Dependability and Computer Engineering:

Concepts for Software-Intensive Systems

Luigia Petre
Åbo Akademi University, Finland

Kaisa Sere
Åbo Akademi University, Finland

Elena Troubitsyna
Åbo Akademi University, Finland

ENGINEERING
SCIENCE REFERENCE

Senior Editorial Director: Kristin Klinger
Director of Book Publications: Julia Mosemann
Editorial Director: Lindsay Johnston
Acquisitions Editor: Erika Carter
Development Editor: Michael Killian
Production Editor: Sean Woznicki
Typesetters: Keith Glazewski, Natalie Pronio, Jennifer Romanchak
Print Coordinator: Jamie Snavely
Cover Design: Nick Newcomer

Published in the United States of America by
Engineering Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue
Hershey PA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: http://www.igi-global.com

Library of Congress Cataloging-in-Publication Data

Dependability and computer engineering: concepts for software-intensive systems / Luigia Petre, Kaisa Sere and Elena Troubitsyna, editors.
p. cm.
Summary: “This book offers a state-of-the-art overview of the dependability research, from engineering various software-intensive systems to validating existing IT-frameworks and solving generic and particular problems related to the dependable use of IT in our society”--Provided by publisher.
Includes bibliographical references and index.
ISBN 978-1-60960-747-0 (hardcover) -- ISBN 978-1-60960-748-7 (ebook) -- ISBN 978-1-60960-749-4 (print & perpetual access) 1. Reliability (Engineering) 2. Computer systems--Reliability. 3. Computer engineering. I. Petre, Luigia, 1974- editor. II. Sere, K. (Kaisa), 1954- editor. III. Troubitsyna, Elena, 1970-, editor.
TS173.D47 2011
620’.00452--dc22
2011011401

British Cataloguing in Publication Data
A Cataloguing in Publication record for this book is available from the British Library.

Chapter 16
Uncertainty Handling in Weighted Dependency Trees: A Systematic Literature Review

Aida Omerovic
SINTEF & University of Oslo, Norway

Amela Karahasanovic
SINTEF & University of Oslo, Norway

Ketil Stølen
SINTEF & University of Oslo, Norway

ABSTRACT

Weighted dependency trees (WDTs) are used in a multitude of approaches to system analysis, such as fault tree analysis or event tree analysis. In fact, any acyclic graph can be transformed to a WDT. Important decisions are often based on WDT analysis. Common for all WDT-based approaches is the inherent uncertainty due to lack or inaccuracy of the input data. In order to indicate credibility of such WDT analysis, uncertainty handling is essential. There is however, to our knowledge, no comprehensive evaluation of the uncertainty handling approaches in the context of the WDTs. This chapter aims to rectify this. We concentrate on approaches applicable for epistemic uncertainty related to empirical input. The existing and the potentially useful approaches are identified through a systematic literature review. The approaches are then outlined and evaluated at a high-level, before a restricted set undergoes a more detailed evaluation based on a set of pre-defined evaluation criteria. We argue that the epistemic uncertainty is better suited for possibilistic uncertainty representations than the probabilistic ones. The results indicate that precision, expressiveness, predictive accuracy, scalability on real-life systems, and comprehensibility are among the properties which differentiate the approaches. The selection of a preferred approach should depend on the degree of need for certain properties relative to others, given the context. The right trade off is particularly important when the input is based on both expert judgments and measurements. The chapter may serve as a roadmap for examining the uncertainty handling approaches, or as a resource for identifying the adequate one.

DOI: 10.4018/978-1-60960-747-0.ch016

INTRODUCTION

WDTs are widely used in approaches to system analysis. WDTs are used as a means to understand some artifact, e.g. a system and to make informed decisions regarding its future behaviour. Examples include:

- fault tree analysis (IEC, 2006) – a risk analysis technique based on so-called fault trees;
- event tree analysis (IEC, 1995) – a modeling technique for representing consequences of an event and the probabilities of the respective consequences;
- attack trees (Schneier, 1999) – a notation similar to fault tree analysis for modeling potential attacks on a system with an attack goal as the top node and different ways of achieving that goal as leaf nodes; and
- dependency views in system quality prediction (Omerovic, et al., 2010).

Common for all approaches supported by WDTs is the inherent uncertainty due to lack or inaccuracy of input data. The input data originates from two kinds of sources: expert-judgments-based and measurement-based data acquisition (such as logs, monitoring, experience factories and other measurements). Uncertainty regarding input data (due to for example lack of knowledge, as well as variability or poor quality of the measurement-based data) can lead to errors in relation to both modeling and analysis.

Apart from lack or inaccuracy of input, another source of the uncertainty may be the variability of the system or its usage. However, in this chapter we restrict our attention to artifacts, e.g. systems, whose behavior is deterministic, for which the former type of uncertainty is the only prevailing one. Consequently, we only focus on deterministic uncertainty handling.

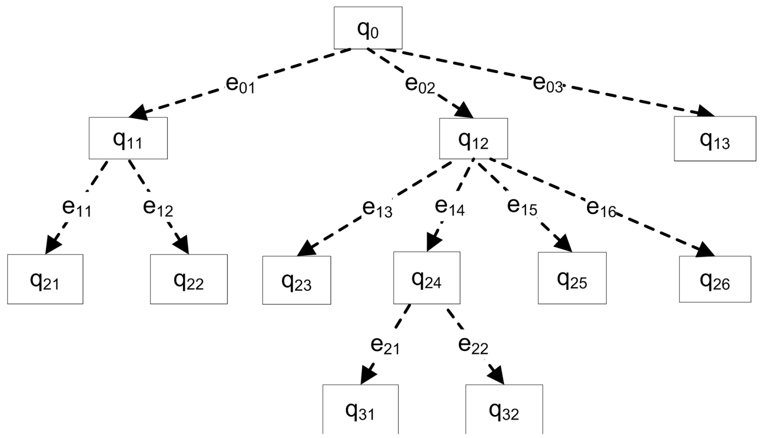
Important decisions are made during and after a WDT-based analysis. The ability of explicitly expressing uncertainty in the input and its implications on the rest of the model is crucial due to the consequences of the input uncertainty in the decision making. The uncertain input may have extensive impact on the rest of the WDT-based model. In worst case, decisions involving unacceptable risk may be taken, without the awareness of the analyst. There are numerous approaches to uncertainty handling. Each of them is motivated by a special need not covered by the competing ones. Their properties such as complexity, expressiveness, propagation and generally the practical applicability vary to a high extent. It is however unclear to what degree the existing approaches are suitable for being adopted for use on WDTs. We have not been able to find a comprehensive evaluation of the uncertainty handling approaches in the context of the WDTs.

The overall objective of this chapter is therefore to identify and evaluate the existing and the potentially useful approaches that could, individually or in combination, be adopted for handling the uncertainty in WDTs. This includes:

1. identification of the established approaches for uncertainty representation, based on a literature review;
2. a high-level evaluation of the identified approaches with respect to the main needs; and
3. a low-level evaluation of a restricted set of the approaches which are assessed as potentially useful during the high-level evaluation.

The chapter is structured as follows: "*Weighted dependency trees and uncertainty*" provides background on the notion of WDT and the issue of uncertainty. The research method is then presented in a section titled "*The research method*". Next, a section titled "*The evaluation criteria*" presents the evaluation criteria, with respect to the practical acceptability of the uncertainty handling approaches. The section titled "*The high-level evaluation*" reports on the high-level

Figure 1. A weighted dependency tree



evaluation of the approaches identified, while the section titled "*The low-level evaluation*" presents the results of a more detailed evaluation of the selected approaches. Finally, we discuss our findings and provide our main conclusions in the section titled "*Conclusion*". Furthermore, there are three appendices providing background on the literature review, the detailed deduction of the evaluation criteria, and threats to validity and reliability, respectively.

WEIGHTED DEPENDENCY TREES AND UNCERTAINTY

This section introduces the two central notions: dependency trees and uncertainty. Firstly, a general definition of a WDT is proposed. Secondly, the major sources of and the two overall categories of uncertainty, are introduced.

Weighted Dependency Trees

Various tree structures, such as for example event trees, attack trees or dependency trees, are used in system analysis. We generalize the notion of the tree-based representations with quantitative parameters such as the already mentioned fault trees, event trees, attack trees or dependency views, by proposing the concept of a WDT. A WDT, as illustrated by Figure 1, is a finite tree in which:

- each node may hold one or more quantitative values or a function,
- each arc may hold one or more a quantitative values or a function, and
- each inferred non-leaf node is a tree-specific function of its immediate children and their connecting arcs.

The tree-specific function is basically what distinguishes one kind of WDT from another. For example, in case of the WDT on Figure 1, the value q_{11} is a function of q_{21}, q_{22}, e_{11} and e_{12}. Note that not all tree-specific functions require assignments on the arcs. Furthermore, we distinguish between prior estimates and the inferred or propagated estimates. The former ones are based on the empirical input which may originate from expert judgments or measurements, while the latter one is inferred based on tree-specific functions.

We distinguish between two variations of WDTs:

- Restricted WDT (RWDT): the same node may occur only once, and
- General WDT (GWDT): the same node may occur more than once.

Any directed acyclic graph (DAG) may be represented as GWDT, by duplicating the nodes in the DAG.

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) (Kiureghian & Ditlevsen, 2009).

The aleatory uncertainty is a property of the system associated with variability. It 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 a discrete nature. It is considered as uncertainty which may be originating from a range of causes that defy pure probabilistic modeling. Examples of such causes are lack or inaccuracy of input, which impedes the specification of a unique probabilistic model. Epistemic quantities have fixed values which may be unknown or poorly known. For example size or cost of an existing system are values which are fixed and existing, but may be difficult to reveal or deduce.

As opposed 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), models of deterministic artifacts are characterized by rather discrete, sudden, non-stochastic and less frequent changes. As a result, the aleatory uncertainty is, in the deterministic artifacts, 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 deterministic artifacts.

THE RESEARCH METHOD

This section provides an overview of the research method. The overall process of research is outlined in the next subsection, which is followed by a subsection presenting the main steps within the literature review.

Overall Process

The overall process is divided into the two stages depicted in Figure 2. Firstly, a literature review is conducted. The literature review consists of a mapping study (that is, a search in several digital resources, based on pre-defined keywords) and a systematic literature review. The relevant publications related to the state of the art with respect to the uncertainty handling approaches, are identified. The main contents of the relevant publications are extracted, indicating the potentially useful approaches for uncertainty handling and their main properties with respect to the objective specified. The systematic literature review provides a preliminary evaluation and an initial list of the evaluation criteria. Based on the results of the systematic literature review and the experienced needs for the evaluation, the evaluation criteria are iteratively refined through multiple discussions among the researchers. The deduction of the criteria is partially guided by an existing taxonomy for system acceptability.

Secondly, an evaluation of the approaches identified and analyzed throughout the systematic literature review is conducted. The evaluation consists of two stages: a high-level evaluation and a low-level evaluation, respectively. The

Figure 2. An overview of the process undergone

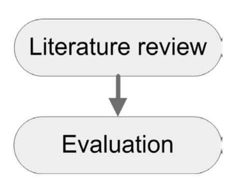


Table 1. The objective of the systematic literature review

Objective:	Identify and assess the potentially useful approaches which satisfy the sub-objectives 1, 2 and 3.
Sub-objective 1:	The approach is practically applicable for supporting uncertainty handling in the context of WDT-based analysis.
Sub-objective 2:	The approach handles the epistemic uncertainty.
Sub-objective 3:	The approach has a well defined propagation within the WDT.

high-level evaluation classifies the approaches identified and summarizes their main properties. The classification is organized according to the kind of the approach (e.g., bayesian, fuzzy, interval-based, hybrid, etc.). Moreover, each classified kind of the approach is assessed as to whether it should be further evaluated as a part of low-level evaluation, that is in the context of the WDTs and with respect to the refined evaluation criteria. In the low-level evaluation each approach is instantiated on a WDT, in order to test its applicability.

Literature Review

This subsection presents an overview of the literature review. It consists of a mapping study performed in order to gain an overview of the main topics and possible keywords, and a systematic literature review. The details regarding the process of both the mapping study and the systematic literature review are provided in Appendix A.

The Mapping Study

The mapping study involved a search in several digital resources, based on pre-defined keywords related to uncertainty and prediction. About fifty publications from mathematics, statistics and computer science which presented either single or combined approaches for uncertainty handling, were extracted. The ones addressing customization of identified approaches for the different domains and the ones focusing on mainly aleatory uncertainty, were excluded, thus reducing the search results to about 33 publications. Moreover, uncertainty and inaccuracy handling in other domains, were searched for in the same resources. About ten additional publications were recorded. A walkthrough of the search results was made, providing additional five publications. Thus, the mapping study resulted in 38 articles addressing the uncertainty handling approaches relevant in the context of WDT-based analysis of deterministic systems.

The Systematic Literature Review

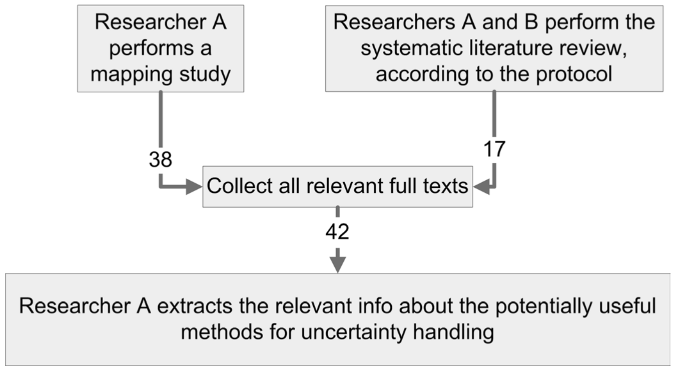
For the systematic literature review we followed a procedure recommended by (Kitchenham, et al., 2007). The first phase of planning a systematic literature review is to specify the research objective that should be satisfied by conducting the review. The objective of this systematic literature review is stated in Table 1.

By an approach we mean a collection of formal and graphical means available, along with the procedures prescribed for applying those means.

The second step of planning a systematic literature review is to develop the search strategy in the form of a review protocol. As the search captures a wide range of publications, we included a stopping heuristic.

The overall process is sketched in Figure 3. Since the search engines used sort the results by relevance, this procedure is effective in producing search results. In most cases, the density of relevant titles steadily decreased as we examined each list of the search results. The systematic literature review resulted in 17 relevant publications. The total number of the unique publications evaluated as relevant during both the mapping study and the systematic literature review was 42.

Figure 3. An overview of the process applied in the literature review



Thus, there were 13 overlapping articles between the mapping study and the systematic literature review, and four of the overall 29 were found through the systematic literature review, while 25 were found through the mapping study. 19 of the 25 relevant articles discovered only through the mapping study were published before 2000 – the timeframe which was excluded from the searches during the systematic literature review.

THE EVALUATION CRITERIA

Based on the objectives of our research, the results of the systematic literature review and the evaluation framework in (Nielsen, 1993), we have identified and iteratively refined a list of evaluation criteria. Table 2 provides an overview of the evaluation criteria. Each criterion is specified in detail in the Tables 3 through 9. Rationale behind the selected criteria is given in Appendix B.

THE HIGH-LEVEL EVALUATION

The results of our literature review show the following trends in terms of publication years and the kinds of publications for the respective approaches. The publications found regarding the imprecise probabilities date from 1991 to 2005 and consist of a book and two scientific chapters. The publications found regarding Bayesian networks date from 1995, and consist of scientific chapters. The publication referenced on the subjective logic is from 1997. Most of the publications found regarding Credal networks are from 2003

Table 2. An overview of the evaluation criteria

Criterion number	Criterion
C1	Complexity of the representation form for the quantitative estimates.
C2	Suitability of the approach for both expert-judgment-based and measurement-based input.
C3	Scalability with respect to number of estimates needed on a WDT.
C4	Precision allowed by the representation form for the quantitative estimates.
C5	Complexity of the propagation.
C6	Expressiveness of the arithmetic operations for propagation.
C7	Empirical evaluation of the approach, as reported in the publications found through the literature review.

Table 3. Specification of Criterion C1

Category number / Category	Criterion C1: Complexity of the representation form for the quantitative estimates.
C1_1	Simple and intuitive. Correspondence with the empirical input without need for a transformation of the notions from the input to the notions of the approach.
C1_2	A schema defining the correspondence of the notions used for the empirical input, is needed and relatively straight-forward to define.
C1_3	A schema defining the correspondence of the notions used for the empirical input, is needed but demanding to define.

to 2007. The fuzzy logic relevant publications date from 1965 to 2008. The Bayesian networks relevant publications date from 1996 to 2009. The interval-based approaches have been addressed from 1994 to 2008 in several technical reports and scientific publications uncovered during the literature review.

This section reports on the results from the high-level evaluation of the uncertainty handling approaches, including analytical approaches for uncertainty representation, hybrid approaches in software engineering, as well as some of the uncertainty handling approaches in statistics, meteorology and software development effort estimation. The list is not exhaustive, but introduces the prevailing approaches which have been identified through the literature review. We categorize the approaches identified and outline each approach with respect to its properties, references and reported evaluation, all extracted from the publications which have been found through the literature review. Then we elaborate on whether the approach can be useful in the context of the WDTs.

Bayesian Networks

Bayesian networks (Neil, Fenton, & Nielson, 2000) (Heckerman, Mamdani, & Wellman, 1995) are quantitative probabilistic networks that allow incorporating both model uncertainty and parameter uncertainty. A Bayesian network (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. In a BN, probability distributions are calculated according to Baye's rule, based on conditional probability tables (CPTs) and the probabilities of the parent nodes. The initial uncertainty is placed in the prior distribution of each input

Table 9. Specification of Criterion C7

Category number	Criterion C7: Empirical evaluation of the approach, as reported in the publications found through the literature review.
C7_1	Promising empirical evaluation reported, in terms of both validity (i.e., size and types of the systems the approach is applied on) and reliability (i.e., the possibility of obtaining the same results in repeated trials) of the evaluation method.
C7_2	The reported empirical evaluation indicates strong validity but weak reliability.
C7_3	The reported empirical evaluation indicates strong reliability but weak validity.
C7_4	Empirical evaluation reported but with considerable limitations with respect to both reliability and validity.
C7_5	No reports on empirical evaluation have been possible to extract from the publications found through the literature review.

parameter. The prior distribution is then updated to a posterior distribution based on the observed data associated with each parameter. Additional data acquisition is undertaken until an acceptable certainty is reached. BNs are however demanding to parametrize and interpret the parameters of, an issue addressed in (Omerovic & Stølen, 2009), where an analytical approach for transforming WDTs to BNs is proposed.

The application of BNs is extensive. Bayesian approach for uncertainty analysis in software reliability modeling is for example applied in (Yin & Trivedi, 1999). Application of the Bayesian networks on large scale systems has been reported in (Fenton & Neil, 1999). The reported empirical evaluation of the Bayesian networks is promising.

The applicability of the Bayesian networks on a WDT will be considered as a part of the low-level evaluation.

Bootstrapping

Bootstrapping may be parametric and nonparametric and is based on resampling methods. The former quantifies the effect of input parameter uncertainty for the parametric formulation. Using the available information, the parameters are first estimated by the maximum likelihood estimation. The estimates are used to draw a new sample of the observations. This process is repeated a number of times obtaining as many estimates as the number of repetitions, to the input parameters. The latter incorporates the error due to input distributions. The nonparametric resampling methods may model the distribution functions of independent input random variables.

The use of a percentile confidence interval is recommended in the bootstrapping approach in the absence of simulation uncertainty. Its use assumes that the statistics of interest are computed deterministically by the resampling methods. When simulation uncertainty is present, percentile confidence intervals are based on a convolution of the input uncertainty and simulation uncertainty. In the bootstrapping simulation method, it is impossible to separate these forms of uncertainty (Batarseh & Wang, 2008).

Bootstrapping is barely presented in the publications found and information on its performance in the empirical studies has not been possible to extract. However, its main disadvantage is the focus on the measurement-based uncertainty, and lack of a representation form suitable for expressing the uncertainty within the expert judgments.

ISO Guide to Handling Measurement Uncertainty

The ISO guide to handling measurement uncertainty (ISO, 2008) uses a probabilistic representation with normal distribution, and treats both aleatory and epistemic uncertainty equally. It also contains instructions on the propagation of the uncertainties. We argue however that such an approach does not explicitly account for the notion of ignorance about the estimates.

As argued in relation to the deduction of the evaluation criteria, discrete representations are better suited than continuous ones to capture epistemic uncertainty.

Delta Method

The Delta method developed by Cheng and Holloand (Cheng & Holloand, 1997) for input uncertainties assumes that the model is known while input parameters are uncertain. The true values of the parameters are estimated by the maximum likelihood estimation, assuming that the parameters follow a normal distribution. The total simulation output variance is estimated by simulation variance and the input parameter variance. No report on the evaluation of the performance of this approach in the empirical studies is found within the literature review.

As argued above, discrete representations are better suited than continuous ones to capture epistemic uncertainty.

Qualitative Probabilistic Networks

Qualitative probabilistic networks are qualitative abstractions of probabilistic networks. A qualitative probabilistic network comprises an acyclic digraph which models variables and the probabilistic relationships between them. Instead of conditional probability distributions, a qualitative probabilistic network associates with its digraph qualitative influences and qualitative synergies (Wellman, 1990).

Qualitative probabilistic networks are barely present in the publications considered by us. We have not been able to extract information on their performance in the empirical studies. A drawback of the qualitative networks is the lack of precision in the representation of the uncertain quantitative parameters. Measurement-based quantities are for example more exact than what can be expressed by the qualitative networks.

Credal Networks

A Credal network (Cozman, 2005) is a graphical model that associates nodes and variables with sets of probability measures. The motivation for Credal networks is the need for more relaxed Bayesian networks. In a Bayesian network, the Markov property implies that we must specify a unique probability distribution for every variable conditional on any configuration of the variable's parents. However, existing beliefs may be incomplete or vague, or there may be no resources to gather or process the input so as to reach a precise probability assessment. The point estimates may not be exact or the domain experts may disagree. Hence, probability intervals may be selected as estimates. As a result, a set of probability distributions is specified for every variable conditional on the variable's parents, thus obtaining a Credal network.

Credal networks are like discrete Bayesian networks, except that they specify closed convex sets of probability mass functions, so-called Credal sets (Levi, 1980), instead of single probability mass functions. Every Credal network variable is associated with a collection of local conditional Credal sets, which do not interfere with each other. (Haenni, 2007) introduces an approximate propagation algorithm for Credal networks. The algorithm represents the Credal network as a compiled logical theory. The resulting graphical structure is the basis on which the subsequent steepest-ascent hill-climbing algorithm operates. The output of the algorithm is an inner approximation of the exact lower and upper posterior probabilities.

Real applications of a complex Credal network constructed both from expert opinions and data, have been reported in (Antonucci, Salvetti, & Zaffalon, 2004). A Credal network constructed from data and used for classification in a medical scenario is presented in (Zaffalon, Wesnes, & Petrini, 2003).

A drawback of the Credal networks is lack of scalability due to the need for providing an extensive number of estimates even for relatively small networks. Another weakness is the interpretation of the probability intervals and their correspondence with the non-probabilistic notions on the WDTs. We consider the quantitative representation of the uncertain estimates on the Credal networks to be too complex in terms of both the number of the estimates needed and their probability-interval-based form, which is non-trivial to interpret by the domain experts. Moreover, the measurement-based input needs to be related to the probability-interval-based representation form.

Intervals

Definitions of intervals, their arithmetics and central tendency operators are provided by (Ferson, Kreinovich, Hajagos, Oberkampf, & Ginzburg, 2007). Additional comprehensive references on the interval-based approach are (Majumdar, et al., 2001), (Kearfott, 1994) and (Kreinovich, et al., 2008). The practical interval solution depends on the monotonicity properties (the relationship between model input and output) of the model. The probability of occurrence of any value within an interval can follow any given arbitrary distribution.

As argued in (Majumdar, et al., 2001), interval arithmetics can serve as a tool to obtain interval extensions of real functions. However, due to a so-called overestimation effect, also known as dependency problem, interval arithmetic does not provide the exact range of a function. 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 occurrences of interval parameters cannot always be avoided, the dependency problem often causes overestimation of the actual range of an evaluated function. A way to handle this issue is to use interval splitting (Majumdar, et al., 2001), where the input intervals are divided and the arithmetics are performed 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. In (Skelboe, 1974) it is shown that the results obtained from the interval splitting converge to the actual range when the width of the subintervals approaches zero. The application of this techniques on performance models is reported in (Majumdar & Ramadoss, 1995).

(Majumdar, et al., 2001) present intervals and extended histograms for characterizing system parameters that are associated with uncertainty and variability. Histograms are expressed by a series of intervals with associated probabilities. Adaptation of the existing approaches for analytic performance evaluation for this interval-based parameter characterization is described. Its application is illustrated by two examples: a hierarchical model of a multicomputer system and a queueing network model of an EJB (Enterprise Java Bean) server implementation.

An approach to assembling, structuring and representing evidence in a decision, based on hierarchical modeling of the process leading up to a decision, is presented in (Davis & Hall, 2003). The propagation is achieved through the evidence hierarchy, using interval probability theory. Case studies within oil and civil engineering industries are used to demonstrate the approach.

The applicability of the interval-based approach on a WDT will be considered as a part of the low-level evaluation.

Subjective Logic

Subjective logic (Jøsang, 1997) is a framework for formal reasoning, which consists of a belief model called *opinion* and a 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 ("$\wedge$"), disjunction ("$\vee$"), negation ("$\neg$"), consensus ("$\oplus$"), recommendation ("$\otimes$") and ordering ("$\uparrow$").

The subjective logic is suited for domain expert judgments, but how the established representations of the automatically acquired input can be related to the concepts of the subjective logic, is unclear. In the context of a WDT, the leaf node estimates could be expressed in terms of opinions and propagation can be based on the algebra of subjective logic. Subjective logic can generally be applied when taking advices from different sources.

The applicability of the subjective logic – based approach on a WDT will be considered as a part of the low-level evaluation.

Fuzzy Logic

Fuzzy logic provides a simple way of drawing definite conclusions from vague, ambiguous or imprecise information, and allows for partial membership in a set. Furthermore, fuzzy logic facilitates the modeling of complex systems using higher levels of abstraction originating from the analyst's knowledge and experience (Weber, 1994). 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 (L.A Zadeh, 1965).

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. The ordinate of a graphical representation of the membership distribution is the membership function, which may be thought of as "level of presumption", while the abscissa is the interval of confidence. Thus, a crisp number is represented by (a,a,0,0) where *a* is a sharp value on the x-axis with maximum level of presumption; a crisp interval is represented by (a,b,0,0) where the interval on the x-axis ranges from *a* to *b*, while the lever of presumption is flat and equal to one for the entire range of values from *a* to *b*; a fuzzy number is represented by (a,a,x,y) where the level of presumption continuously increases from zero to one for the x-axis values starting at *a-x* and ending at *a*, while from *a* to *a+y* values on the x-axis the level of presumption continuously decreases from one to zero. Thus, sketch of the membership function of a fuzzy number forms a triangle. A fuzzy interval is represented by (a,b,x,y) where for the range of values from *a-x* to *a* on the x-axis the level of presumption continuously increases from zero to one, for the range of values from *a* to *b* on the x-axis the level of presumption is one, and for the range of values from *b* to *b+y* on the x-axis the level of presumption continuously decreases from one to zero. In the approach of fuzzy logic, the algebraic operations are easy and straightforward, as argued and elaborated in (Suresh, Babar, & Raj, 1996). Fuzzy logic is non-statistical in nature, but with a well defined arithmetics.

The degrees of membership in a fuzzy set are certain. Therefore, this type of fuzzy sets can not deal, for example, with uncertainty that arises when more than one expert provides judgments. In addition, this type of fuzzy sets can not accommodate different experts' definitions of the low size fuzzy set. As a result, the corresponding fuzzy logic is incapable of dealing with uncertainties of this kind. This is addressed by an extended definition of fuzzy sets which incorporates the notion of uncertainty discussed above, the so-called type-2 fuzzy sets (Ahmed & Muzaffar, 2009). In the type-2 fuzzy sets the membership function is three-dimensional. The first two criteria allow handling imprecision via modeling the degree of membership; while the third criterion allows handling uncertainty via modeling the amplitude distribution of the degree of membership.

Application of fuzzy logic for uncertainty handling in software effort prediction has been reported in (Ahmed, Omolade Saliu, & AlGhamdi, 2005). A graph-based representation with learning algorithms is presented. A validation experiment has been carried out on artificial datasets as well as the COCOMO public database. Additionally, an experimental validation of the training procedure is presented. The authors argue that the fuzzy logic – based framework incorporates expert knowledge, historical data, imprecision handling and adaptability. The uncertain fuzzy estimates are expressed in terms of triangular membership functions.

A case study regarding a type-2 fuzzy system has been conducted on COCOMO ll data for software effort prediction (Ahmed & Muzaffar, 2009). The evaluation is based on root mean square error. Each attribute for prediction is divided into a selected number of fuzzy sets. The evaluation results are promising.

The applicability of the fuzzy logic – based approach on a WDT will be considered as a part of the low-level evaluation.

Dempster-Shafer Structures

The Dempster-Shafer structures (Ferson, et al., 2007) offer a way of representing uncertainty quantified by mass function distributions. A mechanism for aggregation of such representation stored in distributed relational databases, is proposed in (Scotney & McClean, 2003). The Dempster-Shafer approach characterizes uncertainties as intervals with degrees of certainty (that is, sets of values with weights which add up to one). 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.

Dempster-Shafer evidence theory is a possible approach for epistemic uncertainty analysis. It relaxes the assumptions of the probability theory and allows for combining conflicting or ambiguous evidence from multiple sources. In probability theory, however, evidence is associated with only one possible source or event. Dempster-Shafer evidence theory models the epistemic uncertain input variables as sets of intervals. Each variable may be defined by one or more intervals. The user specifies a probability assignment to each interval. The probability assignment indicates how likely it is that the uncertain input falls within the interval. The probability assignments for a particular uncertain input variable must sum to one. The intervals may overlapp, chain or gap relative to one another. Dempster-Shafer has two measures of uncertainty, belief and plausibility. The intervals are propagated to calculate belief and plausibility. Together, belief and plausibility define an interval-valued probability distribution.

Dempster-Shafer interval calculation is computationally expensive due to the number of the interval combinations (within and accross each variable) to be included in the propagation. Minimum and maximum function values are searched for within each interval combination. The minimum and maximum values are aggregated to create the belief and plausibility. The accuracy of the Dempster-Shafer approach depends on the number of samples and the number of interval combinations.

Implementing the Dempster-Shafer theory in the context of WDTs involves 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 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.

Although it is more precise than e.g. the intervals, the main weakness of the Dempster-Shafer structures in the WDT context is the complexity of the representation of the quantitative estimates. This may compromise the comprehensibility of the uncertainty representation and therefore the correctness of the input provided by the domain experts.

Evidence Theory

Both possibility theory, Dempster-Shafer theory and fuzzy sets theory are special cases of the evidence theory. It relies on the notion of random set, each set-valued realization representing an incomplete information item. The evidence theory allows combining evidence from different sources and arriving at a degree of belief (represented by a belief function) that takes into account all the available evidence. In evidence theory, likelihood is assigned to sets, as opposed to probability theory where likelihood is assigned to a probability density function. In the evidence theory, two main functions help us to obtain information about the uncertainty in the knowledge, the plausibility and the belief functions.

The evidence theory of (Shafer, 1976) is a branch of the mathematics of uncertain reasoning that has been presented as an alternative decision theory, as a purely subjectivist, non-statistical approach to uncertain evidence. (Fioretti, 2004) presents the principles of evidence theory and its application in creative decision making. Details on epistemic uncertainty representation using evidence theory can be retrieved from (Simon, Weber, & Levrat, 2007), which argues for shortcomings of the Bayesian networks in representing uncertain knowledge, and proposes an integration of the Bayesian networks model of reliability and the evidence theory, resulting in a so-called Evidential Network.

We consider the evidence theory as too complex for being adopted on the WDTs. The representation form is therefore not necessarily suited for the expert-judgment-based input.

Imprecise Probabilities

In the imprecise probability theory, developed in (Walley, 1991), sets of probability functions capture the notion of partial lack of probabilistic information. The imprecise probabilities approach the probability theory through lower and upper probabilities, rather than probabilities. (Cozman, 2005) presents an overview of graphical models that can handle imprecision in probability values. A review of the algorithms for local computation with imprecise probabilities is presented in (Cano & Moral, 1999), where the objective is to carry out a sound global computation by mainly using the initial local representation. These algorithms try to solve problems of propagation (calculation of conditional or unconditional probabilities) in cases in which there is a large number of variables. There are two main types depending on the nature of the assumed independence relationships. In both of them the global knowledge is composed of several pieces of local information. It has not been possible for us to extract information on performance of imprecise probabilities in the empirical studies.

A drawback of the imprecise probabilities is the interpretation of the probability intervals and their correspondence with the non-probabilistic notions of the WDTs. We consider the quantitative representation of the uncertain estimates on the imprecise probabilities to be too complex. The probability interval – based form is non-trivial to interpret by the domain experts. Consequently, the actual estimates available by the domain experts may be difficult to provide in a precise and correct form. Due to the number of estimates needed (as a consequence of the propagation based on Baye's rule), the imprecise probabilities – based approach also suffers from low scalability.

Hybrid Approaches

The propagation of imprecise probabilities in Bayesian networks is addressed in (Kleiter, 1996), where the imprecision is handled by second order probability distributions. The Dirichlet distributions are used to express the uncertainty about probabilities. The problem of how to propagate point probabilities in a Bayesian network is transformed into the problem of how to propagate Dirichlet distributions in Bayesian networks.

(Renooij & van der Gaag, 2002) present a new type of network in which both signs and numbers are specified (a combination of a qualitative and quantitative network). An associated algorithm for probabilistic propagation is also presented. Building upon these semi-qualitative networks, a probabilistic network can be quantified and studied in a stepwise manner. As a result, modeling inadequacies can be detected and amended at an early stage in the quantification process.

(Ferson, et al., 2007) propose considering a hybrid approach comprising both probabilistic and possibilistic (based on intervals) representation, in order to account for both aleatory and epistemic uncertainty.

(Baudrit, Couso, & Dubois, 2007) propose mathematical models for joint uncertainty propagation involving quantities respectively modeled by probability and possibility distributions in the context of risk analysis.

The Maximum-entropy principle and the Bayesian approach have in a hybrid approach (Dai, Xie, Long, & Ng, 2007) been proposed for uncertainty analysis in software reliability modeling. The approach proposed integrates the capability of the maximum entropy principle the Bayesian approach to derive the priori distribution, which can incorporate historical data with expert judgments, constraints and expected values. This is specifically appropriate for highly reliable software, where only a few failure data are available from a test.

A hybrid Monte Carlo and possibilistic approach for representation and propagation of aleatory and epistemic uncertainty is presented in (Baraldi, Popescu, & Zio, 2008). The approach 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 limited cumulative distributions requires the introduction of a degree of confidence directly connected with the confidence on the value of epistemic parameters.

A simulation mechanism which takes into account both aleatory and epistemic uncertainty in an interval-based approach, is proposed in (Batarseh & Wang, 2008). It concentrates on stochastic simulations as input for the interval estimates, when significant uncertainties exist. Probabilistic distributions represent variabilities and intervals capture epistemic uncertainty. The interval width is proportionally related to the uncertainty. The calculations are based on interval arithmetic, and interval statistics methods are used to report the mean and the standard deviation to provide a concise summary of the results.

(Baraldi, et al., 2008) illustrate a hybrid Monte carlo and possibilistic approach for the representation and propagation of aleatory and epistemic uncertainty. The approach obtained promising results in a case study addressing the lifetime of a component which is randomly degrading in time.

The hybrid approaches are mainly based on stochastic simulation or probabilistic notions based on continuous representations, which are, as argued earlier, not suited for handling the epistemic uncertainty. That is also why the hybrid approaches have not been selected for the low-level evaluation.

Uncertainty Handling in Other Domains

The approaches to uncertainty handling in other domains, such as weather forecasting (Palmer, 2000), electricity demand forecasting (Taylor & Buizza, 2003), correlations between wind power and meteorological conditions (Lange & Heinemann, 2002), forecast uncertainty in power system planning (Douglas, Breipohl, Lee, & Adapa, 1998), and forecast uncertainty in supply industry (Lo & Wu, 2003) are mainly based on probabilistic representations and stochastic simulations using continuous representations.

Uncertainty representation in approaches to effort estimation in software development effort like (Grimstad & Jørgensen, 2007) and (Gruschke & Jørgensen, 2005) are based on intervals for prior estimates, without a requirement for their propagation within the model. The approach to effort estimation in (Ahmed, et al., 2005) uses fuzzy logic with a learning algorithm for the structure and the parameters. (Sarcia, Basili, & Cantone, 2009) show that uncertainty can be used as an invariant criterion to figure out which effort estimation model should be preferred over others. This work is mainly empirical, applying Bayesian prediction intervals to some COCOMO model variations with respect to a publicly available cost estimation data set coming from the PROMISE repository.

In (Davis & Hall, 2003), an approach to assembling, structuring and representing evidence in a decision, based on hierarchical modeling of the processes leading up to a decision, is presented. Uncertainty in the available evidence is represented and propagated through the evidence hierarchy using interval probability theory. Case studies in the oil and civil engineering industries are used to demonstrate how the approach facilitates developing shared understanding of the implications of uncertainty.

THE LOW-LEVEL EVALUATION

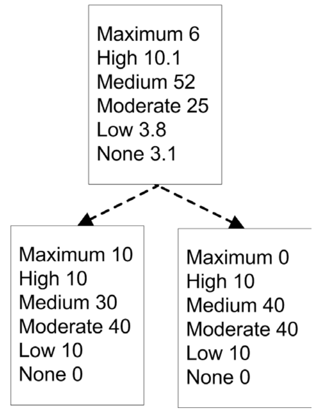
This section presents the results of the low-level evaluation. Each approach identified for the low-level evaluation is:

- instantiated on a WDT, in order to test how its representation form and propagation can be applied, and
- analyzed further with respect to the evaluation criteria.

Bayesian Networks: Based Approach on a WDT

One way to handle uncertainty in a WDT using a BN representation is sketched in Figure 4, where each node is assigned six states. The uncertainty representation is contained in both the granularity (the number of states on the nodes) and their respective probabilities. Then, Baye's rule is applied for the propagation. The prior estimates consist of the probabilities associated with each state on the leaf nodes, as well as the dependencies estimated through the values in the conditional probability table (CPT). The probabilities of the states on each node have to sum up to 100. The CPT in the case of the WDT in Figure 4 consists of 216 (that is,, the product of the number of states

Figure 4. Instantiation of a BN on a WDT



in the children and parent nodes) parameters and must be estimated before the root node can be propagated according to Baye's rule.

Due to the number of the estimates needed, the BN-based approach does not scale. Furthermore, since the uncertainty representation of several states for each node requires extensive CPTs and a demanding propagation, the BN-based approach suffers from lower comprehensibility. The precision depends on the number of states introduced for each node, which means that e.g. new evidence of higher precision than the one defined through the existing states, may be difficult to represent. Hence, combining evidence of different uncertainty (e.g., expert judgments and measurements) on a node may be problematic. The propagation is well-defined but complex.

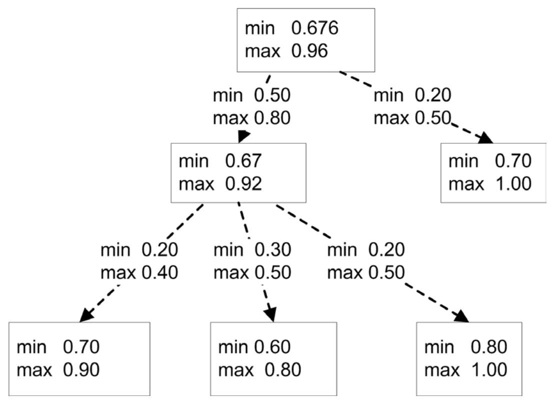
Interval-Based Approach on a WDT

One way of using intervals to capture uncertainty in a WDT is illustrated by Figure 5. The prior intervals are represented by values of type

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

The propagation in Figure 5 is based on a linear model where the parent node interval is obtained by multiplying the respective values on

Figure 5. Uncertainty handling in a WDT using interval-based approach



the arcs and the nodes of the immediate children, and summing up these products. Optimization is performed to find the upper and the lower limits of the inferred interval, so that the quantities on the arcs still sum up to one (Omerovic & Stølen, 2010.). The only two interval arithmetic operations needed for propagation in this example are addition and multiplication. In case of two intervals denoted by x and y (of the form given specified above), addition and multiplication are defined as:

$x \circ y = [x \circ y; x \circ y]$,

where $\circ$ denotes the operation symbol.

The interval-based approach has several advantages. It is relatively straight-forward with respect to both estimate representation and propagation. Moreover, it is comprehensible and easy for the domain experts to relate to, regardless of their formal background. The number of estimates needed is lower than in most of the competing approaches. At the same time, measurement-based input can easily be converted to intervals by relating standard deviation to the interval width. There are no restrictions on the interval width, so the precision is adjustable and can vary among the estimates. Interval arithmetics and interval statistics are well defined and expressive. An issue for this kind of propagation is however overestimation as discussed in relation to the high-level evaluation. Moreover, interval splitting is difficult due to its computational complexity.

Subjective Logic: Based Approach on a WDT

Figure 6 illustrates how we may introduce uncertainty in a WDT using a subjective logic – based representation in the form of opinions on the nodes and applying the arithmetics of subjective logic for the propagation. Ordering was applied on π_{31} and π_{32} to obtain π_{24}. Furthermore, recommendation was applied to obtain π_{12}, consensus

Figure 6. Uncertainty handling in a WDT using subjective logic

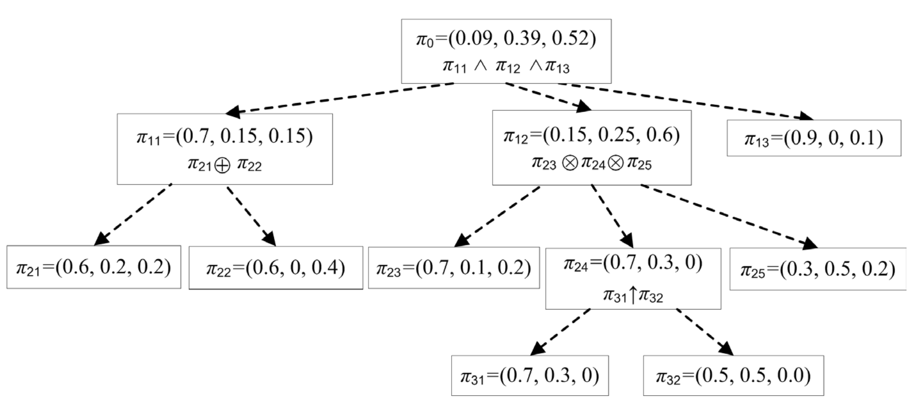
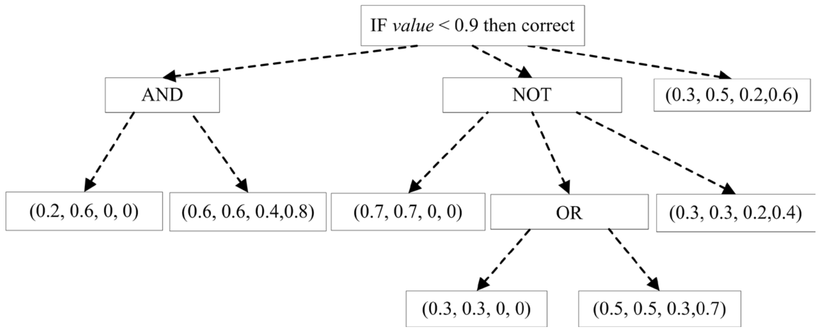


Figure 7. Uncertainty handling in a WDT using fuzzy logic



to obtain π_{11} and conjuction to obtain the opinion of the root node.

The main challenge with the subjective logic approach is relating the notions and values within each opinion to the empirical input. There are no recommendations that we are aware of regarding the interpretation of the notions from the subjective logic to the measurements. How a measurement represented by a mean and a standard deviation can be converted to an opinion, is therefore unclear. However, subjective logic includes the "base rate" (atomicity) that might be interpreted as the deviation whereas the mean is an opinion with low uncertainty. Still, lack of guidelines for interpretation and conversion may degrade comprehensibility, result in misleading input, and disable combining the expert-judgment-based input with the measurement-based one.

The arithmetics are rich and well suited for expert-judgment-based input. The propagation is however demanding, and requires tool support. Given that the interpretation is well defined, the estimate representation based on the subjective logic is precise.

Fuzzy Logic: Based Approach on a WDT

A possible way of using the Fuzzy logic – based approach is illustrated by Figure 7. In this example, the values on the leaf nodes define the membership functions, while the internal nodes express the operators to be applied on the "value", which is a variable expressed on the root node. In practice, there may be many "value" variables expressed within the WDT, and they can then be applied on the operators specified. Several kinds of membership functions can be combined, as illustrated by Figure 7.

The fuzzy approach is expressive in terms of both estimate representation and arithmetics. There is a wide variety of arithmetic operators which can be selected and applied. Although the propagation is non-trivial, tool support is available. The precision level can be selected.

The extensive expressiveness that the membership functions and arithmetic operators offer may however degrade comprehensibility, since their variety also increases complexity if compared to for example the interval-based approach (which can be considered as a special case of the fuzzy one). The increased complexity of the estimate representation may not necessarily contribute to the precision of the parameter values, but rather introduce misinterpretations and incorrectnesses in the provisioning of the expert-judgment-based input. The interpretation of the membership distributions and their correspondence to the practical settings may be demanding.

The measurement-based input may however benefit from the different membership distributions. Since the membership functions can be selected and combined, the preferred one can be used in each case, thus giving a high precision. Thus,

Table 10. Summary of the low-level evaluation

Approach	Criterion						
	C1: Complexity of the representation form for the quantitative estimates.	C2: Suitability of the approach for both expert-judgment-based and measurement-based input.	C3: Scalability with respect to number of estimates needed on a WDT.	C4: Precision allowed by the representation form for the quantitative estimates.	C5: Complexity of the propagation.	C6: Expressiveness of the arithmetic operations for propagation.	C7: Empirical evaluation of the approach, as reported in the publications found through the literature review.
Bayesian networks – based approach	C1_2	C2_4	C3_3	C4_2	C5_3	C6_1	C7_1
Interval-based approach	C1_2	C2_3	C3_1	C4_1	C5_4	C6_3	C7_4
Subjective logic – based approach	C1_3	C2_4	C3_1	C4_1	C5_4	C6_2	C7_5
Fuzzy logic – based approach	C1_1	C2_1	C3_2	C4_1	C5_3	C6_1	C7_1

both expert-judgment-based and measurement-based kinds of input are fully supported.

Summary of the Low-Level Evaluation

Table 10 summarizes the approaches which have undergone low-level evaluation. The properties of each approach are classified according to the definitions of the evaluation criteria and their respective categories, and based on the findings from the evaluation.

The values assigned to the approaches represent the category of the criterion which the approach fits into. Note that in the case of each criterion, the categories have been ordered from highest possible level of fulfillment, to the lowest one. Where appropriate, the domain-expert-judgment-related categories are ranking higher among categories of a criterion, than the measurement-related ones. Thus, the rightmost digit in the category indicates the level of fulfillment of the criterion.

CONCLUSION

Based on a literature review, the potentially useful approaches for uncertainty handling in WDTs are identified and evaluated with respect to a set of evaluation criteria. The evaluation criteria are deduced from the objective of the systematic literature review and related to an established framework for practical system acceptability. Hence, in addition to the literature review and the evaluation, this chapter contributes with a proposal of a set of evaluation criteria regarding selection of uncertainty handling approaches in a WDT context. The proposed criteria are concerned with:

- Complexity of the representation form for the quantitative estimates.
- Scalability of the approach for both expert-judgment-based and measurement-based input.
- Scalability with respect to number of estimates needed on a WDT.
- Precision allowed by the representation form for the quantitative estimates.
- Complexity of the propagation.
- Expressiveness of the arithmetic operations for propagation.

- Empirical evaluation of the approach, as reported in the publications found through the literature review.

The chapter may serve to researchers and practitioners as a roadmap for examining the uncertainty handling approaches, or as a resource for identifying the adequate one. The summary presented in Table 10 indicates a high degree of variation with respect to the scores of the approaches evaluated at low level.

A weakness of all the approaches evaluated at low level is the assumption that the model is certain, while only the input parameters are uncertain. For deterministic systems, this assumption is justifiable. However, in most practical cases, both the model and the input parameters are uncertain. Here, we also have aleatory uncertainty. A model-general uncertainty measure should therefore be introduced in relation to the WDTs, in order to express the uncertainty at an overall level due to for example context uncertainty or usage profile uncertainty.

We argue that the epistemic uncertainty is suited for possibilistic uncertainty representations, that is the representations where pieces of information take the form of fuzzy sets (L. A. Zadeh, 1999) of possible values. The merit of this framework lies in its simplicity, which enables incomplete probabilistic information on the real line to be encoded in the form of fuzzy intervals (Dubois & Prade, 1987). Possibility distributions can also straightforwardly accommodate linguistic information on quantitative scales. As argued by (Möller & Beer, 2008), in many cases of engineering practice, not even subjective probabilistic information is available. Examples are uncertain quantities for which mere bounds or linguistic expressions are known. A probabilistic modeling would then introduce unwarranted information in the form of a distribution function that is totally unjustified.

Selection of the appropriate approach for uncertainty handling implies finding the right balance between the practical applicability on the one hand and the functional properties of the approach on the other hand.

By practical applicability we mean the usability of the approach for the domain experts, who have in-depth system knowledge, but not necessarily a profound insight into formal uncertainty representation. That is, a good approach should offer a representation form for empirical input and uncertain parameter estimates that domain experts find easy to interpret and use. This is partly evaluated through criteria C1 and C5. In addition, usability implies the possibility of combining both expert-judgment-based and measurement-based input, as captured by criterion C2. Another aspect of the practical acceptability is the scalability of the approach with respect to the number of prior estimates needed for generating a WDT, as captured by criterion C3.

The preciseness and expressiveness of the approach are captured by criteria C2, C4 and C6, which also address the functional properties of the approach, that is, handling of epistemic uncertainty and well defined propagation.

The needs regarding practical applicability and functionality are obviously not independent. For example, an approach which is overly advanced in terms of propagation or quantitative estimate representation may be too complex for the domain experts to relate to. Such an approach may be formally sound and precise, but not applicable in a practical setting.

The comprehensibility of an approach may degrade with the increased level of the detail or the size of the model. A coarse grained representation may be more comprehensible, but its model accuracy may degrade and the available input may be lost due to the low expressiveness.

On the other hand, an approach which is well suited for representing measurement-based input may not support equally well the representation of the expert judgments. Similarly, an approach which is very expressive may not scale in an industrial setting, due to the size of the model or the necessity of providing too many parameter estimates.

Based on our evaluation, we can of course not claim that one approach is better than the others for all purposes. The approach must be selected based on the concrete needs and how they are reflected by our evaluation criteria. The criteria should be weighted based on the needs of the context in question. The appropriate trade off is particularly important when the input is based on both expert judgments and measurements, since presence of both kinds of input strengthens the interaction between the functional and the non-functional properties mentioned above.

Nevertheless, it can not be denied that the interval-based approach performs well, due to its well defined formalisms and comprehensibility. The width of the interval is proportional to the uncertainty, and propagation is based on interval arithmetics. 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-based approach can easily be applied on both measurement-based and expert-judgment-based input simultaneously, since the conversion of the uncertain estimate representation of the former to the latter is well defined.

In the case of the Dempster-Shafer structures, the uncertainty may vary across the fractions of the intervals. This additional expressiveness however weakens the comprehensibility due to an extensive increase in the level of detail and may result in compromised correctness of the input.

We argue that Bayesian networks are expressive and precise, but in their general form they are demanding to parametrize and interpret. Thus, they suffer from lower scalability and comprehensibility.

In fuzzy logic the algebraic operations are easy and straightforward. The interval-based approach may be seen as 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 increases complexity of the estimate representation and may introduce misinterpretations and incorrectness in the provisioning of input. The interpretation of the membership distributions and their correspondence to the practical settings may be demanding and therefore unlikely to increase accuracy. Still, the fuzzy logic – based approach performs very well in the evaluation, due to its high expressiveness in terms of representation of the quantitative estimates, rich arithmetics and flexibility in terms of complexity needed to be adopted.

A subjective logic – based approach may provide additional accuracy for domain expert-judgment-based estimates, but how measurement-based input can be converted to the notions within subjective logic, is not clear.

The hybrid approaches address both epistemic and aleatory uncertainty and the propagation is based on stochastic simulation. In most cases, both possibilistic and probabilistic representations are adopted for handling the two respective kinds of uncertainty. Although the hybrid approaches address both kinds of uncertainty, their major weakness is the high complexity, which degrades the overall comprehensibility. They generally focus equally much on both kinds of uncertainty, something that may represent a bias in our case where the epistemic uncertainty is the prevailing one.

Based on our results we recommend the further development of the uncertainty handling approaches to focus on the practical aspects such as scalability, comprehensibility and support for soft computing, in order to facilitate human interaction in the input acquisition. Integrated approaches supporting handling of both measurement-based and expert-judgment-based input in a manner which is comprehensible and accurate, will be particularly useful. Moreover, the existing approaches should be exposed to further empirical investigations in a real life setting.

ACKNOWLEDGMENT

This work has been conducted within the DIGIT (180052/S10) project, funded by the Research Council of Norway. The authors acknowledge the help received from Olav Ligaarden, in performing the systematic literature review.

REFERENCES

Ahmed, M. A., & Muzaffar, Z. (2009). Handling imprecision and uncertainty in software development effort prediction: A type-2 fuzzy logic based framework. *Information and Software Technology*, *51*(3), 640–654. doi:10.1016/j.infsof.2008.09.004

Ahmed, M. A., Omolade Saliu, M., & Al Ghamdi, J. (2005). Adaptive fuzzy logic-based framework for software development effort prediction. *Information and Software Technology*, *47*(1), 31–48. doi:10.1016/j.infsof.2004.05.004

Antonucci, A., Salvetti, A., & Zaffalon, M. (2004). *Assessing debris flow hazard by credal nets. Soft methodology and random Information Systems*.

Armstrong, J. S. (1985). *Long-range forecasting:From crystal ball to computer*. New York, NY: Wiley.

Baraldi, P., Popescu, I. C., & Zio, E. (2008). *Predicting the time to failure of a randomly degrading component by a hybrid Monte Carlo and possibilistic method.* International Conference on Prognostics and Health Management, 2008.

Batarseh, O. G., & Wang, Y. (2008). Reliable simulation with input uncertainties using an interval-based approach. *Proceedings of the 40th Conference on Winter Simulation*.

Baudrit, C., Couso, I., & Dubois, D. (2007). Joint propagation of probability and possibility in risk analysis: Towards a formal framework. *International Journal of Approximate Reasoning*, *45*(1), 82–105. doi:10.1016/j.ijar.2006.07.001

Cano, A., & Moral, S. (1999). *A review of propagation algorithms for imprecise probabilities*. First International Symposium on Imprecise Probabilities and Their Applications.

Cheng, R. C. H., & Holloand, W. (1997). Sensitivity of computer simulation experiments to errors in input data. *Journal of Statistical Computation and Simulation*, *57*(1), 219–241. doi:10.1080/00949659708811809

Cozman, F. G. (2005). Graphical models for imprecise probabilities. *International Journal of Approximate Reasoning*, *39*(2-3), 167–184. doi:10.1016/j.ijar.2004.10.003

Dai, Y.-S., Xie, M., Long, Q., & Ng, S.-H. (2007). Uncertainty analysis in software reliability modeling by Bayesian analysis with maximum-entropy principle. *IEEE Transactions on Software Engineering*, *33*(11), 781–795. doi:10.1109/TSE.2007.70739

Davis, J. P., & Hall, J. W. (2003). A software-supported process for assembling evidence and handling uncertainty in decision-making. *Decision Support Systems*, *35*(3), 415–433. doi:10.1016/S0167-9236(02)00117-3

Douglas, A. P., Breipohl, A. M., Lee, F. N., & Adapa, R. (1998). Risk due to load forecast uncertainty in short term power system planning. *IEEE Transactions on Power Systems*, *13*(4), 1493–1499. doi:10.1109/59.736296

Dubois, D., & Prade, H. (1987). The mean value of a fuzzy number. *Fuzzy Sets and Systems*, *24*(3), 279–300. doi:10.1016/0165-0114(87)90028-5

Fenton, N. E., & Neil, M. (1999). A critique of software defect prediction models. *IEEE Transactions on Software Engineering*, *25*(5), 675–689. doi:10.1109/32.815326

Ferson, S., Kreinovich, V., Hajagos, J., Oberkampf, W., & Ginzburg, L. R. (2007). *Experimental uncertainty estimation and statistics for data having interval uncertainty (no. SAND2007-0939)*. Sandia National Laboratories.

Fioretti, G. (2004). Evidence theory: A mathematical framework for unpredictable hypotheses. *Metroeconomica*, *55*(4), 345–366. doi:10.1111/j.1467-999X.2004.00197.x

Grimstad, S., & Jørgensen, M. (2007). Inconsistency of expert judgment-based estimates of software development effort. *Journal of Systems and Software*, *80*(11), 1770–1777. doi:10.1016/j.jss.2007.03.001

Gruschke, T., & Jørgensen, M. (2005). Assessing uncertainty of software development effort estimates: The learning from outcome feedback. *Proceedings of the 11th IEEE International Software Metrics Symposium*.

Haenni, R. (2007). *Climbing the hills of compiled credal networks*. 5th International Symposium on Imprecise Probabilities and Their Applications.

Heckerman, D., Mamdani, A., & Wellman, M. P. (1995). Real-world applications of Bayesian networks. *Communications of the ACM*, *38*(3), 24–26. doi:10.1145/203330.203334

Hyman, R. (1982). *Quasi-experimentation: Design and analysis issues for field settings* (vol. 46).

IEC (1995). *Dependability management - Part 3: Application guide - Section 9: Risk analysis of technological systems - Event tree analysis*. International Electrotechnical Commission.

IEC. (2006). *Fault tree analysis (FTA)*. International Electrotechnical Commission.

ISO. (2008). *Uncertainty of measurement -- Part 3: Guide to the expression of uncertainty in measurement*. International Organisation for Standardization.

Jøsang, A. (1997). *Artificial reasoning with subjective logic*. 2nd Australian Workshop on Commonsense Reasoning.

Kearfott, R. B. (1994). Interval computations: Introduction, uses, and resources. *Euromath Bulletin, 1*(2).

Kitchenham, B., Charters, S., Budgen, D., Brereton, P., Turner, M., Linkman, S., … Visaggio, E. (2007). *Guidelines for performing systematic literature reviews in software engineering* (EBSE Technical Report). Keele University

Kiureghian, A. D., & Ditlevsen, O. (2009). Aleatory or epistemic? Does it matter? *Structural Safety*, *31*(2), 105–112. doi:10.1016/j.strusafe.2008.06.020

Kleiter, G. D. (1996). Propagating imprecise probabilities in Bayesian networks. *Artificial Intelligence*, *88*(1-2), 143–161. doi:10.1016/S0004-3702(96)00021-5

Kreinovich, V., Hajagos, J. G., Tucker, W. T., Beck, J., Ginzburg, L. R., & Ferson, S. (2008). *Propagating uncertainty through a quadratic response surface model (Technical report)*. Sandia National Laboratories.

Lange, M., & Heinemann, D. (2002). *Accuracy of short term wind power predictions depending on meteorological conditions.* Global Windpower Conference.

Levi, I. (1980). *The enterprise of knowledge: An essay on knowledge, credal probability, and chance*. Cambridge, MA: MIT Press.

Lo, K. L., & Wu, Y. K. (2003). Risk assessment due to local demand forecast uncertainty in the competitive supply industry. *IEE Proceedings. Generation, Transmission and Distribution*, *150*(5), 573–581. doi:10.1049/ip-gtd:20030641

Majumdar, S., Johannes, L., Haring, G., & Ramadoss, R. (2001). *Characterization and analysis of software and computer systems with uncertainties and variabilities*. Performance Engineering, State of the Art and Current Trends.

Majumdar, S., & Ramadoss, R. (1995). Interval-based performance analysis of computing systems. *Proceedings of the 3rd International Workshop on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems*.

Möller, B., & Beer, M. (2008). Engineering computation under uncertainty - Capabilities of non-traditional models. *Computers & Structures*, *86*(10), 1024–1041. doi:10.1016/j.compstruc.2007.05.041

Neil, M., Fenton, N., & Nielson, L. (2000). Building large-scale Bayesian networks. *The Knowledge Engineering Review*, *15*(3), 257–284. doi:10.1017/S0269888900003039

Nielsen, J. (1993). *Usability engineering*. Boston, MA: Academic Press.

Omerovic, A., Andresen, A., Grindheim, H., Myrseth, P., Refsdal, A., Stølen, K., & Ølnes, J. (2010). *A feasibility study in model based prediction of impact of changes on system quality (SINTEF report No. A13339)*. SINTEF.

Omerovic, A., & Stølen, K. (2009). *Simplifying parametrization of Bayesian networks in prediction of system quality.* Third IEEE International Conference on Secure Software Integration and Reliability Improvement, 2009.

Omerovic, A., & Stølen, K. (2010.). *Interval based uncertainty handling in model based prediction of system quality*. 2nd International Conference on Advances in System Simulation

Palmer, T. N. (2000). Predicting uncertainty in forecasts of weather and climate. *Reports on Progress in Physics*, *63*(2), 71. doi:10.1088/0034-4885/63/2/201

Renooij, S., & van der Gaag, L. C. (2002). *From qualitative to quantitative probabilistic networks*. Eighteenth Conference on Uncertainty in Artificial Intelligence.

Sarcia, S. A., Basili, V. R., & Cantone, G. (2009). Using uncertainty as a model selection and comparison criterion. *Proceedings of the 5th International Conference on Predictor Models in Software Engineering*.

Schneier, B. (1999). Attack trees: Modeling security threats. *Dr. Dobb's Journal*, *24*(12), 21–29.

Scotney, B., & McClean, S. (2003). Database aggregation of imprecise and uncertain evidence. *Information Sciences*, *155*(3-4), 245–263. doi:10.1016/S0020-0255(03)00172-5

Shafer, G. (1976). *A mathematical theory of evidence*. Princeton, NJ: Princeton University Press.

Simon, C., Weber, P., & Levrat, E. (2007). Bayesian networks and evidence theory to model complex systems reliability. *JCP*, *2*(2), 33–43.

Skelboe, S. (1974). Computation of rational interval functions. *BIT*, *14*, 87–95. doi:10.1007/BF01933121

Suresh, P. V., Babar, A. K., & Raj, V. V. (1996). Uncertainty in fault tree analysis: A fuzzy approach. *Fuzzy Sets and Systems*, *83*(2), 135–141. doi:10.1016/0165-0114(95)00386-X

Taylor, J. W., & Buizza, R. (2003). Using weather ensemble predictions in electricity demand forecasting. *International Journal of Forecasting*, *19*, 57–70. doi:10.1016/S0169-2070(01)00123-6

Walley, P. (1991). *Statistical reasoning with imprecise probabilities*. London, UK: Chapman and Hall.

Weber, D. P. (1994). Fuzzy fault tree analysis. *Proceedings of the Third IEEE Conference on Fuzzy Systems, IEEE World Congress on Computational Intelligence*.

Wellman, M. P. (1990). Fundamental concepts of qualitative probabilistic networks. *Artificial Intelligence*, *44*(3), 257–303. doi:10.1016/0004-3702(90)90026-V

Yin, L., & Trivedi, K. S. (1999). Confidence interval estimation of NHPP-based software reliability models. *Proceedings of the 10th International Symposium on Software Reliability Engineering*.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, *8*, 338–353. doi:10.1016/S0019-9958(65)90241-X

Zadeh, L. A. (1999). Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, *100*(Supplement 1), 9–34. doi:10.1016/S0165-0114(99)80004-9

Zaffalon, M., Wesnes, K., & Petrini, O. (2003). Reliable diagnoses of dementia by the naive credal classifier inferred from incomplete cognitive data. *Artificial Intelligence in Medicine*, *29*(1-2), 61–79. doi:10.1016/S0933-3657(03)00046-0

KEY TERMS AND DEFINITIONS

Bayesian Network: A probabilistic graphical model that represents a set of random variables and their conditional independencies via a directed acyclic graph.

Dependency Tree: A finite tree in which: each node may hold one or more quantitative values or a function, each arc may hold one or more a quantitative values or a function, and each inferred non-leaf node is a tree-specific function of its immediate children and their connecting arcs.

Dependency View: A dependency tree in which each node holds a quantitative value representing degree of fulfillment of a system quality characteristic, each arc holds a quantitative value representing the estimated impact of the child node on the parent node, and each inferred non-leaf node is a linear function of its immediate children and their connecting arcs.

Fuzzy Logic: A form of multi-valued logic derived from fuzzy set theory to deal with reasoning that is approximate. In contrast with “crisp logic”, where binary sets have binary logic, fuzzy logic variables may have a truth value that ranges between 0 and 1 and is not constrained to the two truth values of classic propositional logic.

Inaccuracy: The condition of being imprecise or insufficiently exact.

Intervals: A set containing all points (or all real numbers) between two given endpoints.

Modeling: Representing, designing and analyzing a representation of a system to study the effect of changes to system variables.

Simulation: The imitation of some real thing, state of affairs, or process. The act of simulating something generally entails representing certain key characteristics or behaviours of a selected physical or abstract system.

Subjective Logic: A type of probabilistic logic that explicitly takes uncertainty and belief ownership into account.

Uncertainty: The condition in which reasonable knowledge or information regarding the present state or the future is not available.

APPENDIX A

THE LITERATURE REVIEW

This section presents the steps of a literature review we conducted prior to the identification and evaluation of the potentially useful approaches for uncertainty handling. Firstly, we performed a mapping study in order to gain an overview of the main topics and possible keywords. Secondly, we performed a systematic literature review in order to reveal state of the art based on a rigorous walkthrough of the available publications. The remainder of this section presents the process and the results of the mapping study and the systematic literature review, respectively.

The Mapping Study

The pre-review mapping study was performed in February 2010, to help in scoping the research question and to get a preliminary overview of the existing uncertainty handling approaches. The mapping study involved a search in ACM Digital Library, IEEE Xplore, Lecture Notes in Computer Science and Google Scholar with keywords such as “uncertainty”, “inaccuracy”, “imprecise”, “predictability” and “predictability domain”. The keywords applied were general upon the beginning, and then more focused when the follow-up chapters were recorded. No constraints on the time since publishing were used on the resources acquired. About fifty publications from mathematics, statistics and computer science which presented either single or combined approaches for uncertainty handling, were extracted. The ones addressing customization of identified approaches for the different domains and the ones focusing on mainly aleatory uncertainty, were excluded, thus reducing the search results to about 33 publications. Moreover, uncertainty and inaccuracy handling in other domains, were searched for in the same resources, by adding the keywords using as: “meteorology”, “weather forecast”, “effort estimation”, “uncertainty” and “inaccuracy” individually and in combination. About ten additional publications were recorded. A walkthrough of the search results was made. The mapping study resulted in 38 articles addressing the uncertainty handling approaches relevant in the context of system quality prediction based on the weighted dependency trees.

The Systematic Literature Review

For the systematic literature review we followed a procedure recommended by (Kitchenham, et al., 2007). The systematic literature review was performed in March and April 2010. The first phase of planning a systematic literature review is to specify the research objective that should be satisfied by conducting the review. The objective of this systematic literature review is to identify and assess the potentially useful approaches which:

1. are practically applicable for supporting uncertainty handling in the context of WDT-based analysis,
2. have a well defined propagation within the WDT, and
3. handle the epistemic uncertainty.

By an approach we mean the collection of the formal and the graphical means available along with the procedure prescribed for applying those means.

The second step of planning a systematic literature review is to develop the search strategy, as a part of developing the review protocol. As the search specified below captures a wide range of publications, we developed a stopping heuristic and applied the following protocol:

1. The researchers clarify the objective and make sure that the background of the study, the research question and the research method are understood.
2. For each of the search strings, run the search on all five search engines.
3. Researcher A and Researcher B independently record and collect the bibliographic information for any source not previously recorded that appeared to be relevant, based on title and type of publication.
4. Continue through the search results until reaching 10 consecutive, irrelevant results.
5. Researchers record each other's result lists from steps 3 and 4, and vote on the relevant titles based on the objective and the research question.
6. Consolidate the results obtained by Researcher A and Researcher B, from step 5.
7. Researcher A and Researcher B independently record the abstracts of the titles collected in step 6, and vote on the relevance (by assigning binary votes: relevant or not relevant) of each abstract.
8. Consolidate the results obtained by Researcher A and Researcher B, from step 5.
9. Researcher A reviews full texts of the selected publications, based on the criteria (presented in the following section).

In step 2 of the protocol we used the following search engines to conduct our review:

1. Google scholar
2. IEEE explore
3. Science direct
4. ACM digital library
5. ACM guide

The keywords used were deduced from the research objective. Since the different search engines had different search categories, kinds of possible constraints and various degrees of support for logical keyword compositions, the keyword sets and constraints used in step 2 varied as specified below.

Two sets of keywords were searched on Google scholar:

1. (uncertainty OR inaccuracy OR imprecise OR imprecision OR uncertain OR inaccurate) (system OR software OR meteorology OR "weather forecasting" OR "effort estimation") (prediction OR simulation OR model OR models OR modeling OR modelling)
2. Uncertainty (prediction OR simulation OR model OR models OR modeling OR modelling) (software OR system) (tree OR graph OR DAG)

Each search was constrained to articles excluding patents with at least summaries, published since year 2000. The first search resulted in 796000 items, and the second one in 223000 items.

Two sets of keywords were searched on IEEE explore:

1. ((uncertainty OR inaccuracy OR imprecise OR imprecision OR uncertain OR inaccurate) AND (system OR software OR meteorology OR "weather forecasting" OR "effort estimation") AND (prediction OR simulation OR model OR models OR modeling OR modelling))
2. (Uncertainty AND (prediction OR simulation OR model OR models OR modeling OR modelling) AND (software OR system) AND (tree OR graph OR DAG))

Each search was constrained to conferences and journals as types of content within the subject defined as "Computing and Processing (Hardware/Software)", published between years 2000 and 2010. The first search resulted in 329212 items, and the second one in 290038 items.

Two sets of keywords were searched on Science direct:

1. ((uncertainty OR inaccuracy OR imprecise OR imprecision OR uncertain OR inaccurate) AND (system OR software OR meteorology OR "weather forecasting" OR "effort estimation") AND (prediction OR simulation OR model OR models OR modeling OR modelling))
2. (Uncertainty AND (prediction OR simulation OR model OR models OR modeling OR modelling) AND (software OR system) AND (tree OR graph OR DAG))

Each search was constrained to the computer science category published since 1999. The first search resulted in 3629 items, and the second one in 5631 items.

Two sets of keywords were searched on both ACM digital library and ACM guide:

1. uncert* software*
2. uncert* model*

The stars in the keywords allow for arbitrary ending of the keyword. Each search was constrained to items published since 2000. The search in ACM guide with the first and the second keyword set resulted in 607 and 1232 items, respectively. The search in ACM digital library with the first and the second keyword set resulted in 2879 and 8462 items, respectively.

The entire process is sketched on Figure 8. Since the search engines used sort the results by relevance, this procedure was considered to be effective at producing search results. In most cases, the density of relevant titles steadily decreased as we examined each list of the search results.

In the first pass (step 3 of the protocol), the researchers were as inclusive as possible, since the filtering was only based on titles and the types of the publications. All kinds of publications except sole abstracts, books, presentations and white chapters dating no more that ten years, were included. The procedure resulted in the data presented in Table 11. The reviewed titles and the passed titles columns come from executing the steps 3 and 4 of the protocol. The reviewed titles column reports on the sequence number on the search result list, of the last title. The fifth column of Table 11 indicates which researcher the results from columns 3 and 4 were obtained by.

Next, the researchers voted on the relevance of the titles (step 5 of the protocol). Researcher A voted on whether the results found by Researcher B were relevant to the review, and vice versa. Researcher A collected 91 titles, of which Researcher B accepted 64. Researcher B collected 61 sources, of which Researcher A accepted 55. A total of 89 (not the sum of the two due to the overlaps) unique titles from the voting confirmation was accepted to proceed to the next phase of the review. The last column of

Figure 8. The process applied in the literature review

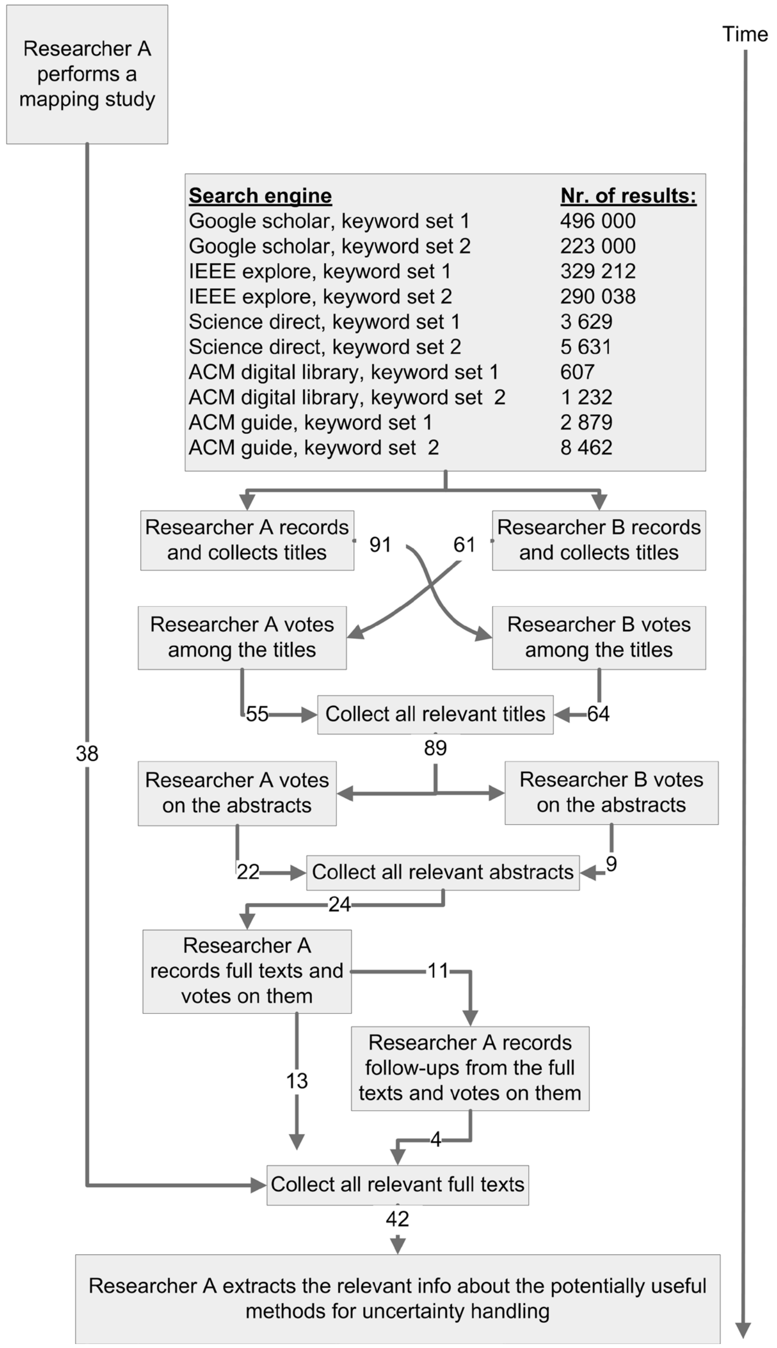


Table 11 reports on the number of the titles deemed relevant by the other researcher during step 5 of the protocol.

Step 6 was done by collecting all titles assessed as relevant by at least one researcher. The 89 unique abstracts of the publications that had passed the step 6 of the protocol were recorded by each researcher

Table 11. Summary of the title votes

Search engine	Keyword set	Reviewed titles	Passed titles	Researcher	Nr of the relevant titles
Google scholar	1	51	14	A	9
Google scholar	2	38	6	A	5
IEEE explore	1	15	11	A	11
IEEE explore	2	44	6	A	6
Science direct	1	20	6	A	6
Science direct	2	34	4	A	2
ACM digital library	1	17	6	A	5
ACM digital library	2	18	10	A	7
ACM guide	1	19	13	A	8
ACM guide	2	19	15	A	13
Google scholar	1	26	9	B	6
Google scholar	2	26	6	B	5
IEEE explore	1	41	15	B	13
IEEE explore	2	35	10	B	10
Science direct	1	12	4	B	4
Science direct	2	9	2	B	2
ACM digital library	1	12	2	B	2
ACM digital library	2	36	7	B	7
ACM guide	1	23	6	B	6
ACM guide	2	0	0	B	0

separately. Each researcher voted whether the abstract was relevant or not. Researcher A voted yes to 22 abstracts and Researcher B voted yes to 9 abstracts. The abstracts that at least one researcher deemed relevant, were collected. This resulted in 24 (not the sum of the two due to the overlaps) relevant abstracts.

Step 8 was done by collecting all abstracts assessed as relevant by at least one researcher. The full texts that had passed step 8 of the protocol were recorded by Researcher A and evaluated for the relevance and applicability in the context of system quality prediction based on the weighted trees. A total of 13 texts were found relevant. Additionally, 11 distinct follow-up articles (referenced from the recorded full texts) were recorded and 4 of them were found relevant. The total number of the unique full texts evaluated as relevant during the mapping study and the systematic literature review was 42. Thus, there were 13 overlapping articles between the mapping study and the systematic literature review, and 4 of the overall 29, were found through the systematic literature review, while 25 were found through the mapping study. 19 of the 25 relevant articles discovered only through the mapping study were published before 2000 – the timeframe which was discluded from the searches during the systematic literature review.

MS Excel sheets were found convenient and used for documenting the votes, counting the results and identifying the overlapping titles.

APPENDIX B

DEDUCTION OF THE EVALUATION CRITERIA

This section presents in detail the deduction of the evaluation criteria specified in tables two through nine. The systematic literature review provided, in addition to a preliminary evaluation, an initial list of the evaluation criteria. Based on the results of the systematic literature review and the practical experiences, the evaluation criteria were iteratively refined through multiple discussions among the authors. The deduction of the criteria was partially guided by an existing taxonomy for system acceptability. Thus, the original list of criteria was incrementally refined during the study.

The objective of the systematic literature review has been decomposed into three sub-objectives specified in Table 1. These sub-objectives are specified and decomposed into aspects. Each aspect implies one or more evaluation criteria. Each criterion was then defined and specified in detail, further decomposing it into a set of distinct categories of fulfillment.

The taxonomy for system acceptability proposed in (Nielsen, 1993) has been adapted to our objective, by mapping the sub-objectives of the systematic literature review to the relevant parts of the taxonomy. Table 12 shows the structure of the model by Nielsen, which we were guided by towards a refined definition of the evaluation criteria. The nodes contained in the model by Nielsen are presented in the form of italic (bold and non-bold) text. The bold italic text represents the adopted ones, while the overall elements from the table which are written in italic are contained in the model by Nielsen but not needed for our objective. In addition, Table 12 shows the aspects we have found necessary to add in order to fully cover the objective specified in Table 1. The added aspects are presented in the form of non-italic text. The aspects adopted from Nielsen or added are annotated with the name of the aspect of one of out three sub-objectives and the criteria (denoted as CX, where X is the number of the criterion) implied by the aspect. Thus, we only focus on two branches of the practical acceptability (usefulness and scalability).

The first sub-objective of the systematic literature is non-functional, while the remaining two are functional. Hence, utility (or functionality) in our context is:

- handling of epistemic uncertainty, and
- well defined propagation.

Table 12. Decomposition of system acceptability, based on our objective and motivated by the model of the attributes of system acceptability provided in (Nielsen, 1993)

<table>
<tr><th colspan="8">System acceptability</th></tr>
<tr><td colspan="7">Practical acceptability</td><td rowspan="4">Social acceptability</td></tr>
<tr><td rowspan="3">Cost, compatibility, reliability, etc.</td><td rowspan="3">Scalability (C3)</td><td colspan="5">Usefulness</td></tr>
<tr><td colspan="3">Usability</td><td colspan="2">Utility</td></tr>
<tr><td>Easy to learn, easy to use, easy to remember, few errors, subjectively pleasing</td><td>Complexity of estimate representation form (C1)</td><td>Suitability of estimate representation form for expressing input (C2)</td><td>Handling epistemic uncertainty (C2, C4)</td><td>Well defined propagation (C5, C6)</td></tr>
</table>

The practical applicability is de-composed into two aspects: scalability and usability. We consider that this influences usability:

- the complexity of the estimate representation form, and
- suitability of estimate representation for expressing both domain expert-judgment-based and measurement-based input.

The aspects identified are related to evaluation criteria. Seven evaluation criteria have been identified, which in conjuction cover the objective and form the basis for evaluation of the potentially useful approaches. Six of the criteria are related to the individual aspects, while the last criterion addresses the empirical evaluation of the approaches and therefore spreads across all the aspects of the three sub-objectives.

Assessing an uncertainty handling approach should not only be based on accuracy of estimates made, but on the practical applicability aspects discussed above. In the sequel we state two necessary prerequisites for the approaches evaluated and define a set of evaluation criteria in relation to the sub-objectives of the systematic literature review.

The epistemic uncertainty is better suited for possibilistic uncertainty representations, than the probabilistic ones. Possibility distributions can also straightforwardly accommodate linguistic information on quantitative scales. A probabilistic modeling may introduce unwarranted information in the form of a distribution function that is not sufficiently justified. Continuous probability distributions are also more challenging to provide and interpret by the domain experts, and the precision expressed by the probability distribution may be higher than what in reality is the case. We only focus on the epistemic uncertainty associated with the empirical input which is acquired from the domain experts and from the measurements. Moreover, practical applicability of the approach is one of the sub-objectives. Therefore, two prerequisites for the approaches evaluated are identified:

- only discretized estimates or the estimates that can be discretized are considered, thus accommodating for a possibilistic approach; and
- the number of the prior estimates (that is, the estimated to be fed into the model before the propagation can be performed) that can be acquired is limited.

Both prerequisites are due to the conditions that the practical settings in which the approaches are to be applied in, define. The prerequisites are motivated by the experiences from practical trials, in particular development and application the models for prediction of system quality, in the form of WDTs (Omerovic, et al., 2009).

An uncertainty handling approach should facilitate provisioning of reliable estimates which are provided by domain experts and through other forms of data acquisition. Finding optimal balance between soundness and practical applicability is challenging due to the mutual interaction between the two. While for example precision (as an attribute of soundness) of the representation may improve efficiency, it may be difficult to understand, achieve and apply the higher precision in a practical setting. The criteria defined are therefore not mutually exclusive, and their interaction and importance may depend on the objective of the system analyzed and the context. Such a weighting of the criteria is however beyond the scope of this chapter.

The objective for the uncertainty handling approaches has been decomposed into the three sub-objectives specified in Table 1. Each sub-objective is elaborated on below, by interpreting it, decomposing it to aspects, and refining the aspects in the form of evaluation criteria. Table 2 provides an overview of the evaluation criteria deduced in the sequel.

Sub-Objective 1: Practical Applicability

The aspects of the practical applicability are first deduced by the authors, based on the experiences from development and application the models for prediction of impacts of architecture design changes on the system quality, in the form of WDTs (Omerovic, et al., 2009). Thereafter, the aspects are related to the relevant elements within the practical acceptability framework proposed in (Nielsen, 1993).

The practical applicability from the point of view of the domain experts implies two aspects:

1. the usability of the approach, and
2. the scalability of the approach.

The first aspect addresses whether the representation form of the quantitative estimates is suitable for the domain experts to relate to. In particular, complexity and suitability of the notions used by the approach are assessed. 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 techniques. We assume that complexity of the quantitative notions influences comprehensibility. The approach should facilitate provisioning of the uncertain WDT estimates which are easy for domain experts to provide and interpret. Thus, for the first aspect of practical applicability we propose the following criteria:

- C1: complexity of the representation form for the quantitative estimates. The criterion is specified in Table 3.
- C2: suitability of the approach for both expert-judgment-based and measurement-based input. The criterion is specified in Table 4.

Table 4. Specification of Criterion C2

Category number	Criterion C2: Suitability of the approach for both expert-judgment-based and measurement-based input.
C2_1	Both domain expert-judgment-based and measurement-based input can be directly related to and expressed by the notions of the approach for representing quantitative estimates.
C2_2	It is easy to directly express the measurement-based input but the domain expert judgments can only be expressed if its correspondence with the notions of the approach is defined.
C2_3	It is easy to directly express the domain expert judgments but the measurement-based input can only be expressed if its correspondence with the notions of the approach is defined.
C2_4	Neither the domain expert judgments nor the measurement-based input can be directly related to and by the notions of the approach regarding representation of the quantitative estimates. Their correspondence with the notions of the approach has to be defined.

Table 5. Specification of Criterion C3

Category number	Criterion C3: Scalability with respect to number of estimates needed on a WDT.
C3_1	Linearly proportional to the number of nodes with a low coefficient. A practically managable number of estimates even on a large WDT.
C3_2	A practically managable number of estimates but only for a restricted WDT with few nodes. Linearly proportional to the number of nodes, but a large coefficient.
C3_3	Exponential relationship between number of nodes and the number of estimates needed. Practically demanding to provide the large number of estimates.

In order to be practically applicable on a realistic system, the approach has to scale with respect to the number of prior estimates that need to be provided in order to create a WDT and before any propagation can start. Thus, for the second aspect of practical applicability we propose criterion C3:

- C3: scalability with respect to number of estimates needed on a WDT. The criterion is specified in Table 5.

Sub-Objective 2: Handling of the Epistemic Uncertainty

This sub-objective is rooted in the need for addressing the deterministic systems and involves explicitly expressing the epistemic uncertainty. The explicit expression involves precisely and correctly representing the quantities of the uncertain estimates. The notions used by the approach have to be related to the empirical input and its representation. This implies again the above specified criterion C2.

How fine grained the expressions of the quantitative estimate and their uncertainty may be, is an important property of the approach. In case the precision is constrained, information about the actual degree of uncertainty may be lost. Each individual parameter estimate and its associated uncertainty needs to be represented in as a precise form as possible to provide by the domain experts or the other sources for data acquisition. The right granularity of the uncertain estimate representation at the level of each parameter is needed, thus we propose criterion C4: precision allowed by the representation form for the quantitative estimates. The criterion is specified in Table 6.

Table 6. Specification of Criterion C4

Category number	Criterion C4: Precision allowed by the representation form for the quantitative estimates.
C4_1	Granularity of the uncertainty related to the empirical input can vary between and within parameters of a WDT.
C4_2	Granularity of the uncertainty related to the empirical input can vary between but not within parameters of a WDT.
C4_3	Granularity of the uncertainty related to the empirical input can be selected at a model level, but is, once defined by a scale for the model, frozen and can not vary within the WDT.
C4_4	Representation of uncertainty related to the empirical input is coarse grained, frozen and can not vary within the WDT.

Table 7. Specification of Criterion C5

Category number	Criterion C5: Complexity of the propagation.
C5_1	Simple propagation which can be performed without tool support.
C5_2	Propagation can be followed by the domain experts without particular formal background. Tool support is needed and available.
C5_3	Propagation can be followed by the domain experts without particular formal background. Tool support is needed but its availability is not reported.
C5_4	Propagation is demanding to understand by the domain experts without formal background. Tool support is needed and available.
C5_5	Propagation is demanding to understand by the domain experts without formal background. Tool support is needed but its availability is not reported.

Sub-Objective 3: Well Defined Propagation

The sub-objective of a well defined propagation addresses the complexity and expressiveness of the propagation techniques provided by the approach. The values on the leaf nodes and arcs of the WDTs have to be propagated according to an applicable technique which is suitable for reasoning on dependencies regarding both expert-judgment-based estimates and measurement-based input.

Thus, we propose the following two criteria:

- C5: complexity of the propagation, and
- C6: the expressiveness of the arithmetic operations for propagation.

The criteria C5 and C6 are specified in Table 7 and Table 8, respectively.

A General Evaluation Criterion

The empirical evaluations provide input on the performance of the approach with respect to all three sub-objectives and the entire objective in general. The existing empirical evaluations of the approach will give an indication of its general applicability in an industrial setting on a realistic system. Thus, this general aspect implies criterion C7: empirical evaluation of the approach, as reported in the publications found through the literature review. The criterion is specified in Table 9.

Table 8. Specification of Criterion C6

Category number	Criterion C6: Expressiveness of the arithmetic operations for propagation.
C6_1	The arithmetic operations fulfill the needs for inference of all kinds of empirical input and its uncertainty.
C6_2	The arithmetic operations are more suited for subjective estimates than for propagation of the measurement-based input.
C6_3	The arithmetic operations are more suited for propagation of measurement-based input than for propagation of the subjective estimates.
C6_4	There are only limited and insufficient kinds of arithmetic operations for propagation of both expert-judgment-based and measurement-based kinds of input and its uncertainty.

APPENDIX C

THREATS TO VALIDITY AND RELIABILITY

The aim of this section is to address the main issues regarding the validity and the reliability of this evaluation. We address the research method, its effects on the evaluation and the validity of the main findings. In addition to reliability threats, four types of validity threats presented in (Hyman, 1982) are addressed below: conclusion validity, construct validity, internal validity and external validity.

Overall, the main threats to validity are:

1. publication bias - there might be unpublished negative experiences with different approaches - this issue was not possible to mitigate, but we do not consider it to represent a problem,
2. selection of articles – there might be other which are not revealed through the literature review or not included in the study,
3. data extraction, and
4. identification of criteria.

The high-level evaluation is based on the detailed information extraction of the approaches from the publications found through the literature review. The high-level evaluation presents exclusively the properties reported within the relevant literature obtained through the mapping study and the literature review. The information otherwise available is not taken into account. The quality of the results is therefore dependent on the thoroughness of the research method. Inspite of the structured process during the literature review, the contextual factors are difficult to fully eliminate. One should therefore view the high-level evaluation within the frames of and the quality of the underlying research method, particularly the literature review and the data extraction.

The low-level evaluation is however based on the instantiations in the WDTs and the discussion of the findings from the high-level evaluation, with respect to the evaluation criteria. The instantiation is based on the presentations of the respective approaches, provided in the literature obtained. Still, the low-level evaluation is to a much lesser extent bound to the literature review.

Reliability is concerned with demonstrating that the study can be repeated with the same results. Two of the factors influencing the reliability of the literature review are: the subjective influence of the researcher A on the mapping study and the step 9 of the protocol. These steps were only performed by one researcher. During the overall steps of the protocol, the consensus in any deviations was made by consolidating all the results, thus a third voter was unnecessary. In addition, the information extraction and the instantiation are, although systematic and based on the evaluation criteria, exposed to minor subjective judgments. However, since the evaluation is driven by the criteria which are clearly defined, the bias should be minimal. The summary provided in Table 10 involved negligible doubt when assigning the criteria categories to the individual approaches.

The main threats to validity for this study are publication selection bias, inaccuracy in data extraction, and misclassification. To help ensure an unbiased selection process, we defined the objective of the evaluation in advance, organized the selection of articles as a multistage process, and involved two researchers in the literature review. Still, the process was complex and we may not have managed to detect all the publications that are relevant for inclusion. Moreover, data extraction from prose is difficult due to lack of both standard terminology and standards for reporting the properties of the approaches.

This may have resulted in some inaccuracy in the data. Another challenge was that there is no up to date keyword standard that we are aware of that exactly captures the topics relevant for this study.

Moreover, except from the framework in (Nielsen, 1993) which we used to deduce the evaluation criteria, we have not revealed any up to date standard for evaluation criteria or generally accepted methodology for evaluation of approaches. Hence, the criteria deduced are, apart from the evaluation itself, another one of the contributions of the chapter. The variation of the assessments in the evaluation summary indicated that the criteria and their respective categories are capable of differentiating between the approaches at in intended detail level.

Construct validity concerns whether we measure what we believe we measure. In this context, construct validity is the traceability of the conjuction of the evaluation criteria to the overall objective. This aspect has been addressed during the deduction of the criteria.

Conclusion validity concerns the composition of participants and the statistical analysis. In this context, conclusion validity is the breadth and quality of the literature review. Given the total number of publications the literature review has covered, we have reason to believe in strength of the conclusion validity. The subjective aspects of the literature review and particularly the parts performed by only one researcher present the major threats to the conclusion validity.

Internal validity concerns matters that may affect the causality of an independent variable, without the knowledge of the researcher. In this context, internal validity is the traceability of the evaluation itself to the respective criteria, that is, whether we actually provide the evaluation asked for by the criteria. Since the evaluation is purely based on the literature uncovered during the literature review and since the criteria are defined with distinct categories, the threats to the internal validity have been addressed.

External validity concerns the generalization of findings of this case study to other contexts and environments. The evaluation is limited to the systems where the epistemic uncertainty is the dominating one. The epistemic uncertainty is the crucial one in the context of the deterministic systems and therefore given most attention. (Baraldi, et al., 2008) argues about the limitations associated to a probabilistic representation of epistemic uncertainty under limited information. Being of a discrete and non-stochastic nature, the epistemic uncertainty should be handled by a possibilistic approach. Hence, we have only addressed the epistemic uncertainty. This is however a simplification, since the aleatory uncertainty is present mainly due to the usage profile. Ideally, both the epistemic and the aleatory uncertainty should be addressed by a hybrid approach. Such an approach would handle the epistemic and the aleatory uncertainty by possibilistic and probabilistic approaches respectively, and by weighing the two types of uncertainty in a system. It would however need to be much more complex, which would compromise its usability.