

Interval-based Uncertainty Handling in Model-based Prediction of System Quality

Aida Omerovic*[†] and Ketil Stølen*[†]

*SINTEF ICT, Pb. 124, 0314 Oslo, Norway

[†]University of Oslo, Department of Informatics, Pb. 1080, 0316 Oslo, Norway

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

Abstract—Our earlier research indicates feasibility of applying the PREDIQT method for model-based prediction of impacts of architecture design changes on system quality. The PREDIQT method develops and makes use of the so called prediction models, a central part of which are the “Dependency Views” (DVs) – weighted trees representing the relationships between system design and the quality notions. The values assigned to the DV parameters originate from domain expert judgments and measurements on the system. However fine grained, the DVs contain a certain degree of uncertainty due to lack and inaccuracy of empirical input. This paper proposes an approach to the representation, propagation and analysis of uncertainties in DVs. Such an approach is essential to facilitate model fitting, identify the kinds of architecture design changes which can be handled by the prediction models, and indicate the value of added information. Based on a set of criteria, we argue analytically and empirically, that our approach is comprehensible, sound, practically useful and better than any other approach we are aware of.

Keywords—uncertainty; system quality; prediction; modeling; architecture design.

I. INTRODUCTION

An important aspect of quantitative prediction of system quality lies in the appropriate representation, propagation and interpretation of uncertainty. We have developed and tried out the PREDIQT method [1] aimed for predicting impacts of architecture design changes on system quality characteristics and their trade-offs. Examples of quality characteristics include availability, scalability, security and reliability. One of the main artifacts of the PREDIQT method are the DVs. The DVs currently rely on sharp parameter values which are based on empirical input. As such, the parameters assigned to the DVs are not very reliable, thus providing predictions of unspecified certainty.

Since the input to the DVs is based on both data acquisition (measurements, logs, monitoring, historical data, or similar) and expert judgments, the representation of the uncertain input should be intuitive, as exact as possible and provide a well defined (complete and sound) inferring mechanism. In a real-life setting, making the right balance between the accuracy and the practical feasibility is what characterizes the overall challenge in selecting the appropriate approach to uncertainty handling in prediction models. We propose an approach to deal with uncertainty which, as we will argue, is both formally sound and practically applicable. Our approach is based on intervals with associated

confidence level, and allows representation, propagation and analysis of all the parameters associated with uncertainty.

The paper is organized as follows: The challenge of uncertainty handling in the context of the PREDIQT method is characterized in Section II. We define the frame within which the approach should be applicable, by providing an introduction to the DVs, and outlining a set of success criteria. The approach is presented in Section III. Section IV argues for soundness and practicability of the approach, by evaluating it with respect to the success criteria. An extensive number of the candidate methods for uncertainty handling have been systematically reviewed prior to proposal of our approach. Section V substantiates why our approach, given the criteria outlined in Section II, is preferred among the alternative ones. The concluding remarks and the future work prospects are given in Section VI.

II. THE CHALLENGE

Our earlier work indicates feasibility of applying the PREDIQT method for model-based prediction of impacts of architecture design changes, on the different quality characteristics of a system. The PREDIQT method produces and applies a multi-layer model structure, called prediction models, which represent system relevant quality concepts (through “Quality Models”) and architecture design (through “Design Models”). In addition, the prediction models comprise DVs, which are deduced from the design models and the quality models of the system under analysis, and represent the dependencies between architecture design and quality in the form of multiple weighted and directed trees.

A. Dependency views

The DVs are the central part of the prediction models, quantifying the fulfillment of the quality notions at the various abstraction levels of the architecture design, as well as the degrees of impact between the dependent model elements. One DV is, according to the process of the PREDIQT method, developed for each quality characteristic defined by the quality models of the system. A DV comprises two notions of parameters:

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

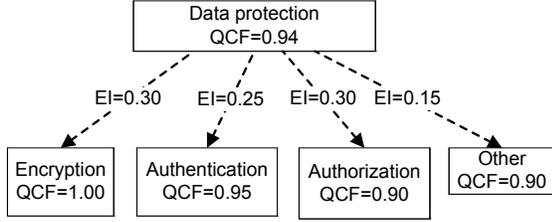


Figure 1. Excerpt of a DV

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 1 shows a small excerpt of a DV. In the case of the *Encryption* node of Figure 1, 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 [2]. A QCF value on a DV expresses to what degree the node (representing system part, concern or similar) is realised 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. The EI of an arc captures the impact of the child node on its parent node, with respect to the sub-characteristics (defined in the quality models) of the quality characteristic under consideration. Once a sum of the contributions of the sub-characteristics is obtained on each arc pointing to children nodes with a common parent, the EI values are normalized so that they sum up to 1. “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.

Once the parameter values are assigned, the QCF value of each non-leaf node is recursively (starting from leaf nodes and moving upwards in the tree) propagated by multiplying the QCF and EI value for each immediate child and summing up these products for all the immediate children. This is referred to as the general DV propagation algorithm. For example, with respect to *Data protection* node on Figure 1 (denoting: DP: Data protection, E: Encryption, AT: Authentication, AAT: Authorization, and O:Other):

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

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 repre-

sented 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 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 with 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. The orthogonality requirement puts additional demands on the DV structuring, which in practice is performed by the analyst, by deducing the DVs based on the underlying design models, quality model and the conceptual model, as prescribed by the PREDIQT process. It has however shown to represent a significant advantage during parametrization.

When an architecture design change is introduced in the design models, it is with aid of a conceptual model and according to the definitions in the quality model, reflected to the relevant parts of the DV and then propagated. Thereafter, the DV provides quantitative predictions of the new quality characteristic values. See [3] for further details on the PREDIQT method.

B. Uncertainty

The empirical input is always associated with a degree of uncertainty. Uncertainty is generally categorized into two different types: aleatory (due to inherent randomness of the system or variability of the usage profile) and epistemic (due to lack of knowledge or information about the system) [4]. The aleatory uncertainty is irreducible even by additional measurements. Aleatory uncertainty is typically represented by continuous probability distributions and forecasting is based on stochastic models. Epistemic uncertainty, on the other hand, is reducible, non-stochastic and of discrete nature. The epistemic uncertainty is therefore best suited for possibilistic uncertainty representations. For a detailed classification of the types and sources of imperfect information, along with a survey of methods for representing and reasoning with the imperfect information, see [5]. As op-

posed to for example weather forecasting models which are of stochastic and continuous nature and where the aleatory uncertainty is the dominating one (due to uncontrollable variabilities of many simultaneous factors), system quality models are characterized by rather discrete, sudden, non-stochastic and less frequent changes. In majority of the system quality prediction models, aleatory uncertainty is negligible in terms of magnitude and impact, while the epistemic one is crucial. It is therefore the epistemic uncertainty we focus on when dealing with the parameters on the DVs.

C. Success criteria

Since expert judgments are a central source of input during development of the prediction models, and also partially during the model verification, it is crucial that the formal representation of uncertainty is comprehensible to those who have in-depth system knowledge, but not necessarily a profound insight into the formal representation. The representation of uncertain estimates should facilitate provision of the uncertain DV parameter estimates which are easy for domain experts to provide and interpret.

Simultaneously, each individual parameter estimate should express the associated uncertainty so that it is as exact as possible. That is, the parameter and uncertainty values provided should be as fine grained as possible to provide, but without restricting the representation from being comprehensible. Thus, the right granularity of the uncertainty representation at the level of each parameter is needed.

Moreover, the input representation should facilitate combining both expert judgment based and measurement based input at the level of each parameter.

The DV propagation algorithm has a number of associated prerequisites (e.g., completeness, independence of the nodes having a common parent and ranges that the EI and QCF values can be expressed within). Together, they restrict the inference and the structure of the DVs so that the DVs become sound and comprehensible. When the parameters with the uncertainty representation are propagated within and across the DVs, the inference must still be well defined and sound.

When applied on real-life cases, the uncertainty handling approach should propagate to practically useful predictions, in the sense that the approach can be applied on realistic DVs with limited effort and give valuable output.

Statistical and sensitivity analyses are currently performed on the DVs, during model fitting and model application, respectively. Therefore, the uncertainty handling approach should also allow deduction of the central tendency measures such as mode, median, arithmetic mean, geometric mean, and variance.

Given the overall objective and context, the main success criteria for the uncertainty handling approach can, in a prioritized order, be summarized into:

- 1) The representation of a parameter estimate and its uncertainty should be comprehensible for the domain experts involved in the development and use of the prediction models.
- 2) The representation of each parameter estimate and its uncertainty should be as exact as possible, in terms of expressing both the parameter estimate and the associated uncertainty.
- 3) The approach should facilitate combining both expert judgment based and measurement based input.
- 4) The approach should correctly propagate the uncertain estimates.
- 5) The approach should provide practically useful results.
- 6) The approach should allow statistical analysis.

III. OUR SOLUTION

This section presents an interval-based approach to representation and propagation of uncertainties on the DVs.

A. Uncertainty representation

All prior estimates (the terms “prior estimate” and “initial estimate” are used interchangeably, and regard the intervals directly assigned to the EIs and leaf node QCFs, i.e., the parameters based on the empirical input and assigned before the non-leaf node QCFs may be inferred) are expressed in terms of intervals within which the correct parameter values should lie. The width of the interval is proportional to the uncertainty of the domain experts or deduced from the standard deviation of the measurement based input represented with probabilistic notions. In the latter case, the standard deviation indicates the accuracy of the measurements associated with each initially estimated parameter. Thus, the interval width may vary between the individual parameters. The representation of the uncertain estimates is exemplified through an excerpt of a DV shown in Figure 2.

In addition to the quantifiable uncertainty associated with each initially estimated parameter, there may exist sources of uncertainty which are general for the context or the system itself, but to a lesser degree expressive or measurable. Examples include the presence of the aleatory uncertainty, the competence of the domain experts, data quality, statistical significance, etc. Such factors contribute to the overall uncertainty, but are (due to their weak expressiveness or generality) not explicitly taken into account within the initially estimated EIs and the leaf node QCFs. Another reason for not accounting them within the intervals is because they are unavailable or may be biased at the individual parameter level. The domain experts may for example be subjective with respect to the above exemplified factors, or the tools for data acquisition may be unable of providing the values regarding data quality, statistical significance, etc. Therefore, the context related uncertainty should, from an impartial perspective (e.g., by a monitoring system or a panel, and

based on a pre-defined rating), be expressed generally for all prior estimates.

Hence, we introduce the “confidence level” as a measure of the expected probability that the correct value lies within the interval assigned to a prior estimate. The confidence level is consistent and expresses the overall, uniform, context or system relevant certainty, in terms of a percentage. The confidence level regards the prior estimates only. The confidence level dictates the width of the intervals of the prior estimates, i.e., the certainty with which the exact value is within the interval assigned to a prior estimate. For example, a confidence level of 100% guarantees that the exact values lie within the intervals assigned to the prior estimates. Obviously, a requirement for increased confidence level will result in wider intervals of the prior estimates. In the case of Figure 2 the prior estimates are assigned with a confidence level of 90%. Let QCFs and EIs be represented by intervals of type x :

$$x = [\underline{x}; \bar{x}] = \{X \in [0; 1] : \underline{x} \leq X \leq \bar{x}\} \quad (2)$$

where \underline{x} is the minimum estimated parameter value above which the exact value should (the term “should” is intentionally used in order to account for the confidence level of the prior estimates which is below 100%) lie, while \bar{x} is the maximum parameter value below which the exact value should lie. Both \underline{x} and \bar{x} are represented by real values. The interval x of a prior estimate is assigned with the confidence level specified. Due to model completeness, EIs on the arcs pointing to the nodes with a common parent must satisfy:

$$(\sum_{i=1}^I \underline{x}_i) \leq 1 \wedge (\sum_{i=1}^I \bar{x}_i) \geq 1 \quad (3)$$

where i denotes index of an arc, I denotes the total number of the arcs with outsprung from a common parent, and x_i denotes the interval estimate for the EI on arc i . That is, there must exist at least one subset of scalars from within each one of the intervals (representing EIs on the arcs to nodes with a common parent), whose sum is equal to 1.

B. Uncertainty propagation

The initial estimates are provided in the form of intervals with respect to a confidence level, as specified above. The propagation of the initially estimated intervals on the non-leaf node QCFs is given by the existing DV propagation algorithm (exemplified by Eq. 1 in Section II), the interval arithmetics [6], [7], and the algorithms for non-linear optimization [8], [9]. The result of the propagation is in the form of intervals of QCF values on the non-leaf nodes.

The confidence level itself is not propagated but only used in the context of the assignment of the initial estimates. Therefore, the confidence level is only associated with the initial estimates and not the inferred ones (non-leaf node QCFs). The confidence level does however affect the width of the inferred parameters through the width of the initial estimates. That is, since a requirement for a higher confidence

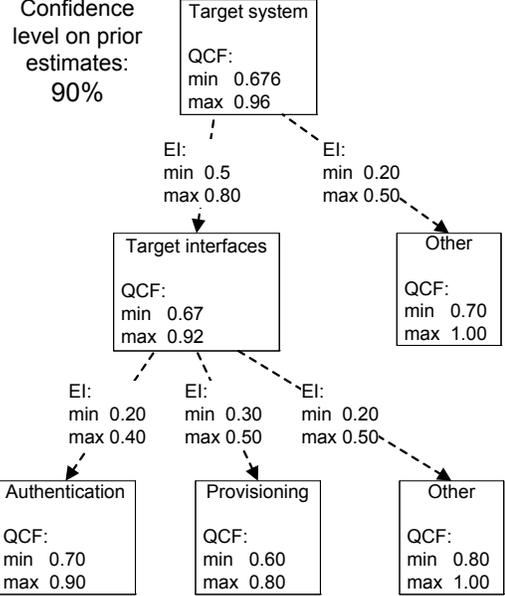


Figure 2. Excerpt of a DV with intervals and confidence level

level implies wider intervals of the initial estimates, the propagation will, as specified below, result in wider intervals on the non-leaf node parameters.

The only two interval arithmetic operations needed for propagation in a DV are addition and multiplication. In case of two intervals denoted by x and y (of the form given by Eq. 2), addition and multiplication are defined as:

$$x \circ y = [\underline{x} \circ \underline{y}; \bar{x} \circ \bar{y}] \quad (4)$$

Where \circ denotes the operation symbol.

The optimization is necessary for obtaining the extreme values (the maximum and the minimum) of the interval of a parent node in the cases when several combinations (within the propagated intervals) give a sum of the EIs (on the arcs pointing to the immediate children) equal to 1. The scalar points (from within the intervals involved), which provide the extreme values, are identified by the non-linear optimization algorithms and then inferred to the parent node QCF in the form of an interval, according to the general DV propagation algorithm.

For a set of EI intervals whose total sum of the upper interval values is more than 1, there may be infinitely many combinations (the number of the combinations depends on the number of decimal digits, which the scalars from the intervals are represented with) of scalar points from within all the intervals, which together sum up to 1. Regardless of how many EIs (or nodes) there are, finding the min and the max values of the interval resulting from the propagation (sum of products of QCF and EI values associated with respectively the immediate children and the arcs pointing to them) is a feasible optimization problem [9], [6]. Since the number of unknowns is equal to the number of equations involved, the only condition for the feasibility of the algorithm is the one

expressed by Eq. 3.

Let $\underline{qcf}, \overline{qcf} \in [0; 1]$ denote the interval limits of the QCFs on the immediate children and let $\underline{ei}, \overline{ei} \in [0; 1]$ denote the EIs on their respective interconnecting arcs. We propose the utility functions for the inferred min and max for the intervals of the parent node QCFs, which are given by respectively:

$$\min \left\{ \sum_{i=1}^I \underline{qcf}_i \cdot ei_i \mid \forall i \in I : \underline{ei}_i \leq ei_i \leq \overline{ei}_i \wedge \sum_{i=1}^I ei_i = 1 \right\} \quad (5)$$

$$\max \left\{ \sum_{i=1}^I \overline{qcf}_i \cdot ei_i \mid \forall i \in I : \underline{ei}_i \leq ei_i \leq \overline{ei}_i \wedge \sum_{i=1}^I ei_i = 1 \right\} \quad (6)$$

I and i denote the same notions as in Eq. 3. The inference starts from the lowest internal nodes, and proceeds recursively upwards the tree.

The sensitivity of the inferred interval width of a dependent node, on the interval width of a dependee (node or arc), can be deduced by:

- 1) estimating the initial uncertain parameters and propagating them
- 2) obtaining the inferred interval width W of the selected dependent node
- 3) removing (or partially reducing) the interval width of the selected dependee node or arc D
- 4) obtaining the new inferred interval width W' of the dependent node
- 5) calculating the sensitivity S between the dependent node W and the dependee parameter D , with respect to uncertainty. We define the sensitivity measure $S_{W,D}$ as:

$$S_{W,D} = \left(1 - \frac{W'}{W}\right) \quad (7)$$

In the context of predicting the quality characteristic values, the natural choice of the dependent node will be the root node, which represents the quality characteristic that the DV is dedicated to, while the dependee will be a leaf node QCF or an EI. The QCF value on the root node will then represent the value of the quality characteristic of the system. The dependee is subject to the initial estimation and whose uncertainty may be directly adjustable (for example, by reducing interval width due to added input). The sensitivity value can be obtained prior to selecting the candidate parameters for uncertainty reduction through added input. Then, the sensitivity value from above is related to the effort needed for acquisition of the additional input.

1) *The propagation in practice:* Currently, we run the optimization in *Matlab*, where the utility function is defined as the sum of products of the QCF and EI intervals (the DV model exemplified by Eq. 1) related to the immediate children nodes, and its constraints are:

- all QCF intervals involved
- all EI intervals involved and
- $\sum_{i=1}^I ei_i = 1$ (where i denotes an arc, I is the total number of the arcs pointing to the nodes with the com-

mon parent under consideration, and ei_i is a variable representing the EI value on the arc i). This constraint ensures the model completeness.

The minimum of the inferred interval is obtained from the utility function, while the maximum of the inferred interval is obtained by inverting the sign on the left hand side of the utility function and re-running the non-linear optimization algorithm. The *Target interfaces* and *Target system* nodes on Figure 2 are examples where such an algorithm had to be run in order to obtain the propagated intervals. In the case of *Target interfaces*, the utility function is specified in *Matlab* as:

```
function f = objfun(x,y)
f = x(1)*x(2)+x(3)*x(4)+x(5)*x(6);
```

Where $x(1)$, $x(3)$ and $x(5)$ represent the EI values on the arcs pointing to the *Authentication*, *Provisioning* and *Other* nodes, respectively; while $x(2)$, $x(4)$ and $x(6)$ represent the QCF values on the *Authentication*, *Provisioning* and *Other* nodes, respectively.

The related nonlinear inequality constraints representing the max and the min interval values of each respective variable specified above are defined in *Matlab* as:

```
c = [-x(1) + 0.2; x(1) - 0.4; -x(2) + 0.7; x(2) - 0.9;
      -x(3) + 0.3; x(3) - 0.5; -x(4) + 0.6; x(4) - 0.8;
      -x(5) + 0.2; x(5) - 0.5; -x(6) + 0.8; x(6) - 1.0];
```

The nonlinear equality constraint specifying that the sum of the EIs has to equal to 1, is defined in *Matlab* as:

```
ceq = [x(1) + x(3) + x(5) - 1];
```

The optimization algorithm is run by the following command in *Matlab*:

```
x0 = [0,0,0,0,0,0]; % Make a starting guess at the solution
options = optimset('LargeScale','on');
[x, fval] = ...
fmincon(@objfun,x0,[],[],[],[],[],[],@confunecq,options)
```

Providing the following result, where the values in the vector x specify the scalar points within the intervals $x(1)$ - $x(6)$, which yield the min value 0.67 of the utility function:

```
x = 0.3000 0.7000 0.5000 0.6000 0.2000 0.8000
fval = 0.6700
```

The max of the inferred interval is specified in *Matlab* by changing the sign of the above shown utility function to:

```
f = -(x(1)*x(2)+x(3)*x(4)+x(5)*x(6));
```

and re-running the command from above. The output obtained is:

```
x = 0.2000 0.9000 0.3000 0.8000 0.5000 1.0000
fval = 0.9200
```

where the values in the vector x specify the scalar points within the intervals $x(1)$ - $x(6)$, which yield the max value of the utility function, namely 0.92.

The propagation results are displayed on Figure 2. We see that the scalar points of the optimization output are in accordance with the Eq. 5 and Eq. 6.

The optimization is performed most efficiently when the estimates are represented in the form of matrices, where the utility functions are given by the DV propagation algorithm at the needed abstraction level. The interval-based approach is well suited in global optimization algorithms and is much faster than the competing point methods.

IV. WHY OUR SOLUTION IS A GOOD ONE

This section argues that the approach presented above fulfills the success criteria defined in Section II. Each one of the six criteria is considered in a dedicated subsection.

A. Criterion 1

The interval-based approach extends the DV parameters with the notions of interval widths and confidence level. They are based on fairly intuitive and simple definitions. Hence, the approach should be relatively easy for the domain experts to use and understand, regardless of the degree of their formal background. The simplicity also makes it less prone to unstable over-fitting, as well as bias or inaccuracy of the estimations.

B. Criterion 2

The interval width can be selected at the individual prior estimate level, thus allowing adjustment of granularity of the uncertainty representation. The number of the decimal digits used in estimation and propagation is unlimited.

C. Criterion 3

The domain expert judgments come directly in terms of intervals with a confidence level. However the measurement based input may come in terms of statistical notions.

Given that the measurement based input is normally distributed, the interval end points can be calculated as [10]:

$$\mu \pm t(1 - \text{conf}, n - 1)\sigma\sqrt{\frac{1}{n} + 1} \quad (8)$$

where $t(1 - \text{conf}, n - 1)$ is the two-tailed value of the Students t-distribution for the confidence level $1 - \text{conf}$ and $n - 1$ degrees of freedom, $\mu \in [0; 1]$ is the mean value, σ is the standard deviation of the measurements and n is the number of measurements. The “1” term inside the square root describes the spread of the measurement accuracy, while the “1/n” term describes the spread of the mean measurement accuracy. When n is high, there will be almost no uncertainty about the mean measurement accuracy, but the spread of the measurement accuracy may still be large. One can express both QCFs and EIs in this manner (for the relationship between the DV parameters and the measurements, see [3]), while requiring that Eq. 2 and Eq. 3 are satisfied. Alternatively, one can represent the QCF values in this manner, and the EI value of each related arc as a probability $p \in [0; 1]$, while enforcing $\sum p = 1$ for all nodes having a common parent. Thus, both kinds of input are transformable to intervals, which then can be propagated as defined in Section III and exemplified below.

D. Criterion 4

A consequence of the inequality and equality constraints is that all the inferred values are guaranteed to lie within the interval $[0; 1]$. As a result of this and the fact that the rating of a quality characteristic is defined so that all possible values

always must lie within this interval, the prediction output is an interval within which the exact value should lie, because the propagation algorithm calculates both the upper and the lower extreme values. Two aspects are hindering from guaranteeing that the factual value lies within the inferred interval:

- the confidence level with which the prior estimates are provided, and
- the aleatory uncertainty, which unless accounted for in the confidence level, is not quantified within the intervals.

E. Criterion 5

The interval-based approach has also been applied by providing example values of uncertain estimates on a real DV structure. The DV structure was originally used in a feasibility study of the PREDIQT method [3], performed on an extensive, real system. The uncertain estimates were straight forward to provide by referring to the quality characteristic rating definition and expressing the estimates in terms of the intervals. The interval width was mostly subject to observability of the parameter and existence of relevant historical input. The size of the DVs was 38 leaf nodes, 9 internal nodes and a root node. The number of EIs on the arcs was 47. Thus, the number of initial (empirical input based) estimates is 85, in this case. All initial estimates were expressed with intervals of reasonable and varying widths, within 90% confidence level. Once the initial estimates were in place, the propagation was quick and straightforward.

Table I summarizes the intervals applied. The first column lists statistics: number of elements, the maximum interval width, the minimum interval width, average interval width and standard deviation of the interval width. The second and the third columns show the value of each statistic for the initial estimates of the leaf node QCFs and all the EIs, respectively. The fourth column shows the value of each statistic for the initial estimates of both the leaf node QCFs, and all the EIs. The last column shows the value of each statistic for all the propagated QCFs (on the internal nodes and the root node). The resulting interval width of the root node QCF was 0.032. Given the attempt to provide as realistic and as variable interval widths of the initial estimates as possible, the example should be an indication of the expected findings in similar settings. Note that, while the interval widths reflect the realistic uncertainty, all values assigned are fictitious due to their confidentiality. The obtained root node interval width can be considered as a promising result, since the predictions are still likely to be associated with limited and acceptable uncertainty.

To test impact of uncertainty elimination on one leaf node (a child node of the root node) on the above presented DV, its QCF was changed from $[0.90; 0.95]$ to $[0.925; 0.925]$. The resulting root node QCF width became 0.0295 and the value of Eq. 7 became 0.081.

	Prior estimates 90% conf. level			Propagated
	QCFs	EIs	QCFs and EIs	QCFs
Count	38	47	85	10
Max	0.05	0.15	0.15	0.0645
Min	0.00	0.00	0.00	0.0141
Avg	0.027	0.02	0.025	0.0366
StDev	0.0199	0.023	0.022	0.014

Table I

SUMMARY OF THE INTERVALS APPLIED ON A REAL DV STRUCTURE

In a real-life setting, not all the estimates will be expressed with uncertainty, since some of the nodes have no impact or no uncertainty. The evaluation of the above mentioned feasibility study showed that the uncertainty of the input and the deviations between the PREDIQT based and the empirical predictions are relatively low. The experience from the feasibility study is that the interval widths would be quite small. Most of the nodes of the DVs were placed on the second or the third level, which considerably limits the vertical propagation of the uncertainties.

Reducing the confidence level and further model fitting (through additional input) are the measures when the inferred values are too uncertain. The candidate parameters for reduction of uncertainty can be identified evaluating the sensitivity measure proposed in Section III in relation to the effort needed for the uncertainty reduction in question. Alternatively, a sensitivity analysis supported by charts and central tendency measures can be pursued for observing the impact that a reduction of uncertainty of the individual estimates would have on (the root node of) the DV.

F. Criterion 6

The analysis of the central tendency measures of the interval-based estimates relies on the existing fully defined interval arithmetics and interval statistics [11]. Both can, in their existing well-established form, be directly applied in our context. For arithmetic mean, geometric mean, harmonic mean, weighted mean, median, standard deviation and variance, see [11]. In addition, [11] provides guidance regarding identification of outliers, trade-off between sample size and precision, handling of measurement uncertainty, handling of dependencies among the sources of uncertainty (correlation and covariance) and accounting for incertitude.

In case of the measurement analysis with respect to uncertainty, it is crucial to ensure reliability and validity [12]. In particular, although some of the estimates on a DV are structured so that they are independent, the measurements performed for their validation may not necessarily be such. In addition, the causes of uncertainty related to these measurements may correlate. The validation of the sub-characteristics, which the EIs are based on, may for example involve correlating measurements with to some degree same causes of uncertainty. The central tendency measures are the possible mechanisms for analyzing the validity. However, dependency and correlation are not the same notion, as argued by [13].

V. WHY OTHER APPROACHES ARE NOT BETTER IN THIS CONTEXT

A ratio scale is a measurement scale in which a certain distance along the scale means the same thing no matter where on the scale we are, and where “0” on the scale represents the absence of the thing being measured. Statistical analysis and arithmetics are supported for the ratio scale. The ratio scale is in fact used in Section II. We may for example introduce uncertainty representation by defining fixed increments on the scale from 0 to 1, and relating their meaning to the quality characteristic rating. The input would have to be expressed in the form of the increments defined, and the uncertainty would per definition range half the way to the neighboring increments. Obviously, this is a special case of the interval approach where the increments and their granularity are frozen at the model (and not parameter) level. By using a ratio scale in the PREDIQT context, the schema of the increments would have to be model general (in order for the uncertainty propagation to be meaningful) and therefore either too coarse grained or too fine grained. The variation of uncertainty between parameters would not be supported, thus violating criterion 2 from Section II.

The Dempster-Shafer structures [11] offer a way of representing uncertainty quantified by mass distribution functions. A mechanism for aggregation of such representation stored in distributed relational databases, is proposed by [14]. The Dempster-Shafer approach characterizes uncertainties as intervals with degrees of certainty (that is, sets of values with weights which add up to 1). It can be seen as a generalization of both interval analysis and probability theory. Weights of evidence are put on a collection of intervals and the structures may overlap. Implementing the Dempster-Shafer theory in our context would involve solving two issues: 1) sorting the uncertainties in the empirical input into a priori independent items of evidence, and 2) carrying out Dempster’s rule computationally. The former one leads to a structure involving input elements that bear on different but related concerns. This structure can be used to make computations based on Dempster’s rule, feasible. Our solution is a special case of the Dempster-Shafer approach, where the intervals of the prior estimates have a general confidence level, and the structure of the DV allows for a linear propagation. The additional expressiveness that the Dempster-Shafer structures offer is not needed in our context, since the certainty is highly unlikely to vary across the fractions of the intervals. In fact, such a mechanism will, due to its complex representation on subsets of the state space, in the PREDIQT context only compromise the comprehensibility of the uncertainty representation and therefore the correctness of the input.

Bayesian networks (BNs) [15], [16] allow incorporating both model uncertainty and parameter uncertainty. A BN is a directed acyclic graph in which each node has an

associated probability distribution. Observation of known variables (nodes) allows inferring the probability of others, using probability calculus and Bayes theorem throughout the model (propagation). BNs can represent and propagate both continuous and discrete uncertainty distributions. BNs are however demanding to parametrize and interpret the parameters of, which violates our first criterion. This issue has been addressed by [17] where an analytical method for transforming the DVs to Bayesian networks is presented. It also shows that DVs are, although easier to relate to in practice, compatible with BNs. It is possible to extend this method so that our interval-based approach is transformed to a BN before a further BN-based analysis can proceed. Such an extension would introduce several states on the BN nodes, and assign probabilities to them. In that manner, the extension would resemble to the Dempster-Shafer structures. BNs do not score sufficiently on our criteria 1 and 5.

Fuzzy logic provides a simple way to draw definite conclusions from vague, ambiguous or imprecise information, and allows for partial membership in a set. It allows modelling complex systems using higher levels of abstraction originating from the analyst's knowledge and experience [18]. A fuzzy set is a class of objects with a continuum of grades of membership. Such a set is characterized by a membership function, which assigns to each object a grade of membership ranging between zero and one [19]. Using the fuzzy membership functions, a parameter in a model can be represented as a crisp number, a crisp interval, a fuzzy number and a fuzzy interval. In fuzzy approach the algebraic operations are easy and straightforward, as argued and elaborated by [20]. The interval-based approach is a special case of the fuzzy approach, where only the crisp intervals are used as the membership function. The additional expressiveness that the overall types of the membership functions offer is in fact not needed in the PREDIQT context, since the increased complexity of the estimate representation would not contribute to the accuracy of the parameter values, but rather introduce misinterpretations and incorrectnesses in the input provision. The interpretation of the membership distributions and their correspondence to the practical settings in the PREDIQT context would be demanding.

Subjective logic [21] is a framework for artificial reasoning, which consists of a belief model called *opinion* and set of operations for combining opinions. A single opinion π is uniquely described as a point $\{b,d,i\}$ in an "Opinion Triangle", where b, d and i designate belief, disbelief and ignorance, respectively. For each opinion, the three notions sum up to unity. The operations formally defined include: conjunction, disjunction, negation, consensus, recommendation and ordering. The subjective logic is suited for the domain expert judgments, but how the measurement based input can be related to the concepts of the subjective logic, needs to be defined. Applying the subjective logic

in the PREDIQT context would however only increase the fulfillment of our second criterion beyond the needs, and on the expense of degrading the third criterion.

Uncertainty representation in software development effort estimation [22], [23] is most comparable to ours. However, they do not have as a strict criterion of propagation, and can therefore introduce different notions to the uncertainty representation.

It should be pointed out that the interval propagation based on the extreme values suffers from the so-called overestimation effect, also known as dependency problem. The dependency problem is due to the memoryless nature of interval arithmetic in cases when a parameter occurs multiple times in an arithmetic expression, since each occurrence of an interval variable in an expression is treated independently. Since multiple occurrence of interval parameters cannot always be avoided, the dependency problem may cause crucial overestimation of the actual range of an evaluated function. A way to approach this issue is to use interval splitting [24], where the input parameter intervals are subdivided and the arithmetics are preformed on the subintervals. The final results are then obtained by computing the minimum of all lower bounds and the maximum of all upper bounds of the intermediate results. Skelboe [25] has shown that the results obtained from the interval splitting converge to the actual range if the width of the subintervals approaches zero. Our solution does not use interval splitting, as it would significantly increase complexity of the entire approach, thus compromising the criterion 1.

The epistemic uncertainty is the crucial one and therefore given all the attention in our context. Being of a discrete nature, the epistemic uncertainty should, as argued in Section II, be handled by a purely possibilistic approach. The approaches mentioned in the remainder of this section focus to a high degree on the stochastic uncertainties, which is mainly what makes them less suited in the PREDIQT context.

The ISO approach to handling measurement uncertainty [26] uses a probabilistic representation with normal distribution, and treats both aleatory and epistemic uncertainty equally. Such an approach however does not explicitly account the notion of ignorance about the estimates, thus failing to intuitively express it.

A simulation mechanism, which takes into account both aleatory and epistemic uncertainty in an interval-based approach, is proposed by [27]. It concentrates on stochastic simulations as input for the interval estimates, when significant uncertainties exist. Moreover, [11] proposes considering a hybrid approach comprising both probabilistic and interval representation, in order to account for both aleatory and epistemic uncertainty. Neither of these two approaches would in the PREDIQT context increase fulfillment of our criteria. In fact, the systematic sources of uncertainty would not be represented more accurately, while comprehensibility

would degrade.

A hybrid Monte Carlo and possibilistic method for representation and propagation of aleatory and epistemic uncertainty is presented by [28]. The method is applied for predicting the time to failure of a randomly degrading component, and illustrated by a case study. The hybrid representation captures the aleatory variability and epistemic imprecision of a random fuzzy interval in a parametrized way through α -cuts and displays extreme pairs of the upper and lower cumulative distributions. The Monte Carlo and the possibilistic representations are jointly propagated. The gap between the upper and the lower cumulative distributions represents the imprecision due to epistemic variables. The possibility distributions are aggregated according to the so called Ferson method. The interpretation of the results in the form of limiting cumulative distributions requires the introduction of a degree of confidence directly connected with the confidence on the value of epistemic parameters. As compared to this approach, our solution is more comprehensible but less suited for handling the aleatory uncertainty. However, given our criteria, the former aspect outranges the latter one.

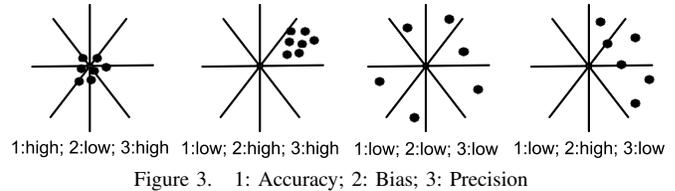
The approaches to uncertainty handling in other domains, such as weather forecasting [29], electricity demand forecasting [30], correlations between wind power and meteorological conditions [31], power system planning [32] and supply industry [33] are mainly based on probabilistic representations and stochastic simulations. They focus mainly on the aleatory uncertainty, which in the PREDIQT context is of secondary relevance.

Hence, given the criteria presented in Section II, the interval-based approach prevails as the most appropriate one in the PREDIQT context.

VI. CONCLUSION AND FUTURE WORK

Our earlier research indicates feasibility of the PREDIQT method for model-based prediction of impacts of architecture design changes on system quality. The PREDIQT method produces and applies a multi-layer model structure, called prediction models, which represent system design, system quality and the interrelationship between the two. A central part of the prediction models are the DVs, which are parametrized with respect to quality notions and impacts among the elements. Due to its empirical nature, input into the DVs is associated with uncertainty. By handling the uncertainty in the DVs, quality of the prediction models and accuracy of the predictions are made explicit, thus indicating which changes are predictable and whether further model fitting is needed.

Based on a set of criteria identified with respect to the PREDIQT method, we have proposed and evaluated an approach to uncertainty handling in the DVs. The approach relies on intervals with a confidence level, and covers representation, propagation and analysis of the uncertain param-



eters. The interval-based approach allows comprehensible representation of uncertainty on all kinds of parameters, with the needed accuracy. Estimation, propagation and analysis in the interval-based approach are scalable and efficient. Both the interval arithmetics, algorithms for non-linear optimization, and the statistical analysis of intervals are already fully established and can be applied in the PREDIQT context in their existing forms. The approach is entirely compliant with the existing version of the PREDIQT method. The evaluation argues for the correctness and practicability of our approach, as well as its outranging appropriateness relative to the alternative uncertainty handling approaches.

Further work will address analysis of the prediction accuracy, that is the deviation between the predicted and the actual quality characteristic values. The notions of magnitude of average deviation AD [3], balanced relative error BRE [34] and hit rate (i.e., the percentage of the correct values lying within the predicted intervals) can be used as measures of prediction accuracy. For an accurate prediction model, the hit rate should be consistent with the confidence level. The BRE allows analysis of bias and precision (see Figure 3) of the predictions. Thus, systematic and random variance of the prediction accuracy can be distinguished in a meta analysis of our uncertainty handling approach. The prospects of further work also include additional, case based evaluations of practicability and accuracy of the approach. Moreover, identifying and categorizing the variables that impact the uncertainty of the estimates is important for improving uncertainty management.

ACKNOWLEDGMENT

This work has been conducted within the DIGIT (180052/S10) project, funded by the Research Council of Norway.

REFERENCES

- [1] 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 *Proceedings of International Symposium on Engineering Secure Software and Systems ESSOS10*, vol. LNCS 5965. Springer, 2010, pp. 231–240.
- [2] International Organisation for Standardisation, "ISO/IEC 9126 - Software engineering – Product quality," 2004.
- [3] 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," SINTEF A13339, Tech. Rep., 2010.

- [4] A. D. Kiureghiana and O. Ditlevsenb, "Aleatory or epistemic? Does it matter?" *Structural Safety*, vol. 31, no. 2, pp. 105–112, 2009.
- [5] S. Parsons, "Current Approaches to Handling Imperfect Information in Data and Knowledge Bases," *IEEE Trans. on Knowl. and Data Eng.*, vol. 8, no. 3, pp. 353–372, 1996.
- [6] R. B. Kearfott, "Interval Computations – Introduction, Uses, and Resources," *Euromath Bull.*, vol. 2, pp. 95–112, 1996.
- [7] V. Kreinovich, J. G. Hajagos, W. T. Tucker, L. R. Ginzburg, and S. Ferson, "Propagating Uncertainty through a Quadratic Response Surface Model," Sandia National Laboratories Report SAND2008-5983, Tech. Rep., 2008.
- [8] J. Nocedal and S. J. Wright, *Numerical Optimization*. New York: Springer, 1999.
- [9] A. R. Ravindran, *Operations Research and Management Science Handbook*. CRC Press, 2008.
- [10] T. Wonnacott and R. Wonnacott, *Introductory Statistics*. New York: Wiley, 1990.
- [11] S. Ferson, V. Kreinovich, J. Hajagos, W. Oberkampf, and L. Ginzburg, "Experimental Uncertainty Estimation and Statistics for Data Having Interval Uncertainty," Sandia National Laboratories Report SAND2007-0939, Tech. Rep., 2007.
- [12] T. D. Cook and D. T. Campbell, *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Houghton Mifflin Company, 1979.
- [13] S. Ferson, R. Nelsen, J. Hajagos, D. Berleant, J. Zhang, W. T. Tucker, L. Ginzburg, and W. L. Oberkampf, "Myths about Correlations and Dependencies and their Implications for Risk Analysis," 2008.
- [14] B. Scotney and S. McClean, "Database Aggregation of Imprecise and Uncertain Evidence," *Inf. Sci. Inf. Comput. Sci.*, vol. 155, no. 3-4, pp. 245–263, 2003.
- [15] M. Neil, N. Fenton, and L. Nielsen, "Building Large-Scale Bayesian Networks," *Knowledge Engineering Rev.*, vol. 15, no. 3, pp. 257–284, 2000.
- [16] D. Heckerman, A. Mamdani, and W. M. P., "Real-World Applications of Bayesian Networks," *ACM Communications*, vol. 38, no. 3, pp. 24–26, 1995.
- [17] A. Omerovic and K. Stølen, "Simplifying Parametrization of Bayesian Networks in Prediction of System Quality," in *Proceedings of Third IEEE International Conference on Secure Software Integration and Reliability Improvement*. IEEE, 2009, pp. 447–448.
- [18] D. P. Weber, "Fuzzy Fault Tree Analysis," in *Proceedings of the 3rd IEEE Conference on EEE World Congress on Computational Intelligence*. IEEE, 1994, pp. 1899–1904.
- [19] L. A. Zadeh, "Fuzzy Sets," *Information and Control*, vol. 8, pp. 338–353, 1965.
- [20] P. V. Suresh, A. K. Babar, and V. Raj, "Uncertainty in fault tree analysis: a fuzzy approach," *Fuzzy Sets Systems*, vol. 83, no. 2, pp. 135–141, 1996.
- [21] A. Jøsang, "Artificial Reasoning with Subjective Logic," in *Proceedings of the 2nd Australian Workshop on Common-sense Reasoning*. Australian Computer Society, 1997.
- [22] S. Grimstad and M. Jørgensen, "Inconsistency of Expert Judgment-Based Estimates of Software Development Effort," *Journal of Systems and Software*, vol. 80, no. 11, pp. 1770–1777, 2007.
- [23] T. M. Gruschke and M. Jørgensen, "Assessing Uncertainty of Software Development Effort Estimates: Learning from Outcome Feedback," *ACM Transactions on Software Engineering and Methodology*, vol. 17, no. 4, pp. 20–35, 2008.
- [24] S. Majumdar, L. Johannes, G. Haring, and R. Ramadoss, "Characterization and Analysis of Software and Computer Systems with Uncertainties and Variabilities," in *Proceedings of Performance Engineering, State of the Art and Current Trends*, vol. LNCS 2047. Springer, 2001, pp. 202–221.
- [25] S. Skelboe, "Computation of Rational Interval Functions," *BIT Numerical Mathematics*, vol. 14, no. 1, pp. 87–95, 1974.
- [26] International Organisation for Standardisation, "Guide to the Expression of Uncertainty in Measurement," 1993.
- [27] O. G. Batarseh and Y. Wang, "Reliable Simulation with Input Uncertainties using an Interval-Based Approach," in *Proceedings of the 40th Conference on Winter Simulation*, 2008, pp. 344–352.
- [28] P. Baraldi, I. C. Popescu, and E. Zio, "Predicting the Time To Failure of a Randomly Degrading Component by a Hybrid Monte Carlo and Possibilistic Method," in *Proceedings of International Conference on Prognostics and Health Management 2008*. IEEE, 2008, pp. 1–8.
- [29] T. N. Palmer, "Predicting Uncertainty in Forecasts of Weather and Climate," *Rep.Prog.Phys.*, vol. 63, pp. 71–116, 2000.
- [30] J. W. Taylor and R. Buizza, "Using Weather Ensemble Predictions in Electricity Demand Forecasting," *International Journal of Forecasting*, vol. 19, no. 1, pp. 57–70, 2003.
- [31] M. Lange and D. Heinemann, "Accuracy of Short Term Wind Power Predictions Depending on Meteorological Conditions," in *Proceedings of Global Windpower Conference*, 2002.
- [32] A. P. Douglas, A. M. Breipohl, F. N. Lee, and R. Adapa, "Risk due to Load Forecast Uncertainty in Short Term Power System Planning," in *IEE Transactions on Power Systems*, vol. 13, no. 4. IEEE, 1998, pp. 1493–1499.
- [33] K. L. Lo and Y. K. Wu, "Risk Assessment due to Local Demand Forecast Uncertainty in the Competitive Supply Industry," in *IEE Proceedings on Generation, Transmission and Distribution*, vol. 150, no. 5. IEEE, 2003, pp. 573–581.
- [34] J. Armstrong, *Long-Range Forecasting*. Wiley, 1985.