) deduced and motivated in [16]. Thus, this section addresses the last stage of the research method depicted by Figure 3.

SC1: *The PREDIQT-based analysis facilitates predictions providing sufficient understanding of the impacts of architectural design changes on system quality characteristics, so that informed decisions can be made.*

The ability of simulating a realistic change during meeting W5 and the assessment reported in Section V, indicate that we have been able to develop an understandable and harmonized model of the system, communicate around the model, identify the dependencies and simulate the impacts of changes.

By performing thought experiments on the root node, the change propagation and its impact from the leaves throughout the different parts of the DVs, was evaluated. Whether the deviation reported is sufficiently small, is up to the customer

to assess. The answers obtained in Section V suggest that this is the case.

The thought experiment-based evaluation of the predictions resulted in no deviation on the first DV, and some degree of overestimation during the thought experiments. This can be due to varying quality of the specific models or optimism of the domain experts. We observe however that the deviation between the simulated (based on the DV models) and the estimated (through the thought experiments) root node values during both model validation and the evaluation of the predictions, has no repeatable pattern but considerably high variance. Therefore, we do not have reason to assume bias in the relationship between the simulation and the thought experiments.

Many different parts of the DVs were affected during the evaluation, which ensured both variation and complexity in the change propagation – that is, coverage of the evaluation. The number of parameters (QCFs and EIs) in each one of the four different DVs was around 60-70.

The evaluation of the degree to which the simulated and the thought experiment-based estimates coincide would have been more reliable if uncertainty [17] had been expressed in the estimates. Then, one could have based the evaluation on whether the deviation is within the already present uncertainty of the estimates. Due to the limited time and the extent of the prediction models, we did not have the resources for also including the uncertainty handling in the analysis.

SC2: *The PREDIQT-based analysis is cost-effective.*

The analysis indicates that the PREDIQT method is feasible in a fully realistic setting and within the limited resources allocated. The process of the PREDIQT method was undergone, addressed the whole target of analysis and resulted in prediction models that, as the assessment indicates, provide the customer organization with useful basis for understanding the impacts of changes, capturing the propagation paths and obtaining the predictions.

The feedback from R1 and R2 (customer management representatives) presented in Section V, indicates cost-effectiveness of the analysis. The analysis has required approximately 215 (see Table I) man-hours (apart from the reporting), which is within the resources allocated. There are, however, some issues that must be taken into consideration when evaluating these numbers. Firstly, this was the second time the PREDIQT-based analysis was performed on a real industrial case. Hence, even though the analysis team included one of the inventors of the PREDIQT method, the process is not fully streamlined yet, due to limited empirical experience with PREDIQT. It can reasonably be assumed that the process will be more effective as the analysts gain experience with applying the PREDIQT method.

Furthermore, the process of the PREDIQT method assumes that the Design Models are in place prior to the analysis. Since this was not the case, considerable time had to be spent on modeling the system. Based on the experience gained and given that the Design Models are available as input to the analysis, we believe that it should be possible to carry out this

kind of analysis within a time frame of approx. 60 man-hours spent by analyst (not including writing a final report) and ca. 50 man-hours spent by the overall participants. Hence, the success criterion appears to be fulfilled in this case. There is however still a need for a reference/baseline for comparing our results with the results from possible alternative methods. The future studies should address this, as well as cost-effectiveness per DV/quality characteristic/Design Model. Reusability of results (e.g. through experience factories) also contributes to the cost-effectiveness and should be examined in the future work.

SC3: *The prediction models are sufficiently expressive to adopt the relevant architectural design changes and analyze their effects on quality.*

The diversity of changes in the demo and the validation, the ability of simulating a realistic change during meeting W5 and the assessment, indicate that we have been able to develop a harmonized model of the system and use it for identifying the dependencies and simulating the impacts of all proposed changes. The participants provided a lot of information about the target during the analysis process. There were no instances where we were not able to capture the relevant information in the prediction models. Further application of the prediction models is however needed in order to evaluate their expressiveness and whether they can be maintained and used during the needed time period.

SC4: *The prediction models are sufficiently comprehensible to allow the domain experts to be actively involved in all phases of the PREDIQT process and achieve the goals of each phase with a common understanding of the results.*

The number of diagrams and parameter estimates was considerable. Still, the multi-disciplinary domain expert panel affiliated with several departments of the customer organization managed to discuss and agree upon the the different parts of the eventually harmonized and approved prediction models. The fact that the domain experts actively participated and continuously made progress according to the schedule of the analysis, managed to perform thought experiments and apply the models, indicates comprehensibility of the models. One of the most demanding parts of the analysis – development of the DVs, was entirely performed by the domain experts and only facilitated by the analyst.

The available prediction models were presented by the analyst during the meetings, in order to validate the correctness of the models or use them as basis for the forthcoming stages. There were many occasions where the participants suggested modifications, explained their rationale, or asked relevant questions about some detail in a model. This indicates that the models were in general comprehensible for the participants, and the postmortem review suggests that the models served well as an aid in establishing a common understanding of the target.

Still, comprehensibility of the models may vary among the participants and between the models depending on the knowledge of the system and the modeling notation. The fact that all the participants in this analysis had a strong

technical background may have contributed to making the models easier for them to understand than would be the case for an even more diverse group. It is still necessary to have an analyst explain the method and the models, as well as facilitate and manage the process, since the current tool support is insufficient for ensuring a structured process and since an adequate PREDIQT manual currently does not exist. The analyst has played a rather active part during the analysis. A disadvantage is that the active role may have influenced the analysis. However, the involvement of the analyst is openly reported and reflected upon. It has also allowed better insight into the process and a more detailed evaluation of the results.

SC5: The PREDIQT-based analysis facilitates knowledge management and contributes to a common understanding of the target system and its quality.

The answers reported in Section V consistently suggest that the PREDIQT-based analysis facilitates knowledge management. The models have served as a means of documenting the system, triggering discussions and exchanging knowledge. The means of triggering the discussions and further increasing participation of the domain experts can still be developed as a part of the method. It is for example essential that the analyst does not too actively develop any models or uses the tools alone, which would make it more demanding for the domain experts to use and maintain the models.

More structured process, improved traceability between the models, documentation of assumptions and rationale, as well as improved tool support (in terms of flexibility of modifications, usability, process guidance, as well as documentation of traces, rationale and assumptions) would facilitate the knowledge exchange and certainty of the models.

VII. CONCLUSIONS

The PREDIQT method makes use of models that capture the system design, the system quality notions and the interplay between system architecture and quality characteristics. The predictions result in propagation paths and the modified values of the parameters which express the quality characteristic fulfillment at different abstraction levels. PREDIQT aims at establishing the right balance between the practical usability of the models, and the usefulness of the predictions. We are not aware of other approaches that combine notions of architectural design and quality in this way. However, the issues of metrics estimation, system quality and the various notations for modeling system architecture, have received much attention in the literature [1-13].

The paper has presented experiences from using the PREDIQT method in an industrial case study. The contributions of the paper include a detailed account of how the PREDIQT method [14] scales in an industrial context, as well as an evaluation of the performance of the method in an industrial context.

The experiences and results obtained indicate that the PREDIQT method can be carried out with limited resources, on a real-life system and result in useful prediction models.

Furthermore, the observations indicate that the method, particularly its process, facilitates understanding of the system architecture and its quality characteristics, and contributes to structured knowledge management through system modeling. All stakeholders, including the customer, the domain experts and the analyst gained a better and a more harmonized understanding of the target system and its quality characteristics, during the process. The knowledge management in the context of this case study has concerned acquisition, exchange and documentation of the knowledge available, regarding the architectural design of the system, non-functional (quality) characteristics of the system and the interplay between the architectural design and the system quality. Four evaluation methods have been used that complement each other and that to a varying degree are used during the discussion of the success criteria.

The issue of method scalability concerns two aspects which our results indicate have been achieved and balanced: resources required to perform the analysis and the usefulness of the prediction models. In particular, the evaluation substantiates that:

- the PREDIQT-based analysis facilitates predictions providing sufficient understanding of the impacts of architectural design changes on system quality characteristics, so that informed decisions can be made,
- the PREDIQT-based analysis is cost-effective,
- the prediction models are sufficiently expressive to adopt the relevant architectural design changes and analyze their effects on quality,
- the prediction models are sufficiently comprehensible to allow the domain experts to be actively involved in all phases of the PREDIQT process and achieve the goals of each phase with a common understanding of the results, and
- the PREDIQT-based analysis facilitates knowledge management and contributes to a common understanding of the target system and its quality

within the scope of the characterized target and objectives.

Full documentation of the case study exists, but its availability is restricted due to confidentiality required by the customer. Hard evidence in the form of measurements to validate the correctness of the predictions would have been desirable, but this was unfortunately impossible within the frame of this case study. Instead, we have relied on extensive documentation and the domain expert group with solid background and diversity. Still, thought experiment-based validation of models and evaluation of the predictions have weaknesses compared to the measurement-based ones. Particularly, we can not exclude that possible undocumented or inconsistent assumptions have been made in model development, although the Quality Models and the active participation of the domain experts in all model development should prevent this. Statistical power was limited, due to low number of participants. The careful selection of experienced participants and the variety of the changes specified during model validation, compensated for some of

this. Another weakness is that the same domain expert group has developed and validated the prediction models. However, given the complexity of the prediction models, the variation of the changes applied and variance of the deviation pattern obtained (between the simulations and the thought experiment-based estimates), we can not see any indication of bias due to the same expert group.

Although the above mentioned threats to validity and reliability are present in such a study, we argue that the results indicate the feasibility and usefulness of the method in a real-life setting. The study has also provided useful insight into the strengths and weaknesses of the method, as well as suggested directions for future research. Particularly, the needs for improved traceability, even more structured process guidelines and better tool support have been highlighted.

Note that PREDIQT has only architectural design as the independent variable – the Quality Model itself is, once developed, assumed to remain unchanged. This is of course a simplification, since system quality prediction is subject to more factors than architectural design. Usage profile, quality definitions and process are examples of the factors whose variation PREDIQT does not address. Although this case study has evaluated PREDIQT in a different domain compared to the one reported in [14], many more evaluations are needed for evaluating the external validity of the method.

The target system is representative for the systems intended to be within the scope of the PREDIQT method. This is the second trial of PREDIQT in a real-life setting and both trials have given strong indications of feasibility of the method, reported similar benefits (understanding of system architecture and its quality, usefulness of estimates particularly when interpreted relative to each other, and usefulness of the process) and undergone the same stages of the PREDIQT process. There is no significant difference in the size or complexity of the prediction models between the two case studies. No particular customizations of the method were needed for this trial. Thus, we have reason to believe that it should be possible to reapply PREDIQT in another context.

Acknowledgments: This work has been conducted as a part of the DIGIT (180052/S10) project funded by the Research Council of Norway, as well as a part of the SecureChange project and the NESSoS network of excellence both funded by the European Commission within the 7th Framework Programme.

REFERENCES

- [1] International Organisation for Standardisation: ISO/IEC 9126 - Software engineering – Product quality. 2004.
- [2] V. Basili, G. Caldiera, and H. Rombach. The Goal Question Metric Approach. *Encyclopedia of Software Engineering*, 1994.
- [3] V. R. Basili. Software Modeling and Measurement: the Goal/Question/Metric Paradigm. Technical Report TR-92-96, University of Maryland, 1992.
- [4] C. Byrnes and I. Kyrazoglou. Applying Architecture Tradeoff Assessment Method (ATAM) as Part of Formal Software Architecture Review. Technical Report 07-0094, The MITRE Corporation, 2007.
- [5] L. Dobrica and E. Niemel. A Survey on Software Architecture Analysis Methods. *IEEE Transactions on Software Engineering*, 28(7):638–653, 2002.
- [6] C. Ebert, R. Dumke, M. Bundschuh, A. Schmietendorf, and R. Dumke. *Best Practices in Software Measurement*. Springer Verlag, 2004.
- [7] N. Fenton and M. Neil. A Critique of Software Defect Prediction Models. *IEEE Transactions on Software Engineering*, 25:675–689, 1999.
- [8] N. E. Fenton and S. L. Pfleeger. *Software Metrics: A Rigorous and Practical Approach*. PWS Publishing Co., 1998.
- [9] D. Heckerman, A. Mamdani, and W. M. P. Real-World Applications of Bayesian Networks. *ACM Communications*, 38(3):24–26, 1995.
- [10] R. Kazman, M. Barbacci, M. Klein, S. J. Carriere, and S. G. Woods. Experience with Performing Architecture Tradeoff Analysis. *International Conference on Software Engineering*, 0:54, 1999.
- [11] R. Kazman, M. Klein, M. Barbacci, T. Longstaff, H. Lipson, and J. Carriere. The Architecture Tradeoff Analysis Method. In *Fourth IEEE International Conference on Engineering of Complex Computer Systems*, pages 68–78, 1998.
- [12] M. Mattsson, H. Grahn, and F. Mårtensson. Software Architecture Evaluation Methods for Performance, Maintainability, Testability and Portability. In *Second International Conference on the Quality of Software Architectures*, June 2006.
- [13] M. Neil, N. Fenton, and L. Nielsen. Building Large-Scale Bayesian Networks. *Knowledge Engineering Rev.*, 15(3):257–284, 2000.
- [14] A. Omerovic, A. Andresen, H. Grindheim, P. Myrseth, A. Refsdal, K. Stølen, and J. Ølnes. A Feasibility Study in Model Based Prediction of Impact of Changes on System Quality. Technical Report A13339, SINTEF, 2010.
- [15] A. Omerovic, A. Andresen, H. Grindheim, P. Myrseth, A. Refsdal, K. Stølen, and J. Ølnes. A Feasibility Study in Model Based Prediction of Impact of Changes on System Quality. In *International Symposium on Engineering Secure Software and Systems*, volume LNCS 5965, pages 231–240. Springer, 2010.
- [16] A. Omerovic, B. Solhaug, and K. Stølen. Evaluation of Experiences from Applying the PREDIQT Method in an Industrial Case Study. Technical Report A17562, SINTEF, 2011.
- [17] A. Omerovic and K. Stølen. Interval-Based Uncertainty Handling in Model-Based Prediction of System Quality. volume 0, pages 99–108. IEEE Computer Society, 2010.
- [18] R. K. Yin. *Case Study Research: Design and Methods, Third Edition, Applied Social Research Methods Series, Vol 5*. Sage Publications, Inc, 3 edition, 2002.