A Practical Approach to Uncertainty Handling and Estimate Acquisition in Model-based Prediction of System Quality

Aida Omerovic*† and Ketil Stølen*†
*SINTEF ICT, P.O. Box 124, 0314 Oslo, Norway
†University of Oslo, Department of Informatics, P.O. Box 1080, 0316 Oslo, Norway
Email: {aida.omerovic,ketil.stolen}@sintef.no

***Abstract*—Our earlier research indicated the feasibility of applying the PREDIQT method for model-based prediction of impacts of architectural design changes on system quality. The PREDIQT method develops and makes use of so called prediction models, a central part of which are the "Dependency Views" (DVs) – weighted trees representing the relationships between architectural design and the quality characteristics of a target system. The values assigned to the DV parameters originate from domain expert judgements 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 (that is, adjustment of models during verification), identify the kinds of architectural 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 uncertainty handling approach is comprehensible, sound, practically useful and better than any other approach we are aware of. Moreover, based on experiences from PREDIQT-based analyses through industrial case studies on real-life systems, we also provide guidelines for use of the approach in practice. The guidelines address the ways of obtaining empirical estimates as well as the means and measures for reducing uncertainty of the estimates.**

***Keywords*-uncertainty, system quality prediction, modeling, architectural design, change impact analysis, simulation.**

I. Introduction

An important aspect of quantitative prediction of system quality lies in the appropriate representation, propagation and interpretation of uncertainty. Our earlier work has addressed this issue by proposing an interval-based approach to uncertainty handling in model-based prediction of system quality [1]. This paper extends the interval-based approach to uncertainty handling with two major tightly related issues:

- uncertainty analysis, and
- practical guidelines for use of the interval-based approach, addressing both the uncertainty handling and the estimate acquisition.

We have developed and tried out the PREDIQT method [2], [3] for model-based prediction of impacts of architectural design changes on system quality characteristics and their trade-offs. Examples of quality characteristics include availability, scalability, security and reliability. Among the main artifacts of the PREDIQT method are the Dependency Views (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 measurement-based data acquisition (measurements, logs, monitoring, historical data, or similar) and expert judgements, 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, finding the right balance between accuracy and practical feasibility is the main challenge when 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 in the PREDIQT context. Our approach is based on intervals with associated confidence level, and allows representation, propagation and analysis of all the parameters associated with uncertainty.

Input acquisition is in this context concerned with how the DV estimates and their uncertainty measures are obtained in practice. An overview of the practical means and measures for 1) acquiring the input and 2) achieving a specified minimum level of uncertainty, is clearly a prerequisite for applicability of the uncertainty handling approach. Therefore, we also provide guidelines for practical use of our solution, covering both the issues of estimate acquisition and uncertainty handling. The guidelines build on the experiences from the empirical evaluations of the PREDIQT method.

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 overview of the PREDIQT method and in particular the DVs, introducing the notion of uncertainty, and outlining a set of success criteria. The interval-based approach to uncertainty handling is presented in Section III. Section IV argues for the usefulness and practical applicability of the approach by evaluating it with respect to the success criteria. An extensive

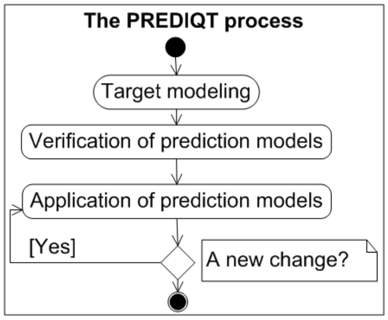

Figure 1. The overall PREDIQT process

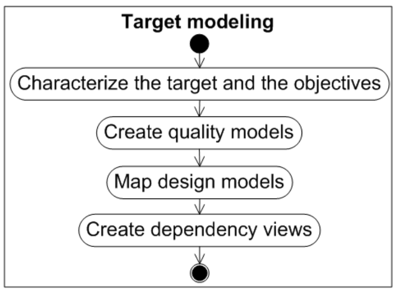

Figure 2. Target modeling phase

number of the candidate methods for uncertainty handling have been systematically reviewed prior to the proposal of our approach. Section V substantiates why our approach, given the criteria outlined in Section II, is preferred among the alternative ones. Practical guidelines for use of our solution, based on lessons learned from PREDIQT-based analyses on real-life systems, are provided in Section VI. The concluding remarks and the future work prospects are given in Section VII.

II. The challenge

Our earlier work indicates the feasibility of applying the PREDIQT method for model-based prediction of impacts of architectural design changes, on the different quality characteristics of a system. The PREDIQT method produces and applies a multi-layer model structure, called prediction models. The PREDIQT method is outlined in the next subsection. Uncertainty and the evaluation criteria for the uncertainty handling approach are thereafter presented in dedicated subsections.

A. Overview of the PREDIQT method

The PREDIQT method defines a process and a structure of prediction models. These two perspectives are presented in the following.

1) The process of the PREDIQT method: The process of the PREDIQT method consists of three overall phases as illustrated by Figure 1. Each of these phases is decomposed into sub-phases.

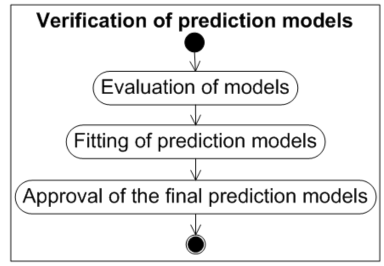

Figure 3. Verification of models – phase

The sub-phases within the "Target modeling" phase are depicted in Figure 2. 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 Model diagrams are created in the form of a tree, by decomposing total quality into the system specific quality characteristics, their respective sub-characteristics and indicators. The Quality Model diagrams represent a taxonomy with interpretations and formal definitions of system quality notions. The initially obtained Design Models are customized so that (1) only their relevant parts are selected for use in further analysis; and (2) a mapping within and across high-level design and low-level Design Models (if available), is made. The mapped models result in a class diagram, which includes the relevant elements and their relations only. A conceptual model (a tree-formed class diagram) in which classes represent elements from the underlying Design Models and Quality Models, relations represent the ownership, and the class attributes represent the dependencies or the properties, is created.

For each quality characteristic defined by the Quality Model, a quality characteristic specific DV is created via the instantiation of the conceptual model. A DV is basically a weighted dependency tree which models the relationships among quality characteristics and the design of the system. The instantiation of the conceptual model into a DV is performed by selecting the elements and relationships which are relevant to the quality characteristic being addressed by the DV. Each set of nodes having a common parent is supplemented with an additional node called "Other" for completeness purpose. The DV parameters are assigned by providing the estimates on the arcs and the leaf nodes, and propagating them according to a pre-defined inference algorithm.

The sub-phases within the "Verification of prediction models" phase are depicted in Figure 3. This phase 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

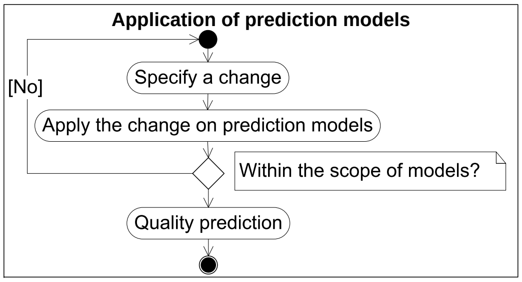

Figure 4. Application of models – phase

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 phase is re-initiated.

The sub-phases within the "Application of prediction models" phase are depicted in Figure 4. This phase involves applying the specified architectural design change on the prediction models and obtaining the predictions. The phase presupposes that the prediction models are approved. During this phase, a specified change is applied on 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 change specification should clearly state all deployment relevant facts necessary for applying the change. The "Apply the change on prediction models" sub-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 DVs. Thereafter, the DVs provide 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 [2] based on MS Excel [4] for simulation and sensitivity analysis of DVs.

The intended application of the prediction models does not include implementation of change on the target system, but only simulation of effects of the independent architectural design changes on quality of the target system (in its currently modelled state). Hence, maintenance of prediction models is beyond the scope of PREDIQT.

2) The prediction models: The PREDIQT method produces and applies a multi-layer model structure, called prediction models, which represent system relevant quality concepts (through "Quality Models") and architectural design (through "Design Models").

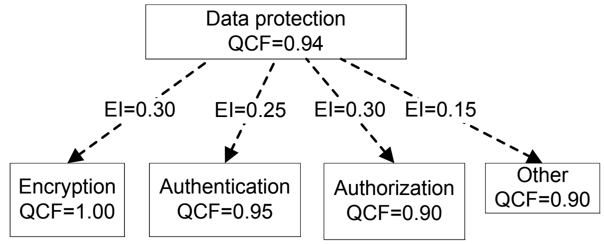

Figure 5. Excerpt of an example DV with fictitious values

The Design Models represent the architectural design of the target system. The models include the parts and the detail level characterized (during the first sub-phase of the PREDIQT process) as a part of the objective of the analysis. Typically, Design Models include diagrams representing the process, the system structure, the dataflow and the rules for system use and operation. The Design Model diagrams are used to specify the target system and the changes whose effects on quality are to be predicted.

A Quality Model is a tree-like structure whose nodes (that is, quality notions at the different levels) are defined qualitatively and formally, with respect to the target system. The total quality of the system is decomposed into characteristics, sub-characteristics and quality indicators. Each of them is, by the Quality Model, defined in terms of a metric and an interpretation with respect to the target system. The definitions of the quality notions may for example be based on ISO 9126 product quality standard [5].

In addition, the prediction models comprise DVs, which are deduced from the Design Models and the Quality Models of the system under analysis. As explained above, the DVs model 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. The values and the dependencies modeled through the DVs are based on the quality characteristic definition provided by the Quality Model. A DV comprises two notions of parameters:

1) EI: Estimated degree of Impact between two nodes, and
2) QCF: estimated 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 5 shows an excerpt of an example DV with fictitious values. In the case of the *Encryption* node of Figure 5, the QCF value expresses the goodness of encryption with respect to the quality characteristic in question, e.g., security. A QCF value on a DV expresses to what degree the node (representing system 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. The EI of an arc captures the impact of the child node on its parent node, with respect to the quality characteristic under consideration.

Input to the DV parameters may come in different forms (e.g., from domain expert judgements, experience factories, measurements, monitoring, logs, etc.), during the different phases of the PREDIQT method. Once the initial 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 in Figure 5 (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 represented by the node), respectively. This constraint is ensured through the normalized definition of the quality characteristic metric. 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 overall concerns are covered by the nodes denoted *Other*, which are included in each set of nodes having a common parent, thus making the DV complete.

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 parameterization 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 been shown to represent a significant advantage during the estimation.

Figure 6 provides an overview of the prediction models, expressed as a UML [6] class diagram. A prediction model is decomposed into a Design Model, a Quality Model and a DV. A Quality Model is a set of tree-like structures. Each tree is dedicated to a target system-relevant quality characteristic. Each quality characteristic may be decomposed into quality sub-characteristics, which in turn may be decomposed into a set of quality indicators. As indicated by the relationship of type aggregation, specific sub-characteristics and indicators can appear in several Quality Model trees dedicated to the different quality characteristics. Each element of a Quality Model is assigned a quantitative normalized metric and an interpretation (qualitative meaning of the element), both specific for the target system. A Design Model represents the relevant aspects of the system architecture, such as for example process, dataflow, structure and rules. A DV is a weighted dependency tree dedicated to a specific quality characteristic defined through the Quality Model. As indicated by the attributes of the Class *Node*, the nodes of a DV are assigned a name and a QCF (that is, value of the degree of fulfillment of the quality characteristic, with respect to what is represented by the node). As indicated by the *Semantic* dependency relationship, semantics of both the structure and the weights of a DV are given by the definitions of the quality characteristics, as specified in the Quality Model. A DV node may be based on a Design Model element, as indicated by the *Based on* dependency relationship. As indicated by the self-reference on the Class *Node*, one node may be decomposed into children nodes. Directed arcs express dependency with respect to quality characteristic by relating each parent node to its immediate children nodes, thus forming a tree structure. Each arc in a DV is assigned an EI, which is a normalized value of degree of dependence of a parent node, on the immediate child node. The values on the nodes and the arcs are referred to as parameter estimates. We distinguish between prior (or initial) and inferred parameter estimates. The former ones are, in the form of empirical input, provided on leaf nodes and all arcs, while the latter ones are deduced using the DV propagation model for PREDIQT exemplified above. For further details on the PREDIQT method, see [2], [3], [7], [1].

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 of knowledge or information about the system) [8]. The aleatory uncertainty is irreducible even by additional measurements. Aleatory uncertainty is typically represented by continuous probability distributions and forecasting is

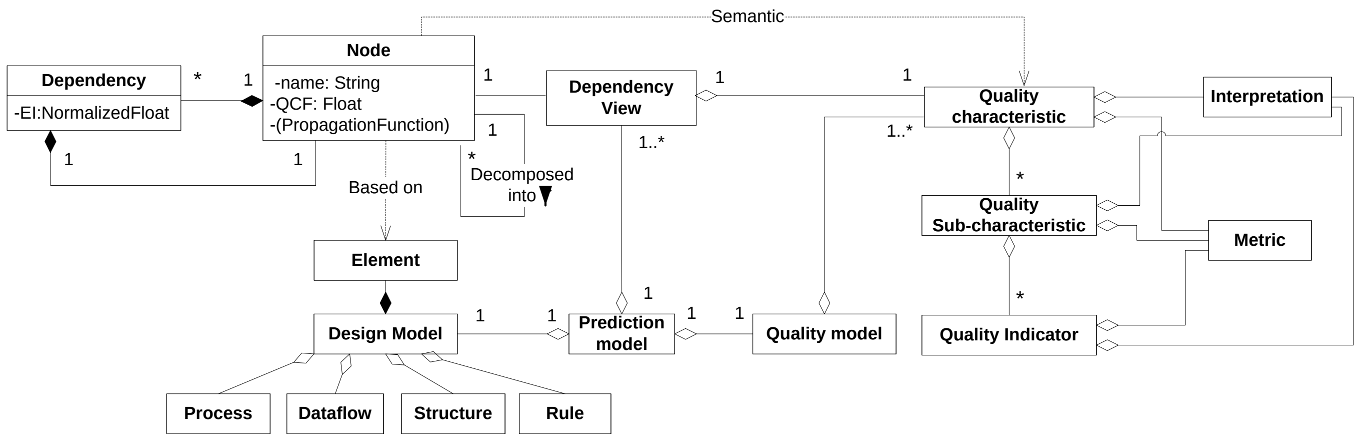

Figure 6. An overview of the elements of the prediction models, expressed as a UML class diagram

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 [9]. For a systematic literature review of the approaches for uncertainty handling in weighted dependency trees, see [10].

Prediction models, as opposed to for example weather forecasting models, are characterized by rather discrete, sudden, non-stochastic and less frequent changes. The weather forecasting models are of stochastic and continuous nature and the aleatory uncertainty is the dominating one (due to uncontrollable variabilities of many simultaneous factors). 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 judgements are a central source of input during the 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 form of uncertainty estimates should make them 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 comprehensibility. 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 judgement-based and measurement-based input at the level of each parameter in a DV.

The DV propagation algorithm has a number of associated prerequisites (e.g., completeness, independence of the nodes which have 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 in the DVs, during the *Fitting of prediction models* sub-phase and the *Application of prediction models* phase (of the PREDIQT process), 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 form of each 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 form 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 judgement-based and measurement-based input.
4) The approach should correctly propagate the estimates and their uncertainty.
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 estimates and their uncertainty is exemplified through an excerpt of a DV (with fictitious values) shown in Figure 7.

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) 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 incapable of providing the values regarding data quality, statistical significance, etc. Therefore, the context related uncertainty should, from an unpartial 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 7 the prior estimates are assigned with a confidence level of 90%. Let QCFs and EIs be represented by intervals of type x:

$$x = [\underline{x}; \overline{x}] = \{X \in [0;1] : \underline{x} \leq X \leq \overline{x}\} \tag{2}$$

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 $\overline{x}$ is the maximum parameter value below which the exact value should lie. Both $\underline{x}$ and $\overline{x}$ are represented by real numbers. 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:

$$\left(\Sigma_{i=1}^{I} \underline{x}_i\right) \leq 1 \wedge \left(\Sigma_{i=1}^{I} \overline{x}_i\right) \geq 1 \tag{3}$$

where i denotes index of an arc, I denotes the total number of the arcs with outspring 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 [11], [12], and the algorithms for non-linear optimization [13], [14]. 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 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:

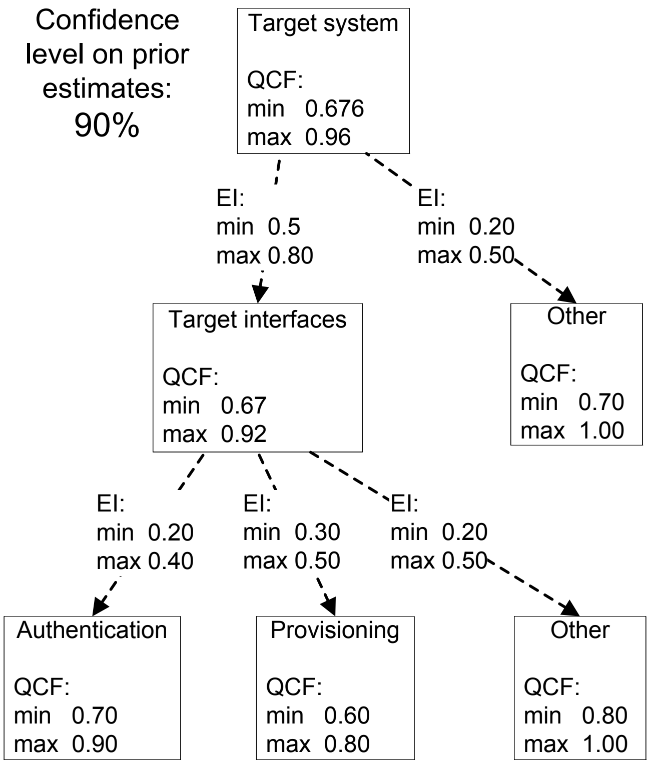

Figure 7. Excerpt of a DV with intervals and confidence level

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

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 nodes and the arcs pointing to them) is a feasible optimization problem [14], [11]. 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:

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

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

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 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 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:

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

In the context of predicting the quality characteristic, 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. Therefore, the uncertainty of a dependee 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. The obtained value of sensitivity (defined by Eq. 7) can in such a case be considered in relation to the effort needed for acquisition of the additional input. That is, higher sensitivity justifies putting more effort in acquiring additional input in order to decrease uncertainty of the dependee (and thus dependent) node.

C. The uncertainty propagation in practice

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

- all QCF intervals involved,
- all EI intervals involved, and

- $\Sigma_{i=1}^{I} ei_i = 1$ (where i denotes an arc, I is the total number of the arcs pointing to the nodes with the common 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 in Figure 7 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,[],[],[],[],[],[],@confuneq,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 in Figure 7. We see that the scalar points of the optimization output are in accordance with the Eq. 5 and Eq. 6.

D. Uncertainty analysis

Statistical analysis of measurements performed prior to model fitting and sensitivity analysis performed in relation to the application of prediction models, require a toolset for analysis of the data sets represented by intervals. The analysis of the central tendency measures of the interval-based estimates relies on the existing fully defined interval arithmetics and interval statistics [15]. Both can, in their existing well-established form, be directly applied in our context.

Apart from the summation and the multiplication presented by Eq. 4, the elementary interval arithmetic functions addition and multiplication (given two intervals denoted by x and y, both of the form given by Eq. 2) include subtraction and division:

$$x - y = [\underline{x} - \underline{y}, \overline{x} - \overline{y}] \tag{8}$$

$$x \div y = [\underline{x}, \overline{x}] \cdot [1/\overline{y}, 1/\underline{y}], \text{ as long as } 0 \notin y. \tag{9}$$

Arithmetic mean is given by:

$$\left[\tfrac{1}{I}\Sigma_{i=1}^{I}\underline{x}_i, \tfrac{1}{I}\Sigma_{i=1}^{I}\overline{x}_i\right]. \tag{10}$$

For geometric mean, harmonic mean, weighted mean, and median, see [15]. Since no two data values are likely to be the same at infinite precision, mode does not generalize to a useful summary for data sets containing interval values. Instead, [15] proposes a substitute statistic, which identifies the places where most values in the data set overlap.

For problems with large sample sizes, computing variance of the interval data is an NP-hard problem. The algorithms for calculating variance presented in [15] solve the issue of infeasibility and make practical calculations of the needed interval statistics.

The standard deviation σ of an interval can be computed immediately from the variance *var* by taking its square root:

$$\sigma = [\underline{\sigma}, \overline{\sigma}] = \sqrt{var} = \left[\sqrt{\underline{var}}, \sqrt{\overline{var}}\right]. \tag{11}$$

Interval statistics for interquartile range, skewness, confidence intervals, regression fitting, maximum likelihood methods, as well as inferential interval statistics are thoroughly presented in [15]. In addition, [15] 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.

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. Both interval width and confidence level 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 judgements are provided 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 [16]:

$$\mu \pm t(1-conf, n-1)\sigma\sqrt{\frac{1}{n}+1} \tag{12}$$

where $t(1-conf, n-1)$ is the two-tailed value of the Student's t-distribution for the confidence level $1-conf$ 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 [2]), 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 will lie within the interval [0;1]. In addition, the normalized quality characteristic metric is defined so that all possible values always must lie within this interval. Moreover, the propagation algorithm calculates both the upper and the lower extreme values. As a result, the inferred prediction is an interval within which the exact (factual) value should lie. Two aspects are hindering from guaranteeing that the factual value lies within the inferred interval:

	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

1) the confidence level with which the prior estimates are provided, and
2) 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 tested by providing example values of estimates and their uncertainty on a real DV structure. The DV structure was originally used in a feasibility study of the PREDIQT method [2], performed on an extensive, real system. The uncertainty estimates were straight-forward to provide by referring to the definition of the rating of the quality characteristic and expressing the estimates in terms of intervals. The interval width was mostly subject to observability of the parameter and existence of relevant historical input. The DV consisted of 38 leaf nodes, 9 internal nodes and 1 root node. The number of EIs on the arcs was 47. Thus, the number of initial (empirical input-based) estimates was 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. Each column lists the number of elements, the maximum interval width, the minimum interval width, the average interval width and the standard deviation of the interval width. The first two columns present the values for the initial estimates of the leaf node QCFs and all the EIs, respectively. The third column presents the values for the initial estimates of both the leaf node QCFs and all the EIs. The last column presents the results for 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 expected uncertainty, all values assigned to parameter estimates 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 interval width of the root node QCF became 0.0295 and the value of Eq. 7 became 0.081. Note that these values, too, are based on fictitious input, due to confidentiality of the actual initial estimates.

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 uncertainties.

Reducing the confidence level and conducting further model fitting (through additional input) are the obvious counter-measures when the inferred values are too uncertain. The candidate parameters for reduction of uncertainty can be identified by using 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 in order to observe 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 [15]. 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 [15]. In addition, [15] 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.

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 apply for the entire model (in order for the uncertainty propagation to be meaningful) rather than being adjustable at the parameter level. As a result, the schema of the increments may be either too coarse grained or too fine grained in the context of certain parameters. The variation of uncertainty between parameters would not be supported, thus violating criterion 2 from Section II.

The Dempster-Shafer structures [15] 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 [17]. 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) [18], [19] may represent 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 in their general form are however demanding to parameterize and interpret the parameters of, which violates our first criterion. This issue has been addressed by [20] where an analytical method for transforming the DVs to Bayesian networks is presented. It also shows that DVs, although easier to relate to in practice, are compatible with BNs. It is possible to generalize this transformation so that our interval-based approach is transformed to a BN before a further BN-based analysis may be conducted. Such an extension would introduce several states on the BN nodes, and assign probabilities to each of them. In that manner, the extension would resemble the Dempster-Shafer structures. BNs in their general form 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 modeling complex systems using higher levels of abstraction originating from the analyst's knowledge and experience [21]. 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 [22]. Using the fuzzy membership functions, a parameter in a model can be represented as a crisp number, a crisp interval, a fuzzy number or a fuzzy interval. In the fuzzy logic approach the algebraic operations are easy and straightforward, as argued and elaborated by [23]. The interval-based approach is a special case of the fuzzy approach, where only the crisp intervals are used as membership functions. 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 [24] is a framework for 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 judgements, but how the measurement-based input can be related to the concepts of the subjective logic, needs to be defined. Thus, applying the subjective logic in the PREDIQT context would increase the fulfillment of our second criterion beyond the needs, while degrading fulfillment of the third criterion.

Uncertainty representation in software development effort-estimation [25], [26] 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 the 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 [27], 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 [28] 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 our first criterion.

The epistemic uncertainty is the crucial one in the context of PREDIQT and therefore given the main 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 makes them less suited in the PREDIQT context.

The ISO approach to handling measurement uncertainty [29] uses a probabilistic representation with normal distribution, and treats both aleatory and epistemic uncertainty equally. Such an approach however does not explicitly account for 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 [30]. It concentrates on stochastic simulations as input for the interval estimates when significant uncertainties exist. Moreover, [15] 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 the context of PREDIQT increase fulfillment of our success 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 [31]. 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 parameterized 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. 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 [32], electricity demand forecasting [33], correlations between wind power and meteorological conditions [34], power system planning [35] and supply industry [36] 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. Lessons learned

This section provides practical guidelines for obtaining the empirical input and reducing the uncertainty of estimates. Firstly, we elaborate on how the maximum acceptable uncertainty objective, that is, an acceptable threshold for uncertainty, can be characterized. Secondly, guidelines for obtaining the prior estimates are summarized. Lastly, means and measures for reducing uncertainty are outlined. The guidelines are based on the authors' experiences from industrial trials of PREDIQT on real-life systems [2], [3]. As such, the guidelines are not exhaustive but may serve as an aid towards a more structured process for uncertainty handling.

A. Characterizing the maximum acceptable uncertainty objective

The maximum acceptable uncertainty objective can to a certain degree be expressed through the confidence level, which is a measure of the expected probability that the correct value lies within the interval assigned to a prior estimate. However, the confidence level is merely concerned with the prior estimates although it indirectly influences the inferred DV estimates. Therefore, if the interval width of a specific non-leaf node is of major concern, it has to be specified directly as a part of the maximum acceptable uncertainty objective, by the stakeholders. Note however that there is still a correlation between the confidence level of the prior estimates and the inferred QCFs, that is, the uncertainty of an inferred QCF is expressed through both width of its interval, as well as the confidence level of the prior estimates which influence the QCF value of the non-leaf node in question.

Consequently, in the case of the prior estimates, the maximum acceptable uncertainty objective can be expressed through the confidence level, and will in that case give interval widths depending on the quality of the empirical input. In the case of the non-leaf node QCF values, the maximum acceptable uncertainty objective should be expressed in terms of both the confidence level of the prior estimates and the interval width of the parameters in question.

B. Obtaining the prior estimates

We recommend obtaining the leaf node QCFs of a subtree prior to obtaining the related EIs. The rationale for this is to fully understand the semantics of the nodes, through reasoning about their QCFs first. In estimating a QCF, two steps have to be undergone:

1) interpretation of the node in question – its contents, scope, rationale and relationship with the Design Models, and
2) identification of the relevant metrics from the Quality Model of the quality characteristic that the DV is addressing, as well as evaluation of the metrics identified.

QCF is the degree of fulfillment of a quality characteristic, with respect to the node in question. Normalization of the values of the above mentioned metrics and their degree of influence, results in a QCF value with an uncertainty interval assigned with respect to the pre-defined confidence level. Alternatively, rating of the characteristic (as formally defined by its Quality Model at the root node level) can be estimated directly with respect to the node under consideration, in order to provide its QCF value.

In estimating an EI, two steps have to be undergone:

1) interpretation of the two nodes in question, and
2) determination of the degree of impact of the child node on the parent node, with respect to the quality characteristic (defined by the Quality Model) that the DV is addressing. The value is assigned relative to the overall EIs related to the same parent node, and with a consistent unit of measure, prior to being normalized (in order to fulfill Eq. 2). The normalized EIs on the arcs from the same parent node have to fulfill Eq. 3, due to the requirement of model completeness.

Hence, EI is the dependency of the parent node on the child node. Estimation of the EI values between a parent node and its immediate children, results in intervals with respect to the pre-defined confidence level.

1) Questions to ask domain experts: The first step in the interaction between the analyst and the domain experts is to clarify the meaning of the node(s) under consideration, their respective rationales and the possible traces to the Design Models. Secondly, the analyst has to facilitate the estimation by reminding the domain experts of the quality characteristic definition – both the qualitative and the formal part of it.

When estimating a QCF the following question is posed: *"To what degree is the quality characteristic fulfilled, given* *the contents and the scope of the node?"* The definition of the quality characteristic (interpretation and the metric) should be recalled.

When estimating an EI the following question is posed: *"To what degree does the child node impact the parent node, or how dependent is the parent node on child node, with respect to the quality characteristic that the DV is dedicated to?"* The definition of the quality characteristic provided by its Quality Model, should be recalled and the estimate is provided relative to the impact of the overall children nodes of this parent. Alternatively, an impact value is assigned using the same unit of measure on all arcs of the sub-tree, and normalized thereafter.

Once one of the above specified questions is posed, depending on the kind of the DV parameter, the domain expert panel is asked to provide the estimate with an interval so that the correct value is within the interval with a probability given by the confidence level. For EIs on the nodes having a common parent, it has to be validated that Eq. 3 is fulfilled.

Furthermore, discussions among the domain experts should be encouraged and all the estimates should be requested during a limited period of time (in the form of tightly scheduled meetings), in order to ensure relative consistency of the estimates. Additionally, for the purpose of the relative consistency of the estimates, the domain expert group should be diverse and representative. There should be continuity in a fraction of the group, and limited turnover between the different meetings. The turnover may however be advantageous for the purpose of the expertise at the different stages of the process of PREDIQT.

Apart from the domain expert judgements, the estimates are also based on measurements. When obtaining measurement-based input, we rely on a measurement plan which relates the practical measurements to the DV parameters and the quality notions. The Goal/Question/Metric [37], [38], [39] approach and the ISO 9126 product quality standard [5] are particularly useful for deducing such relationships. The overall literature on software measurement is extensive [40], [41], [42], [43] and provides useful guidelines for obtaining the measurement-based input.

2) Use of Quality Models: Quality Models are used as a reference in estimation of each prior estimate. The Quality Models assist the domain experts in selecting and evaluating the relevant metrics. The metrics also provide a basis for defining the measurements. The decomposition of the Quality Models is however only based on indicators whose overlaps and degrees of impact on the characteristic may vary. The composition of the degrees of relevance of the various indicators is therefore left to the analyst or the domain experts to determine in the case of each estimate.

3) Use of Design Models: Design Models specify the target of the analysis in terms of scope and the contents. The Design Models serve as a reference for common understanding of the system, prior to and during the estimation. In addition, the appropriate parts of the DVs are traced to the elements of the Design Models, making the contents and the scope of the DV elements more explicit.

4) Determining the uncertainty value: The uncertainty value of a prior estimate is determined through the interval width based on the pre-defined confidence level. In the case of the measurement-based input, the transformation to an interval is presented in Section IV. In that context, confidence level will reflect the data quality (that is, the validity of the measurements).

In the case of the domain expert judgements, however, the interval width is agreed upon by the domain expert panel, while the validity of the panel (that is, mainly representativeness and statistical significance of its composition) is reflected through the confidence level. This ensures consistency of the confidence level in the case of the expert judgements.

In order to also ensure a consistent confidence level in the case of the measurements (where data quality may vary among the measurements related to the different DV estimates), the confidence level can be kept consistent by compensating for the possible variations through the interval width. The relationship between the confidence level and the interval width is however not formalized beyond the fact that the confidence level denotes the probability of the correct value of the estimate lying within the interval.

C. Reducing uncertainty

Since we only consider the epistemic uncertainty, there exist means and measures that can be used to reduce it. The difficulty of reducing uncertainty lies in addressing the unknown sources of uncertainty, which are not explicitly expressed in the estimates. This is however not a major issue in the case of the epistemic uncertainty.

The rest of this section provides guidelines for uncertainty reduction from the different perspectives: process, model granularity, measurement-based input and expert judgements.

1) Process related measures: Among the process related measures are:

- access to the necessary documentation
- access to measurement facilities
- involvement and composition of the domain expert panel in all phases of the process
- common understanding of the modeling language and the terminology
- sufficient understanding of the PREDIQT method, particularly the models (by the domain experts and the analyst)
- use of known notations and modeling frameworks
- use of standards where appropriate
- user-friendly tool support with structured process guidance
- reuse of the existing models where appropriate.

The rationale for these measures is a more structured process which provides the sufficient input and leads towards a harmonized understanding of the models. For a more detailed elaboration of the process related measures, see [3].

2) Granularity of the models: Quality of a model-based prediction is, once the prediction models are developed, subject to the granularity of the models. Increased granularity of all prediction models will potentially decrease uncertainty.

In case of Quality Models, finer granularity can be achieved by further formalization and decomposition of the quality characteristics. In case of Design Models, the more detailed diagrams and traces among them are a means of addressing granularity.

In the case of the DVs, additional traceability of the actions and rationale, as well as increased traceability between DV model elements and the Design Models, will increase the precision and reduce uncertainty. Particularly, the following should be documented during the DV development:

- assumptions
- rationale
- relationships or traces to the Design Models
- traces to the relevant quality indicators and contributions of the relevant quality indicators
- interpretations of the prior estimates
- the supporting information sources (documents, measurement, domain experts) used during the development of DV structure and estimation of the parameters.

3) Quality of measurement data: Increase of validity of the measurement data will directly increase the confidence level. This may be achieved by increasing the statistical significance of the measurements in terms of relevance and amount of the measurement-based input.

4) Quality of expert judgements: The expert judgements are subject to understandability and granularity of the prediction models, composition of the expert panel (representativeness, number of participants, their background and interests), setup and approach to the estimate acquisition. Discussion should be facilitated and possible interest conflicts should be addressed.

VII. Conclusion and Future Work

Our earlier research indicates the feasibility of the PREDIQT method for model-based prediction of impacts of architectural 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 parameterized in terms of fulfillment of quality characteristics and impacts among the elements, with respect to the quality characteristics. The DV elements are representations of architectural design or quality, which are partially traceable to the underlying Design Models and Quality Models. 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.

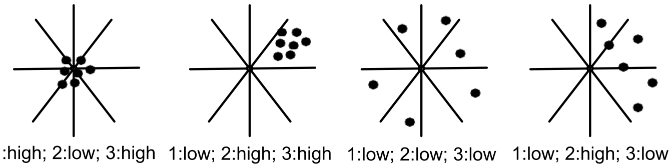

Figure 8. 1: Accuracy; 2: Bias; 3: Precision

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 DV parameters and their respective uncertainty estimates. 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. The interval arithmetics, the 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 evaluation argues for the correctness and practical usefulness of our approach, as well as its outranging appropriateness relative to the alternative uncertainty handling approaches.

The approach is entirely compliant with the existing version of the PREDIQT method. Based on empirical trials of PREDIQT, we have provided guidelines for use of the uncertainty handling approach in practice. The guidelines address the ways of obtaining the empirical estimates as well as the means and measures for reducing uncertainty of the estimates.

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 [2], balanced relative error BRE [44] 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 8) 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 empirical evaluations of practical usefulness 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 as a part of the DIGIT (180052/S10) project funded by the Research Council of Norway, as well as a part of the NESSoS network of excellence funded by the European Commission within the 7th Framework Programme.

REFERENCES

[1] A. Omerovic and K. Stølen, "Interval-Based Uncertainty Handling in Model-Based Prediction of System Quality," in *Proceedings of Second International Conference on Advances in System Simulation, SIMUL 2010*, August 2010, pp. 99–108.

[2] 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.

[3] A. Omerovic, B. Solhaug, and K. Stølen, "Evaluation of Experiences from Applying the PREDIQT Method in an Industrial Case Study," SINTEF, Tech. Rep. A17562, 2011.

[4] "Excel Help and How-to," accessed: June 7, 2011. [Online]. Available: http://office.microsoft.com/en-us/excel-help

[5] International Organisation for Standardisation, "ISO/IEC 9126 - Software engineering – Product quality," 2004.

[6] J. Rumbaugh, I. Jacobson, and G. Booch, *Unified Modeling Language Reference Manual*. Pearson Higher Education, 2004.

[7] 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.

[8] A. D. Kiureghiana and O. Ditlevsenb, "Aleatory or epistemic? Does it matter?" *Structural Safety*, vol. 31, no. 2, pp. 105–112, 2009.

[9] 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.

[10] A. Omerovic, A. Karahasanovic, and K. Stølen, "Uncertainty Handling in Weighted Dependency Trees – A Systematic Literature Review," in *Dependability and Computer Engineering: Concepts for Software-Intensive Systems*, L. Petre, K. Sere, and E. Troubitsyna, Eds. IGI, 2011, accepted to appear as a chapter in the book.

[11] R. B. Kearfott, "Interval Computations – Introduction, Uses, and Resources," *Euromath Bull*, vol. 2, pp. 95–112, 1996.

[12] 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.

[13] J. Nocedal and S. J. Wright, *Numerical Optimization*. Springer, 1999.

[14] A. R. Ravindran, *Operations Research and Management Science Handbook*. CRC Press, 2008.

[15] 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.

[16] T. Wonnacott and R. Wonnacott, *Introductory Statistics*. Wiley, 1990.

[17] 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.

[18] M. Neil, N. Fenton, and L. Nielsen, "Building Large-Scale Bayesian Networks," *Knowledge Engineering Rev.*, vol. 15, no. 3, pp. 257–284, 2000.

[19] D. Heckerman, A. Mamdani, and W. M. P., "Real-World Applications of Bayesian Networks," *ACM Communications*, vol. 38, no. 3, pp. 24–26, 1995.

[20] 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.

[21] 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.

[22] L. A. Zadeh, "Fuzzy Sets," *Information and Control*, vol. 8, pp. 338–353, 1965.

[23] 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.

[24] A. Jøsang, "Artificial Reasoning with Subjective Logic," in *Proceedings of the 2nd Australian Workshop on Commonsense Reasoning*. Australian Computer Society, 1997.

[25] 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.

[26] 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.

[27] 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.

[28] S. Skelboe, "Computation of Rational Interval Functions ," *BIT Numerical Mathematics*, vol. 14, no. 1, pp. 87–95, 1974.

[29] International Organisation for Standardisation, "Guide to the Expression of Uncertainty in Measurement," 1993.

[30] 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.

[31] 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.

[32] T. N. Palmer, "Predicting Uncertainty in Forecasts of Weather and Climate," *Rep.Prog.Phys.*, vol. 63, pp. 71–116, 2000.

[33] 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.

[34] M. Lange and D. Heinemann, "Accuracy of Short Term Wind Power Predictions Depending on Meteorological Conditions," in *Proceedings of Global Windpower Conference*, 2002.

[35] 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.

[36] 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.

[37] V. R. Basili, "Software Modeling and Measurement: the Goal/Question/Metric Paradigm," University of Maryland, Tech. Rep. TR-92-96, 1992.

[38] V. Basili, G. Caldiera, and H. Rombach, "The Goal Question Metric Approach," *Encyclopedia of Software Engineering*, 1994.

[39] N. E. Fenton and S. L. Pfleeger, *Software Metrics: A Rigorous and Practical Approach*. PWS Publishing Co., 1998.

[40] C. Ebert, R. Dumke, M. Bundschuh, A. Schmietendorf, and R. Dumke, *Best Practices in Software Measurement*. Springer Verlag, 2004.

[41] C. Ebert and R. Dumke, *Software Measurement*, 1 ed. Springer-Verlag Berlin Heidelberg, 2007.

[42] S. H. Kan, *Metrics and Models in Software Quality Engineering*, 1st ed. Addison-Wesley Longman Publishing Co., Inc., 1994.

[43] J. C. McDavid and L. R. L. Hawthorn, *Program Evaluation and Performance Measurement : An Introduction to Practice*. Sage Publications, 2006.

[44] J. Armstrong, *Long-Range Forecasting*. Wiley, 1985.