ABSTRACT

ATHE, PARIDHI. A Framework for Predictive Capability Maturity Assessment of Computer Simulation Codes. (Under the direction of Dr. Nam Dinh).

This work presents a formalized and computerized framework for the assessment of decision regarding the adequacy of a simulation tool for a nuclear reactor application. The adequacy of a simulation code for an intended application is determined by verification, validation and uncertainty quantification (VVUQ) of the code. Therefore, the decision regarding code adequacy is dependent on the assessment of different attributes that govern verification, validation and uncertainty quantification of the code. In this work, the focus is on code validation. Therefore, the framework is developed and illustrated from the perspective of decision regarding the validation assessment of code. Code validation assessment is performed based on the validation test results, data applicability and process quality assurance factors. The process quality assurance factors warrant the trustworthiness of the evidence and help in checking people and process compliance with respect to the standard requirements.

The proposed framework is developed using an argument modeling technique called Goal Structuring Notation (GSN). Goal structuring notation facilitates structural knowledge representation, information abstraction, evidence incorporation and provides a skeletal structure for quantitative maturity assessment. The decision schema for the development of the decision model is based on the Predictive Capability Maturity Model (PCMM) and Analytic Hierarchy Process (AHP), and formalized using Goal structuring notation. Each decision attribute is formulated as a claim, where the degree of validity of the claim (attribute's assessment) is expressed using different maturity levels. The GSN representation of the decision model is transformed into a confidence network to provide evidence-based quantitative maturity assessment using the Bayesian network. A metric based on expected utility of maturity levels, called expected

distance metric, is proposed to measure the distance between target maturity and achieved maturity on a scale of 0 to 1. Expected distance metric helps in comparing the assessment of different attributes and identification of major areas of concern in terms of modeling capability, data needs, and quality of assessment process. Practical application of the framework is demonstrated by two case studies. The first case study is focused on validation assessment of a thermal-hydraulic code for a challenge problem called Departure from Nucleate Boiling (DNB). The second case study is focused on assessment of multiphysics codes for another challenge problem called CRUD Induced Power Shift (CIPS).

A Framework for Predictive Capability Maturity Assessment of Simulation Codes

by
Paridhi Athe

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Nuclear Engineering

Raleigh, North Carolina

2018

APPROVED BY:

Dr. Nam Dinh Committee Chair	Dr. Abhinav Gupta
Dr. David Kropaczek	Dr. Maria Avramova
Dr. Robert W. Youngblood	

DEDICATION

It is the sunrise, filling the sky with light and beautiful colors,

It is the little bird, lovingly feeding its young ones,

It is the cute squirrel, jumping and eating all around,

It is the dandelion seed, flying across the endless blues,

It is the tiny sapling, emerging out of a crevice,

It is life, blooming and beating in every corner of the world,

So mysterious, yet so amazing!

I dedicate this thesis to life, the greatest teacher of all.

BIOGRAPHY

Paridhi Athe was born in Raisen, a small town in the central part of India, to Manlata Athe and Prakash Athe. She has two elder siblings. She is very close to her family. Her parents, brother Pratik Athe, and sister Pallavi Athe have always been a source of inspiration and strength to her.

She completed her undergraduate degree in Electronics and Telecommunication engineering from Shri Govindram Seksaria Institute of Technology and Science in 2007. She got a master's degree in Electronics and Communication Engineering with specialization in Semi-conductor devices and VLSI Technology from Indian Institute of Technology Roorkee in 2009. Paridhi got another master's degree in Nuclear Engineering from Indian Institute of Technology Kanpur in 2011. Paridhi started her doctoral study in Nuclear Engineering at North Carolina State University in Fall 2012. She started working with Dr. Nam Dinh as her Ph.D. advisor in Fall 2014. Her current research is focused on the development of framework for predictive capability maturity assessment of computational tools.

ACKNOWLEDGMENTS

I would like to express my gratitude to my thesis supervisor, Dr. Nam Dinh, for his erudite guidance throughout my thesis work. His analytical and methodical way of working has inspired me, and under his guidance, I have learned a lot. The quantum of knowledge I gained during his inspiring guidance would be immensely beneficial for my future endeavors. Next, I want to thank Dr. Robert W. Youngblood, Dr. Abhinav Gupta, Dr. David Kropaczek and Dr. Maria Avramova for agreeing to be my committee member and devoting their time to review my thesis.

I would like to acknowledge the US Department of Energy via the Consortium for Advanced Simulation of Light Water Reactors (CASL) and NCSU college of Engineering for supporting this research. I am also thankful to the developers of Assurance and Safety Case Environment (ASCE 4.2) and AgenaRisk software that has been used in this research.

I appreciate Dr. Hany Abdel-Khalik for his guidance and support during the initial phase of my Ph.D. I am also thankful to Dr. Igor Bolotnov, Dr. John Mattingly, Dr. Joseph Michael Doster, Dr. Dmitriy Anistratov, Dr. Steve Shannon, and Dr. Carl Meyer. Their courses have been a great learning experience for me. I am grateful to Dr. Prabhat Munshi for his encouragement and advice. Next, I want to thank my colleagues, Bassam Khuwaileh, Mohammad Abdo, Young Suk Bank, Jason Hite, Chih-Wei Chang, Yang Liu, Linyu Lin, Botros Hanna, Han Bao, Joomyung Lee, Yangmo Zhu, their comments and feedback on my research were very helpful. I want to thank Shefali Saxena, Arka Datta, Anand Vardhan Mishra, Rudrodip Majumdar, Yuwei Zhu, Jun Fang, Baljinder Kaur, Suman Dhayal, Vandana Dhayal, Nilesh Kumar, Anant Raj and Prerna Prateek, for the wonderful time I spent with them during my stay in Raleigh.

Finally, I want to thank my parents, my brother, and my sister for their unconditional love, support, and encouragement through every endeavor of my life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

AHP	Analytic Hierarchy Process
AIAA	American Institute of Aeronautics and Astronautics
ASCE	Assurance and Safety Case Environment
BC	Boiling Condition
CAE	Claim Argument and Evidence
CAS	Credibility Assessment Scale
CASL	Consortium for Advanced Simulation of Light Water Reactors
CDF	Cumulative Density Function
CF	Cross Flow
CFD	Computational Fluid Dynamics
CHF	Critical Heat Flux
COBRA-TF/CTF	Coolant-Boiling in Rod Arrays-Two Fluids code
CP	Challenge Problem
CSAU	Code Scaling, Applicability, and Uncertainty
CVER	Specifier for Code Verification in PCMM
DBA	Design Basis Accident
DNB	Departure from Nucleate Boiling
DCA	Data Coverage Assessment
DGA	Deterministic and Graphical Assessment
DP	Data pyramid (experiments-based)
DRA	Data Relevance Assessment
DUA	Data Uncertainty Assessment
DVA	Direct Validation Attribute
EAP	Evidence Assessment Process
EM	Evaluation model
EMDAP	Evaluation Model Development and Assessment Process
FOM	Figure of Merit
FP	Fuel Performance
FR	Flow Regime
FRM	Flow Redistribution Mechanism
GSN	Goal Structuring Notation
HLC	High Level Composition
IA	Initial-author Assessment
IAEA	International Atomic Energy Agency
ICME	Integrated Computational Materials Engineering
IET	Integral Effect Test
LBLOCA	Large Break Loss of Coolant Accident
LLC	Low Level Composition
LOCA	Loss of Coolant Accident
LWR	Light Water Reactor
MET	Mixed Effect Test
NA	Not Available
NASA	National Aeronautics and Space Administration
NB	Nucleate Boiling

NC	Natural Circulation
NPP	Nuclear Power Plant
NRC	Nuclear Regulatory Commission
PA	Peer-reviewed Assess
PCMM	Predictive Capability Maturity Model
PD	Pressure Drop
PE	Point Estimate
PIRT	Phenomena Identification and Ranking Table
PM	Model pyramid (code-based)
PMMF	Physics and Material Model Fidelity (a PCMM attribute)
PMO	Plant Measurement and Observations
PNNL	Pacific Northwest National Laboratory
PP	Phenomenology Pyramid
PQA	Process Quality Assurance
PSA	Physics Scaling Assessment
PVM	Probabilistic Validation Metric
QOI	Quantity of Interest
QPIRT	Quantitative Phenomena Identification and Ranking Process
RGF	Representation and Geometric Fidelity (a PCMM attribute)
SA	Specialist-author Assessment
SBLOCA	Small Break Loss of Coolant Accident
SD	Scale distortion
SDG	Selective Dimensionless Groups
SET	Separate Effect Test
SM	Scaling Methodology
SME	Subject Matter Expert
SPF	Single Phase Flow
SQA	Software Quality Assurance
SRQ	System Response Quantity
SVER	Specifier for Solution Verification in PCMM
TBA	To Be Assessed
TH	Thermal-Hydraulics
TM	Turbulent Mixing
TPF	Two Phase Flow
UQ	Uncertainty Quantification
US	United States of America
V&V	Verification and Validation
VAL	Specifier for Model Validation in PCMM
VEA	Validation Evidence Assessment
VP	Validation Pyramid
VTR	Validation Test Results
VVUQ	Verification, Validation and Uncertainty Quantification

CHAPTER 1: INTRODUCTION

1.1. Motivation

Over the past 30 years, the role of M & S tools in the decisions related to design, operation and safety assessment of nuclear power plants have increased at an astounding rate. Computational tools are widely used in nuclear engineering to quantify and characterize the safety margins, perform hazard and fault analysis, and improve the performance of nuclear reactors. Therefore, comprehensive methodologies and systematic processes have been developed, adopted, and applied to guide the development of M & S tools and assess their adequacy for applications in nuclear reactor design, operation or safety analysis.

The design of an M & S tool involves three major phases: complexity resolution, model formulation, and numerical simulation. Each phase of development of the computational tool can be associated with different sources of uncertainty (aleatoric and epistemic). These uncertainties directly impact the code's prediction of the system response quantity of interest. Verification, validation, and uncertainty quantification (VVUQ) are three key processes that help in assessing the reliability of the code prediction for an intended application. Different field of science and engineering developed their own procedures and guidelines for V &V of M & S tools. Remarkable are the Code Scaling, Applicability, and Uncertainty (CSAU) methodology [1], and the Evaluation Model Development and Assessment Process (EMDAP) [2] developed by the U.S. Nuclear Regulatory Commission (NRC). In parallel, nuclear security, defense, and aerospace communities also made efforts in the development of assessment methodology (e.g., Quantification of Margins and Uncertainties, QMU; Predictive Capability Maturity Model, PCMM [3]; NASA Standard for Models and Simulations [4], AIAA CFD V&V Guide [5], etc.) Depending on the intended application area, the implementation of verification validation and uncertainty quantification

(VVUQ) process have some differences in these methodologies. However, the philosophy of all the standards and methodologies is inherently the same.

Based on the nature of information, the reliability of a computational tool for a given decision problem is governed by two set of information or data: Subjective information and objective information. Subjective information includes phenomena identification and ranking process, experiment's relevance information and expert's confidence in the quality of experiment and code simulation. Objective information is based on experiments (available for validation), models and code simulation, and their uncertainty and sensitivity information (see Figure 1.1 for basic illustration).

Subjectivism is eminent in the development and reliability assessment of M & S tools because of the approximate nature of model and data. The situation further complicates due to complex multiscale and multi-physics interactions in nuclear reactor systems. Therefore, the central question of concern in all methodologies is the "adequacy decision" or "fitness of purpose" of the M & S tools. Even though comprehensive methodologies and assessment procedures have been developed to guide the assessment of computational tools, in the end, the "adequacy decision" is still left to engineering judgment. This heuristic approach to adequacy assessment often turns code licensing into an elongated process of extensive review and scrutiny. These challenges and difficulties motivate us to develop a systematic, formalized and computerized framework that can assist the current assessment methodology (PCMM) in the decision regarding the adequacy of a simulation tool for a given nuclear reactor engineering or safety application.

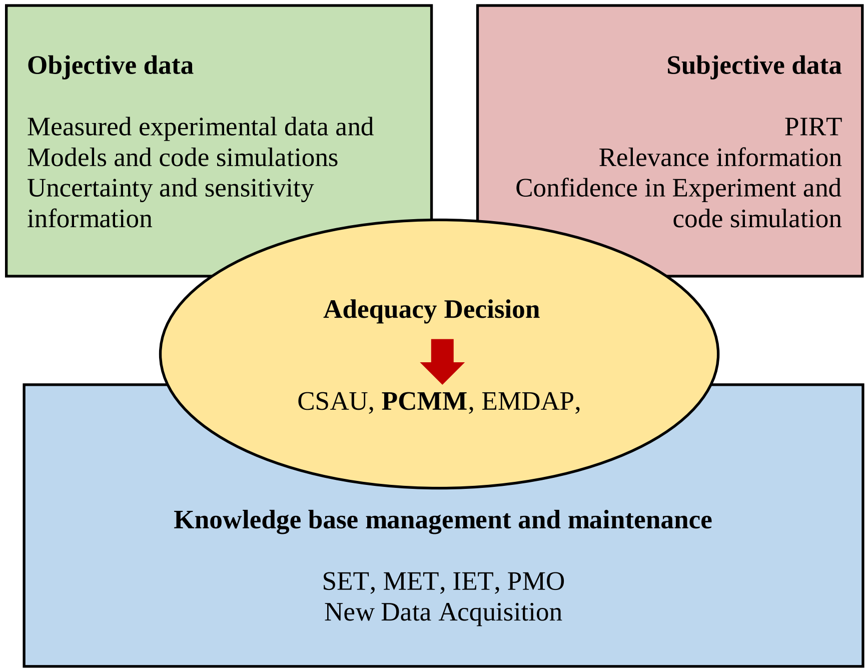


Figure 1.1: Illustration of code adequacy assessment

1.2. Dissertation overview

1.2.1. Objectives

Code V& V (for nuclear reactor applications) can be described as a confidence-building process. It is an iterative process that requires continuous exploration, learning, and assessment. A successful VVUQ process should address all sources of uncertainty and provide sufficient evidence for reliable and robust decision making. The target of the proposed work is to formalize the maturity assessment process and support the implementation of PCMM by providing a framework for structural knowledge representation, evidence incorporation, and maturity quantification. The principal objectives of the proposed research are as follows:

- Facilitate structural knowledge representation, information abstraction, and integration for maturity assessment of a code.

- Provide support for structural organization, classification, and characterization of evidence for code's maturity assessment (primary focus→ code validation assessment).
- Create a formal decision model for code maturity assessment (primary focus→ code validation assessment).
- Facilitate confidence assessment, estimation, and sensitivity analysis using the decision model.
- Complete the framework for predictive capability maturity assessment (PCMA).
- Demonstrate the use of the proposed framework for the maturity assessment of a computational tool (CTF) for an intended application (CASL challenge problems-DNB).

1.2.2. Technical approach

This section describes the technical approach adopted to develop the framework for predictive capability maturity assessment. The key points encompassing the technical approach are described below:

- The current standard and methodologies for credibility assessment in Nuclear engineering, i.e. CSAU [1], EMDAP [2] and PCMM [3], are used to guide the formulation of the proposed framework.
- Structural knowledge representation in the framework is obtained using an argument modeling technique called Goal structuring notation (GSN) [6]. The PIRT-based phenomenology pyramid is used to guide the classification and characterization of evidence for code validation assessment. The Pyramid is constructed using the GSN.
- The decision schema in the proposed framework is based on the PCMM [3] and the Analytical hierarch process (AHP) [7]. The hierarchical decision model is constructed

using the GSN. The number of levels in the hierarchy depends upon the required depth and rigor of the analysis. Each attribute and sub-attribute in the decision model is formulated as a claim (i.e. Goals nodes in the GSN tree) where the degree of validity of the claim is defined by different maturity levels. Evidence are integrated across the lower level attribute in the decision model (using the solution nodes in the GSN tree).

- The GSN based decision model is transformed into a confidence network (Bayesian network) for quantitative maturity assessment. Bayesian network enable abstraction of maturity information from lower level attribute to higher level attributes. It helps in assessing the maturity based on the quality of evidence integrated in the decision model. Subjective data based on the expert opinion is incorporated in to the decision model using condition probability table (CPT) and subjective probabilities based on the criteria of evaluation of the evidence.
- A metric based on the expected utility of the maturity levels is proposed to evaluate the distance between the target level and achieved level of maturity on a scale of 0 to 1 for each attribute and sub-attribute in the decision model.

1.2.3. Dissertation structure

The current chapter describes the motivation, objectives and technical approach of the proposed research. It also provides a glossary of important terminologies used in this thesis. The organization of the rest of the thesis is as follows.

Chapter 2 of the dissertation documents a comprehensive review of different topics that provide necessary background and foundation for the development of the assessment framework. This chapter consists of eight sections. The first section focuses on complexity resolution and the PIRT process. The second section provides significant developments and comparison of different standards for credibility assessment in the M & S tools. The third section discusses the decision process and decision analysis. The fourth section provides a brief overview of CASL codes and activities. The fifth section gives an overview of the scaling techniques and describes the importance of scaling in model development and data applicability analysis. The sixth section is focused on safety case, argumentation and use of Goal Structuring Notation (GSN) in safety case representation. The seventh section provides a review and brief illustration of different techniques, like Fuzzy logic [8], Bayesian networks [9], and Evidential reasoning [10] that can be employed for quantitative maturity assessment. The last section in this chapter illustrate the current techniques used for transforming GSN into a computable network.

The first section of Chapter 3 provides an overview of the process of code development, verification, and validation. The second section illustrates the research approach for formalizing the maturity assessment process in the proposed framework. The third section of this chapter is devoted to the formulation of the proposed framework. It provides explanation and simple illustration of each element of the framework. The framework consists of different elements including structured knowledge representation, evidence classification and characterization, and quantitative maturity assessment. As the primary focus of the framework is code validation, all the elements of the framework are illustrated from the perspective of code validation assessment.

Chapter 4 and chapter 5 provide case studies to illustrate the application of the proposed framework. The case study in chapter 4 is based on validation assessment of CTF for a CASL challenge problem called Departure from Nucleate Boiling (DNB). The case study in chapter 5 is based on the assessment of multiphysics CASL codes for another CASL challenge problem called CRUD Induced Power Shift (CIPS).

Chapter 6 provides the analysis of the proposed framework based on the different sources of uncertainty and sensitivity analysis of nodes in the decision model.

Chapter 7 provides conclusion and recommendation for future work.

1.3. Glossary

This section presents a list of important definition or terminologies that are frequently used in this dissertation:

- Aleatory uncertainty: Uncertainty attributed to the inherent randomness in the system parameters. It is irreducible in nature (or stochastic) and characterized by statistical distribution or probability density function.
- Benchmarking: It is also part of software quality check. Benchmarking is performed by code-to-code comparison. It involves comparison of simulation of an identical problem on different simulation codes.
- Bottom-up approach: The process of combining the smaller block of the system, starting from the base element to form components and subsystem until the complete representation of the system is obtained.
- Code Verification: The process of agglomerating the evidence to evaluate the assertion (or claim) that the numerical algorithms are implemented correctly inside the code [11].Code verification is focused on,
    - Debugging the source code
    - Eliminating errors in the numerical algorithm.
- Epistemic uncertainty: Uncertainty attributed to the lack of knowledge about the system, e.g., uncertainty due to incomplete knowledge about the physical processes or phenomena and

model form (i.e., model form uncertainty). It is characterized by subjective probabilities, or interval estimation (min, max).

- Figure of merit: The figure of merit are those quantitative standards of acceptance that are used to define acceptable answer for the safety analysis or performance evaluation of the specified nuclear reactor safety or engineering application using the M & S tool [2].
- Goal Structuring Notation (GSN): The Goal Structuring notation is an argument modeling technique. It is used for graphical representation of assurance arguments in the safety case. It was developed by Kelly [6].
- Integral Effect Test (IET): Integral Effect Tests are experiments that involve measurement of integral parameters that encompass the effect of multiple phenomena affecting the system behaviour.
- Model form uncertainty: Uncertainty associated with the choice of a suitable (closest to reality) model from a set of candidate models for emulating a physical quantity.
- Model parameter uncertainty: Model parameter uncertainty is the model uncertainty that arises due to uncertainty in the values of the model parameters, It could be aleatory or epistemic. For example, parameters that can be calibrated to experimental data (e.g., closure parameters) would be considered epistemic, while manufacturing uncertainties would be considered aleatory.
- Phenomena/complexity resolution: The process of resolving the complexity of the system by segregating the relevant phenomena or processes that happen in the system during a specified transient or steady state scenario.

- Regression testing: It is a type of software quality check (part of Verification) which verifies that the code did not' underwent any unintended change due to any modification in the source code.
- Safety case: The U.K. Defense Standard 00-56 describe the safety case as, "*a structured argument, supported by a body of evidence that provides a compelling, comprehensive and valid case that a system is safe for a given application in a given environment*" [12].
- Scaling Analysis: The process of assessing the similarity between the reduced scale test facility and the full-scale nuclear reactor application.
- Scale distortion (SD): Scale distortion can be described as the inefficiency in reproducing the full-scale reactor level phenomena and process in the reduced scale test facility. Different methodologies have been adopted to quantify the scaling distortion. The classical similarity-theoretic method evaluates the scale distortion by comparison of dimensionless scaling groups [13] at the referenced plant level and the scaled experiment level. In the case of dynamic processes or transients, scaling distortion is obtained by the ratios of time [14] or effect metrics [15] of the dominant process in the reactor application and the scaled experiment.
- Separate Effect Test (SET): Separate Effect Tests are simple experiments involving measurement of local phenomenon influencing the behaviour of the system.
- Solution verification: The process of agglomerating the evidence to evaluate the assertion (or claim) that the solution to the mathematical functions represented in the simulation is correct (or correct enough) when compared with the true solution of those same functions [11]. Generally, comparison to the analytical solution is considered part of code verification. Solution verification evaluates mesh convergence, but the solution is not known analytically.

- Top-down approach: Decomposition of a complex system into sub-systems, components and so on, until the base elements at the bottom level are determined.
- Uncertainty quantification: The process of agglomerating the evidence that supports the assumption that the statistical variability in the system response quantities (SRQs) of interest due to variation in the input quantities has been adequately captured [11].
- Unit Testing: Units test are part of software quality check (part of Verification). They involve simple test problems to check if small parts or units of the code are working correctly.
- Validation: The process of agglomerating the evidence to evaluate the assertion (or claim) that the numerical simulation of the mathematical function can predict a real physical quantity[11].

CHAPTER 2: TECHNICAL COMPONENTS BACKGROUND

2.1. Introduction

This chapter presents review and perspective on different topics that provide necessary background and foundation for the development of the proposed work. The section is divided into eight parts:

- Complexity resolution
- Standards and methodologies for credibility assessment of M & S tools
- Decision analysis and decision process
- CASL M & S activities
- Scaling techniques
- Safety case, and argumentation
- Candidate tools/techniques for maturity quantification
- Transforming GSN to computable network

2.2. Complexity resolution

Complexity is eminent in nature everywhere. Analysis and understanding of a complex system require segregation of system into less intricate parts or sub-systems with distinctive form or characteristic. Herbert A. Simon in his classic paper on "*the architecture of complexity*" describes how different complex systems exhibit hierarchical structure and similar properties (in the context of architecture or structural organization) regardless of their specific content [16]. He explains that two types of interactions are eminent in a hierarchical system: (1) Interactions within subsystems (or inter-component linkage), (2) Interactions among subsystem (or intra-component linkage). It is the nature of these interactions that guide the decomposition of a complex system. A complex system can be considered nearly decomposable when interactions among subsystem are feeble in strength compared to interactions within subsystem [16].

Phenomenon Identification and Ranking Process is a crucial technique to resolve complexity in the modeling and simulation of complex nuclear reactor applications. It was introduced as part of the Code Scaling, Applicability, and Uncertainty (CSAU) methodology in 1988. Over the past thirty years, it has been successfully applied to resolve several issues, like LBLOCA [1], SBLOCA [17], nuclear power plant fire modeling[18], analysis of CASL challenge problems [19, 20] and design of next-generation nuclear power plants [21]. The PIRT process is based on the subjective data (expert knowledge) and created by joint consensus of a panel of experts having broad understanding and knowledge of the underlying physical processes governing the problem of interest. It involves identification and ranking of different phenomena relevant to the figure of merit [2].

Simon describes two types of descriptors that can be used for solving a problem involving a complex system. These descriptors are called state descriptor and process descriptor. A *state descriptor* provides criteria for identifying an object or state of the system while the *process descriptors* are related to different processes or actions that lead to that particular state of the system. He further explains, "*We pose a problem by giving state description of the solution. The task is to discover a sequence of processes that will produce the goal state from an initial state*" [16]. In the context of the PIRT, the figure of merit (FOM) may be considered as a state descriptor while different phenomena/processes that impact the FOM may be considered as process descriptors. Understanding the sequence and relation of different phenomena becomes crucial for successful formulation of the problem. Structure of PIRT is governed by the nature of the problem being analyzed. The PIRT for accident situations like loss of coolant accidents (LOCA), resolves complexity by dividing the transient scenario into time phases (blowdown, refill and reflood) based on the dominant mechanism or some other factors (operators action or valve opening and closing). The phenomena identified by the PIRT process are arranged hierarchically based on transient phase, system components, and underlying phenomena. The PIRT for simulation of high fidelity CASL challenge problems involves system decomposition with respect to governing physics (Neutronics, Fuel performance, Coolant chemistry and thermal hydraulics) and scale (micro-scale, meso-scale and macro-scale) of the underlying phenomena. Hence, scale separation and physics decoupling are the two elementary principles that guide complexity resolution for CASL Challenge problems. The outcome of PIRT process is governed by the experts' knowledge and understanding about the problem of interest. Therefore, PIRT is subject to large epistemic uncertainty.

In recent years, objective approaches based on scaling analysis like Hierarchical Two-Tiered Scaling (H2TS) and Fractional Scaling Analysis (FSA) have been used to construct Quantitative Phenomena Identification and Ranking Table (QPIRT). Sensitivity analysis of response quantity of interest with respect to the relevant input parameters or boundary and initial condition is also used for creating a QPIRT. Although these approaches sound more robust and efficient, a QPIRT is completely based on the mathematical model of the problem of interest. Therefore, it cannot be directly applied to cases where the mathematical model is inexistent or under-developed with respect to the intended application. For such cases, traditional PIRT is employed for the conception of governing mechanism and underlying physical processes (complexity resolution), guiding model development, identification of issues and data needs. In this way, PIRT helps in prioritizing the research and development needs for any nuclear reactor application. Major steps related to the PIRT process are shown below [22]:

- Define the problem and PIRT objectives.
- Specify the scenario (transient or steady state). In the case of a transient process, the scenario is partitioned into time phases based on the dominant process/mechanism.
- Identify and define the figure of merit (FOM).
- Identify and review all the relevant literature (experimental and analytical data).
- Identify phenomena relevant to the FOM.
- Rank all the Phenomena based on knowledge and importance (with respect to the FOM).
- Document all the findings.

The "Phenomena" in the PIRT process is treated as a general entity and can be anything that impacts the FOM. It equivocally includes mathematical or engineering approximations, system conditions, physical processes, reactor parameters as phenomena in the PIRT process [23]. Such simplification affects proper structuring of information. Therefore, a systematic approach is required to formalize the PIRT process where we can clearly state the objective, assumptions, strategy for complexity resolution and specify the theoretical and experimental evidence that forms the basis of the expert input to the PIRT.

"*Human mind is not capable of considering all the factors and their effects simultaneously*"[24]. Therefore, the organization of knowledge and information in a proper structure becomes essential to render our ability to make a rational decision in a scenario of uncertainty and lack of information. Hierarchical structure provides organization of information in order of relevance/ importance to the quantity of interest. Such decomposition facilitates the solution to complex problems involving multiple criteria decision-making situations. Saaty introduced "*The Analytic Hierarchy Process*" (AHP) [7] to provide techniques for identifying the relevant information and their interrelationship in a complex problem. He emphasizes that "*conception of reality is crucial*" and hierarchical decomposition is valuable in the analysis of problems where "*subjective, abstract or nonquantifiable criteria*" are eminent in the decision [24]. PCMM is another example that illustrates the use of hierarchical approach for resolving decision concerning the reliability of a modeling and simulation tool for an intended application [25] [11]. Hierarchical decomposition is also eminent in the code validation process in the form of validation pyramid. Due to lack of data for validation, the experimental data is organized hierarchically in order of increasing relevance and complexity with respect to the application of interest. The concept of validation pyramid was introduced by AIAA V & V guide for CFD simulation concerning aerospace applications [5]. CASL extended the application of validation pyramid to multiphysics and multiscale challenge problem by adopting component identification and ranking process[26].

2.3. Standards and methodologies for credibility assessment of M & S tools

This section presents a review of different methodologies/standards that have been developed to provide systematic procedures and standard guidelines for comprehensive adequacy assessment of a computational code for an intended use. Although procedures and guidelines in each methodology have some differences depending on the features of M & S tool and their intended use, in the core all the methodologies address similar issues related to verification, validation and uncertainty quantification (VVUQ) of the computational codes. We will start our discussion with the Code Scaling, Applicability, and Uncertainty (CSAU) evaluation methodology

[1] that was developed by the U. S. Nuclear Regulatory Commission in the late 1980's for reliable estimation of reactor safety margins.

2.3.1. Code Scaling, Applicability, and Uncertainty (CSAU) evaluation methodology

In 1998 US NRC introduced the revised emergency core cooling system rule (ECCS rule, 10CFR50) which consist of limiting values of different response quantities of interest (peak clad temperature, oxidation, hydrogen generation, coolable geometry, long-term cooling) that could be used as safety criteria. These rules were introduced to enable the use of computational tools for safety analysis using best estimate plus uncertainty approach. To assess the reliability of computational results, CSAU methodology was developed. It consists of a set of rules (procedure/guideline) to assess and improve a code's predictive capability and assure low probability of violating the safety criteria.

The CSAU methodology was described by three major elements [27]:

- Requirements and code capability,
- Assessment and ranging of parameters,
- Sensitivity and uncertainty analysis.

The first element involves specification of the scenario to be modeled along with the nuclear power plant type, code specification, and phenomena resolution using the PIRT. The second element of CSAU is focused on validation. It involves identification of relevant separate effect tests (SETs) and integral effect tests (IETs) for validation of code. Based on the SETs and IETs, an assessment matrix is created for validation. CSAU emphasizes the assessment of code's scale-up capability. Code scale-up capability is assessed based on simulation of different reduced scale test facilities. As CSAU focusses on system codes, plant nodalization is considered as a dominant source of uncertainty and scale distortion. The last element of CSAU is focused on sensitivity analysis and uncertainty quantification. This step combines the total bias and uncertainty due to all sources to obtain a quantitative estimate of the plant safety margin for the specified transient in the specified nuclear power plant. CSAU uses the response surface method to estimate the overall uncertainties in the prediction of the FOM. Response surface for the FOM is created by varying all the relevant input parameters within their range of uncertainty. Based on the variability of the FOM a PDF is obtained which gives a measure of total uncertainty in the FOM due to all the parameters. As there are obvious limitations in the code (model uncertainty) and data used for validation, an additional margin is added to compensate for the lack of knowledge and information (epistemic uncertainty).

2.3.2. Evaluation Model Development and Assessment Process

In 2005, U.S. Nuclear Regulatory Commission introduced the Evaluation Model Development and Assessment Process [2] to guide development and assessment of evaluation models (M & S tool) for analysis of transient and accident scenarios that comes within the design basis of a nuclear power plant. The EMDAP process consists of four major elements as shown in Figure 2.1. Although CSAU was primarily developed for safety margin characterization (total uncertainty in prediction), the concepts employed to accomplish this task encompass the entire Evaluation Model Development and Assessment Process (EMDAP). One distinctive feature of the EMDAP is the high focus on hierarchical system analysis and scalability analysis for data and code (Figure 2.1).

EMDAP also emphasizes that “the complexity of the problem should determine the level of detail needed to develop and assess an EM.” [2]. Processes and phenomena that acquire higher rank in the PIRT require higher model fidelity or higher level of detail. EMDAP recommends hierarchical system analysis based on the identification of system, component, phases, geometries, fields, and process that are required to be modeled. EMDAP terminates if the decision regarding the adequacy of the code for the intended application is answered in affirmative. If the decision is negative, the process of assessment and improvement is continuously repeated until the adequacy decision becomes positive. The adequacy decision is taken based on the fulfilment of all the requirements put forward by the EMDAP. Although EMDAP does not specify clear criteria for making the adequacy decision, it does recommend formulation of questions that form the basis for the decision. However, the question regarding the adequacy is still governed by engineering judgment or expert's opinion regarding the fulfillment of all the steps mentioned in the EMDAP. CSAU and EMDAP both highlight the importance of VVUQ in the code licensing process.

Both EMDAP and CSAU are focused on credibility assessment of computer codes used for nuclear reactor safety-related application. A standard for the verification and validation of non-safety-related codes for the nuclear reactor application was developed by American Nuclear Society in 2008 [28].

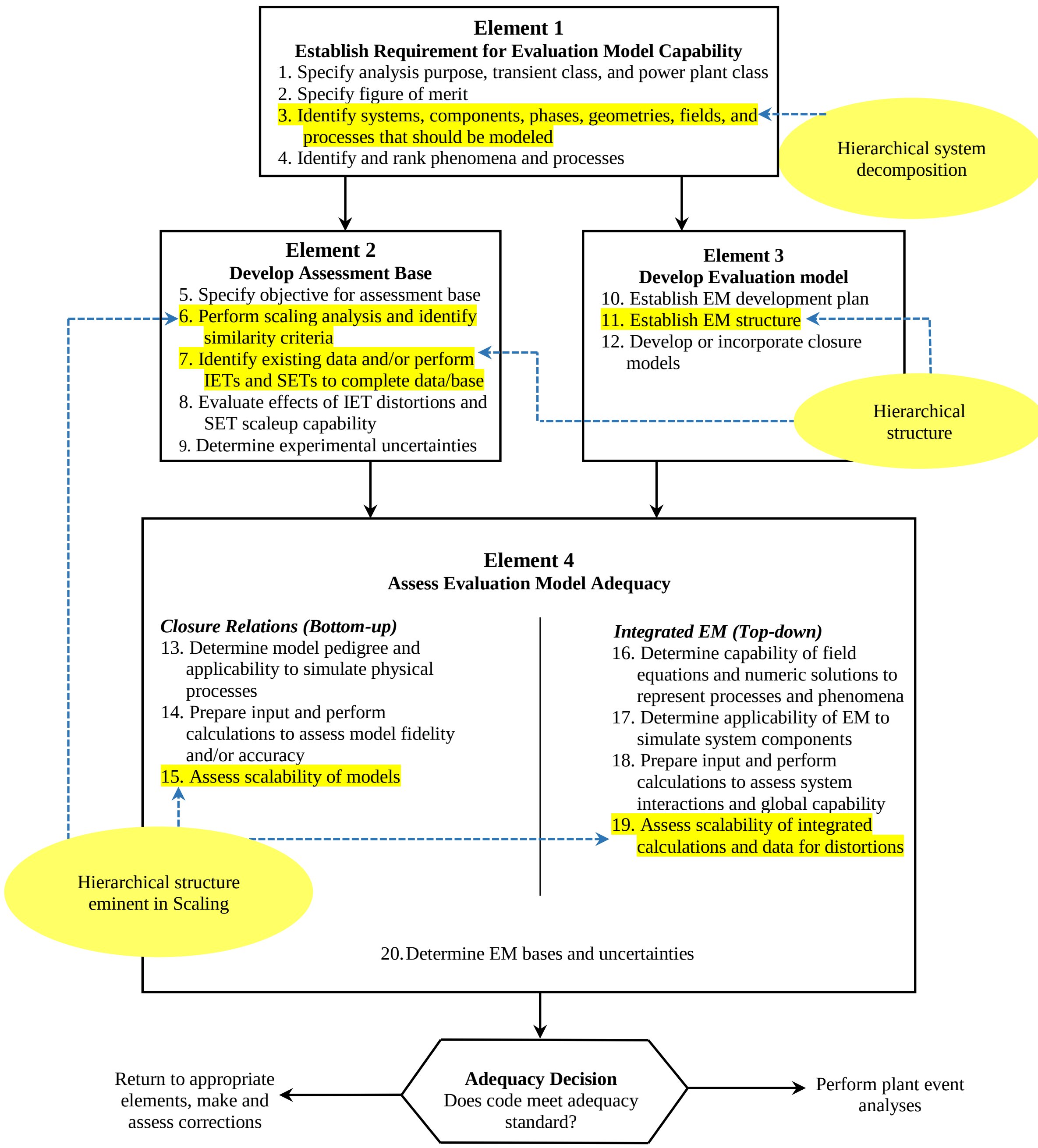


Figure 2.1: Evaluation Model Development and Assessment Process [2]

2.3.3. Contemporary standards for verification and validation of computer codes

Parallel to Nuclear Engineering, U.S. Department of Defense, American Institute of Aeronautics and Astronautics (AIAA), National Aeronautics and Space Administration (NASA) and American Society of Mechanical Engineers (ASME) also developed standards for V & V of

M & S tools. The AIAA V & V guide [5] was focused on CFD simulation concerning aerospace applications. It introduced the concept of Validation Pyramid (VP) (see Figure 2.2) which involves assessment of computational tool using three tiers: Subsystem cases, Benchmark cases, and Unit problems. Complete system lies at the top of the validation pyramid. Complexity and relevance increase as we move up in the validation pyramid while the data for validation becomes scarce. The rationale behind the validation pyramid is to test the M & S tool under different degree of geometric complexity. In the depiction of verification and validation, AIAA guide treats verification and validation separately. ASME guide for V & V in computational solid mechanics [29] provides a more comprehensive view of VVUQ process, depicting verification, validation, and UQ through a single flowchart (see Figure 2.3). At the end of the flowchart, it is determined if an acceptable agreement exists between measurement and code prediction. If not, appropriate changes in models and data are implemented, and the process continues till an acceptable agreement is achieved. Specified accuracy requirement for the SRQs is used as the adequacy criteria.

Assessment of M & S tool in all the methodologies/ standards is focused on VVUQ of the M & S tool. However, NASA's standard for M & S results maturity assessment [4] considers an additional element termed as secondary evidence (see Figure 2.4) in its Credibility Assessment Scale (CAS). It includes use history, M& S management and people qualification as key factors that constitute the secondary evidence in the Credibility Assessment Scale (CAS). The inclusion of secondary evidence provides additional support for assessing the confidence in the M & S tools.

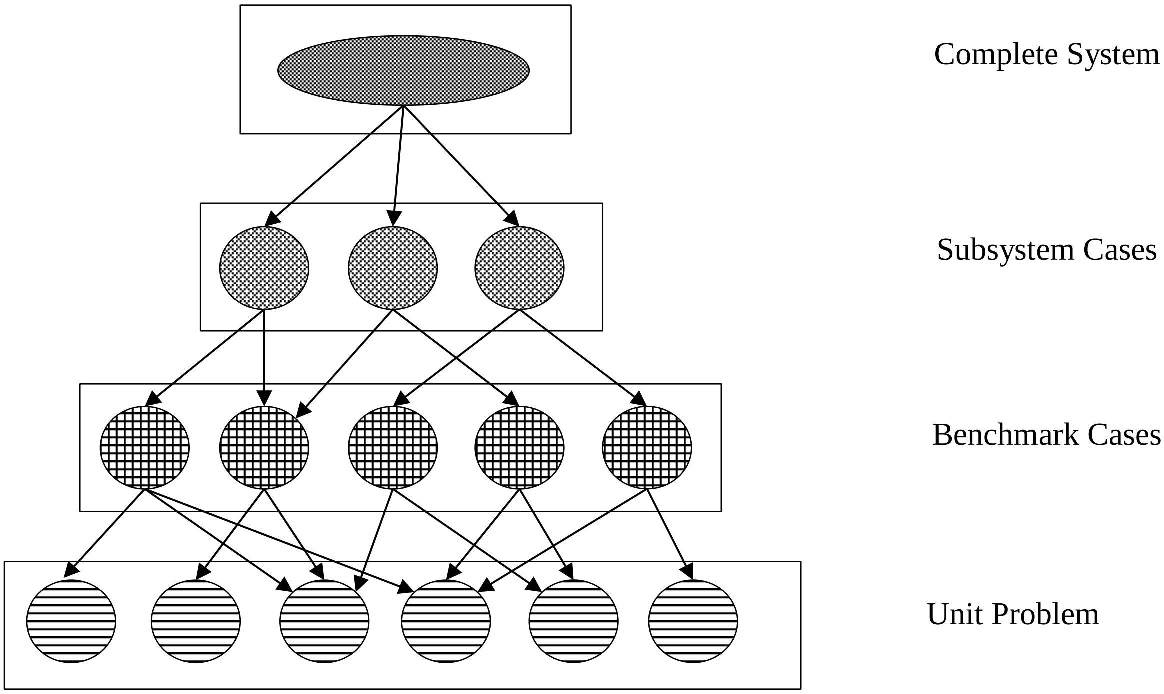


Figure 2.2: AIAA validation pyramid [5]

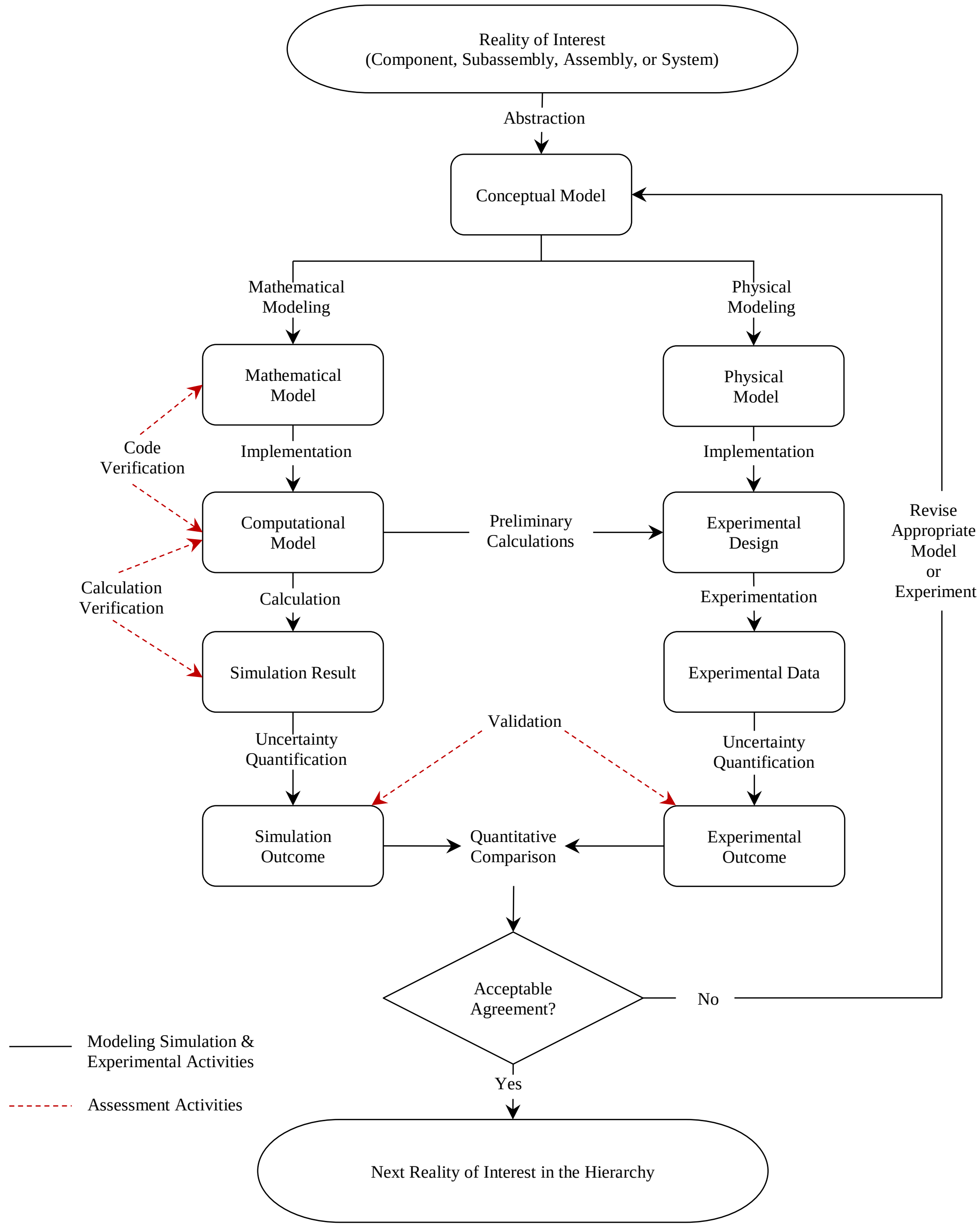


Figure 2.3: ASME Verification and Validation flowchart [29]

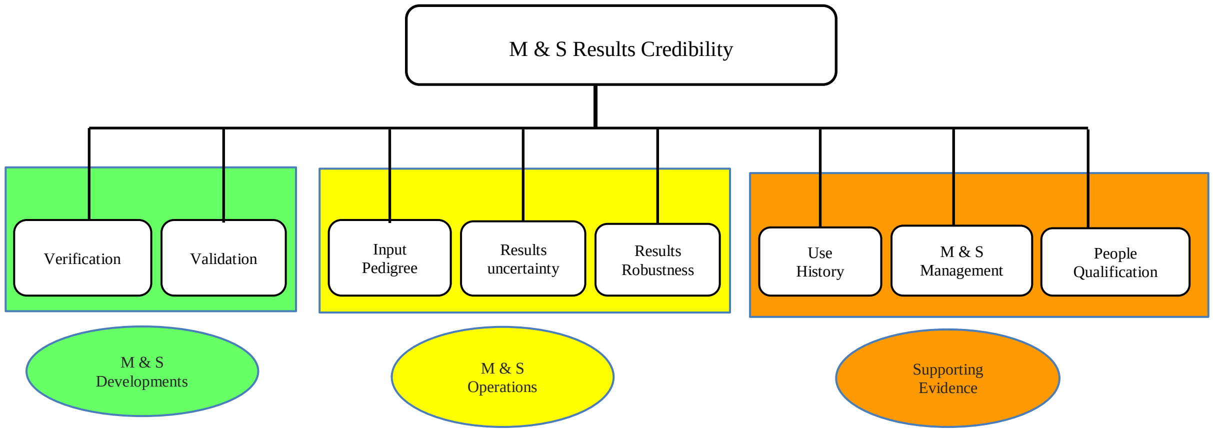


Figure 2.4: M & S results credibility assessment scale used by NASA [4]

The decision of acceptability becomes very important depending on the intended use of the M & S tools. Therefore, all the standards emphasize assessment of M & S tools based on their "intended use" or "fitness for purpose." NASA introduced the M & S Influence-Decision Consequence Risk Matrix (see Figure 2.5) which depicts the M &S results' influence based on the decision consequence. It has three regions marked as red, yellow and green. The region in red indicates the application for which the M & S tool will have the highest impact on the consequence of the decision. Therefore, for these applications, the assessment criteria for the M & S tool must be very stringent. The risk matrix is useful in relating M & S influence with decision consequence. However, it has certain drawbacks [30]:

- It cannot deal with aggregate risk.
- The interaction between risks is not considered.
- Risk matrix does not have the ability to represent uncertainty.
- The tradeoff between likelihood and consequence is fixed.

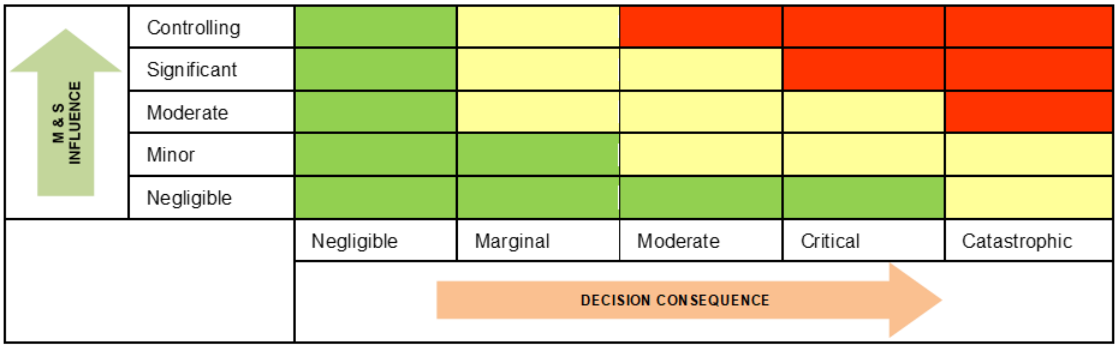


Figure 2.5: M & S Influence-Decision Consequence Risk Matrix used by NASA [30]

2.3.4. Predictive capability maturity model and other maturity assessment methodologies

Originated in nuclear defense applications, the Predictive Capability Maturity Model (PCMM) was developed to assess the maturity of M & S tools based on the decision consequence. In this regard, PCMM can be considered as a decision model for maturity assessment. PCMM was developed by SANDIA national laboratories with the focus on computational simulation concerning nuclear weapon applications. Although PCMM was developed for weapon applications, the elements of PCMM have a broad scope and can be applied to assess the M & S capability for any engineering application. CASL adopted PCMM for assessment of Multiphysics computational tools for different challenge problems related to nuclear reactor operation and safety [20, 31].

The original PCMM matrix consist of six elements (see Table 2.1):

- Representation and geometric fidelity,
- Physics and material model fidelity,
- Code Verification,
- Solution Verification,

- Model Validation,
- Uncertainty quantification and sensitivity analysis.

These elements act as decision attributes and forms the basis for the decision regarding the maturity of a computer simulation code for the intended use. Assessment is performed on the basis of four maturity levels (see Table 2.2). Categorization of each element into these maturity levels is based on the qualitative assessment of constitutive factors that describe that element. In this way, the target level for each element is decided based on the nature of the application of interest. For a high consequence application, more rigorous and stringent assessment criteria are adopted while for low consequence application the assessment criteria are relaxed.

Table 2.3 shows detail PCMM matrix with criteria for assessment of different PCMM element into maturity levels. These criteria provide a qualitative assessment of maturity of the code. PCMM uses the spreadsheet tool and Kiviat (or radar) plots to depict the maturity (see Figure 2.6 and Figure 2.7).

The concept of credibility assessment using maturity level is not new. NASA uses the Technology readiness level (TRL) and credibility assessment level (CAL) (see Table 2.4) while Integrated Computational Materials Engineering (ICME) V & V guide uses the Tool maturity level (TML), both use the concept of maturity level like PCMM. However, TRL is used to express the maturity of technology (material or process) for development of a product (e.g., space shuttle). TML and CAL, like PCMM, were specially developed for assessing the maturity of computational tools.

In 2013 a comprehensive report on fundamentals of scientific commuting was presented by US-NRC [11]. It also emphasizes the use of maturity frameworks like PCMM and NASA maturity assessment framework for credibility assessment of computer simulation tools. It illustrates the use of different maturity assessment set to assess the degree of confidence in verification and validation of a computer simulation.

Table 2.1: Elements of PCMM and maturity levels [3]

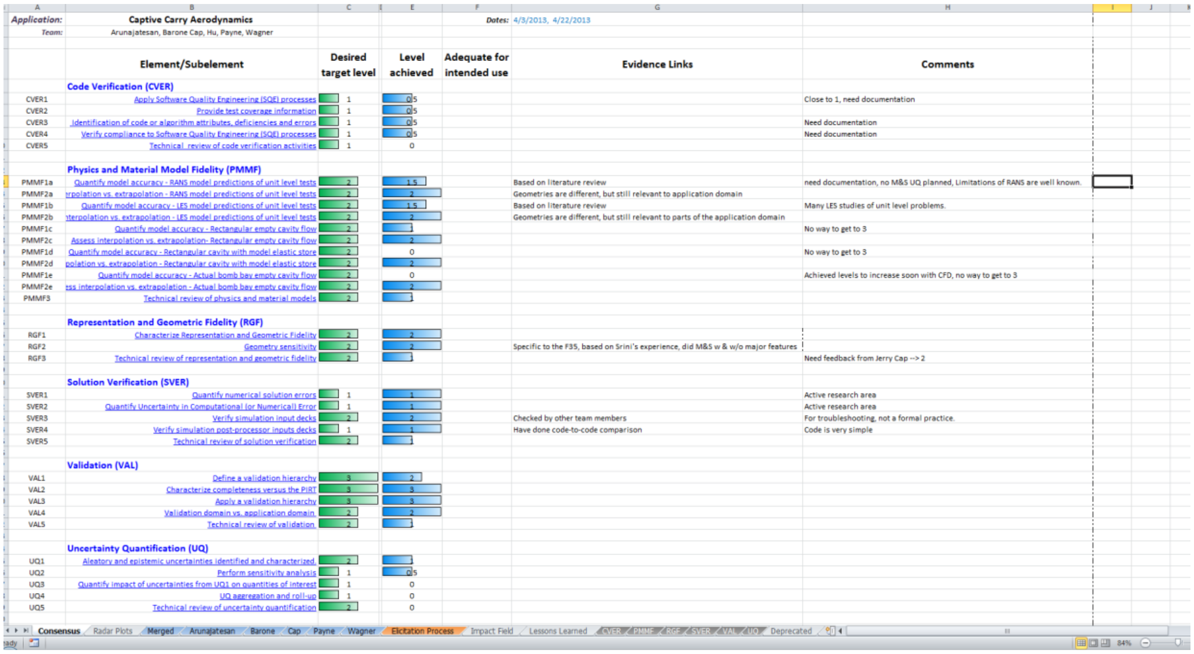
Elements	Maturity Level			
	0	1	2	3
Representation and geometric fidelity (RGF)				
Physics and material model fidelity (PMMF)		MATURITY →		
Code Verification (CVER)				
Solution Verification (SVER)				
Model Validation (VAL)		CONSEQUENCE →		
Uncertainty Quantification and sensitivity analysis (UQSA)				

Table 2.2: Different level of maturity in PCMM as explained by Oberkampf et. al. [3](for detail PCMM matrix see Appendix A)

Level	Description
Level 0	Low consequence, minimum simulation impact, e.g. scoping studies.
Level 1	Moderate Consequence, some simulation impact, e.g. design support.
Level 2	High consequence, high simulation impact, e.g. qualification support.
Level 3	High consequence, decision-making based on simulation, e.g. qualification or certification.

Table 2.3: Detail PCMM matrix [3]

MATURITY / ELEMENT	**Maturity Level 0** Low Consequence, Minimum M & S Impact e.g, Scoping studing	**Maturity Level 1** Moderate Consequence, Some M & S Impact, e.g. Design support	**Maturity Level 2** High-Consequence, High M & S Impact, e.g. Qualifucation support	**Maturity Level 3** High-Consequence, Decision-Making Based on M & S, e.g., Qualification of certification
Representation and Geometric Fidelity What features are neglected becauses of simlifications or stylizations?	• Judgment only • Little or no represenation or geometric fidelity for the system and BCs	• Significant simplification or stylization of the system and BCs • Geometry or representation of major componentss is defined	• Limited simplifictaion or stylization of major components and BCs • Geometry or representation is well defined for major components and some minor components • Some peer review conducted	• Essentially no simplification or stylization of components in the system and BCs • Geometry and representation of all components is at the detail of “as built”, e.g., gaps, meterial interface fasteners • Independent peer review conducted
Physics and Material Model Fidelity How fundamental are the physics and material models and what is the level of model calibration?	• Judgment only • Model forms are either unknown or fully empirical • Few, if any, physics informed models • No coupling of models	• Some models are physics based and are calibrated using data from realated systems • Minimal or adhoc coupling of models	• Physics-based models for all important processes • Significant calibration needed using separate effect tests (SETs) and integral effects tests (IETs) • One way coupling of models • Some peer review conducted	• All models are physics based • Minimal need for calibration using SETs and IETs • Sound physical basis for extrapolation • Full, two-way coupling of models • Independent peer review conducted
Code Verification Are algorithms deficiencies, software errors, and poor SQE practices corrupting the simulation results?	• Judgment only • Minimal testing of any software elements • Little or no SQE procedures specified or followed	• Code is manged by SQE procedures • Unit and regression testing conducted • Some comparisons made with benchmarks	• Some algorithms are tested to determine the observed order of numerical convergence • Some features and capabilities are tested with benchmark solutions • Some peer review conducted	• All important algorithms are tested to determine the observed order of numerical convergence • All important features and capabilities are tested with rigorous benchmark solutions • Independent peer review conducted
Solution Verification Are numerical solution errors and human procedure errors corrupting the simulation results?	• Judgment only • Numerical errors have an unknown or large effect on simulation results	• Numerical effects on relevant SRQs are qualitatively estimated • Input/output (I/O) verified only by the analysts	• Numerical effects are quantitatively estimated to be small on some SRQs • I/O independently verified • Some peer review conducted	• Numerical effects are determined to be small on all important SRqs • Important simulation are independently reproduced • Independent peer review conducted
Model Validation How carefully is the accuracy of the simulation and experimental results assessed at various tiers in a validation hierarchy?	• Judgment only • Few, if any, comparisons with measurements from similar systems or appilcations	• Quantitative assessment of accuracy of SRQs not directly relevant to the application of interest • Large or unknown experimental uncertainties	• Quantitative assessment of predictive accuracy for some key SRQs from IETs and SETs • Experimental uncertainties are well characterized for most SETs, but poorly known for IETs • Some peer review conducted	• Quantitative assessment of predictive accuracy for all important SRQs from IETs and SETs at conditions/geometries directly relevant to the application • Experimental uncertainties are well characterized for all IETs and SETs • Independent peer review conducted
Uncertainty Quantification and Sensitivity Analysis How thorougly are uncertainties and sensitivities characterized and propogated?	• Judgment only • Only deterministic analyses are conducted • Uncertainties and sensitivities are not addressed	• Aleatory and epistemic uncertainties propogated, but without distinction • Informal sensitivity studies conducted • Many strong UQ/SA assumptions made	• Aleatory and epistemic uncertainties segregated, propogated and identified in SRQs • Quantitative sensitivity analyses conducted for most parameters • Numerical propogation errors are estimated and their effect known • Some strong assumptions made • Some peer review conducted	• Aleatory and epistemic uncertainties comprehensively treated and properly interpreted • Comprehensive sensitiviy analyses conducted for parameters and models • Numerical propogation errors are demonstrated to be small • No significant UQ/SA assumptions made • Independent peer review conducted



Application:	Captive Carry Aerodynamics			Dates:	4/3/2013, 4/22/2013	
Team:	Arunajatesan, Barone Cap, Hu, Payne, Wagner					
	Element/Subelement	Desired target level	Level achieved	Adequate for intended use	Evidence Links	Comments
	Code Verification (CVER)					
CVER1	Apply Software Quality Engineering (SQE) processes	1	0.5			Close to 1, need documentation
CVER2	Provide test coverage information	1	0.5			
CVER3	Identification of code or algorithm attributes, deficiencies and errors	1	0.5			Need documentation
CVER4	Verify compliance to Software Quality Engineering (SQE) processes	1	0.5			Need documentation
CVER5	Technical review of code verification activities	1	0			
	Physics and Material Model Fidelity (PMMF)					
PMMF1a	Quantify model accuracy - RANS model predictions of unit level tests	2	1.5		Based on literature review	need documentation, no M&S UQ planned, Limitations of RANS are well known.
PMMF2a	rpolation vs. extrapolation - RANS model predictions of unit level tests	2	2		Geometries are different, but still relevant to application domain	
PMMF1b	Quantify model accuracy - LES model predictions of unit level tests	2	1.5		Based on literature review	Many LES studies of unit level problems.
PMMF2b	nterpolation vs. extrapolation - LES model predictions of unit level tests	2	2		Geometries are different, but still relevant to parts of the application domain	
PMMF1c	Quantify model accuracy - Rectangular empty cavity flow	2	1			No way to get to 3
PMMF2c	Assess interpolation vs. extrapolation- Rectangular empty cavity flow	2	2			
PMMF1d	Quantify model accuracy - Rectangular cavity with model elastic store	2	0			No way to get to 3
PMMF2d	polation vs. extrapolation - Rectangular cavity with model elastic store	2	2			
PMMF1e	Quantify model accuracy - Actual bomb bay empty cavity flow	2	0			Achieved levels to increase soon with CFD, no way to get to 3
PMMF2e	ss interpolation vs. extrapolation - Actual bomb bay empty cavity flow	2	2			
PMMF3	Technical review of physics and material models	2	1			
	Representation and Geometric Fidelity (RGF)					
RGF1	Characterize Representation and Geometric Fidelity	2	2			
RGF2	Geometry sensitivity	2	2		Specific to the F35, based on Srini's experience, did M&S w & w/o major features	
RGF3	Technical review of representation and geometric fidelity	2	1			Need feedback from Jerry Cap --> 2
	Solution Verification (SVER)					
SVER1	Quantify numerical solution errors	1	1			Active research area
SVER2	Quantify Uncertainty in Computational (or Numerical) Error	1	1			Active research area
SVER3	Verify simulation input decks	2	2		Checked by other team members	For troubleshooting, not a formal practice.
SVER4	Verify simulation post-processor inputs decks	1	1		Have done code-to-code comparison	Code is very simple
SVER5	Technical review of solution verification	2	1			
	Validation (VAL)					
VAL1	Define a validation hierarchy	3	2			
VAL2	Characterize completeness versus the PIRT	3	3			
VAL3	Apply a validation hierarchy	3	3			
VAL4	Validation domain vs. application domain	2	2			
VAL5	Technical review of validation	2	1			
	Uncertainty Quantification (UQ)					
UQ1	Aleatory and epistemic uncertainties identified and characterized.	2	1			
UQ2	Perform sensitivity analysis	1	0.5			
UQ3	Quantify impact of uncertainties from UQ1 on quantities of interest	1	0			
UQ4	UQ aggregation and roll-up	1	0			
UQ5	Technical review of uncertainty quantification	2	0			

Figure 2.6: Spreadsheet tool used by PCMM [32]

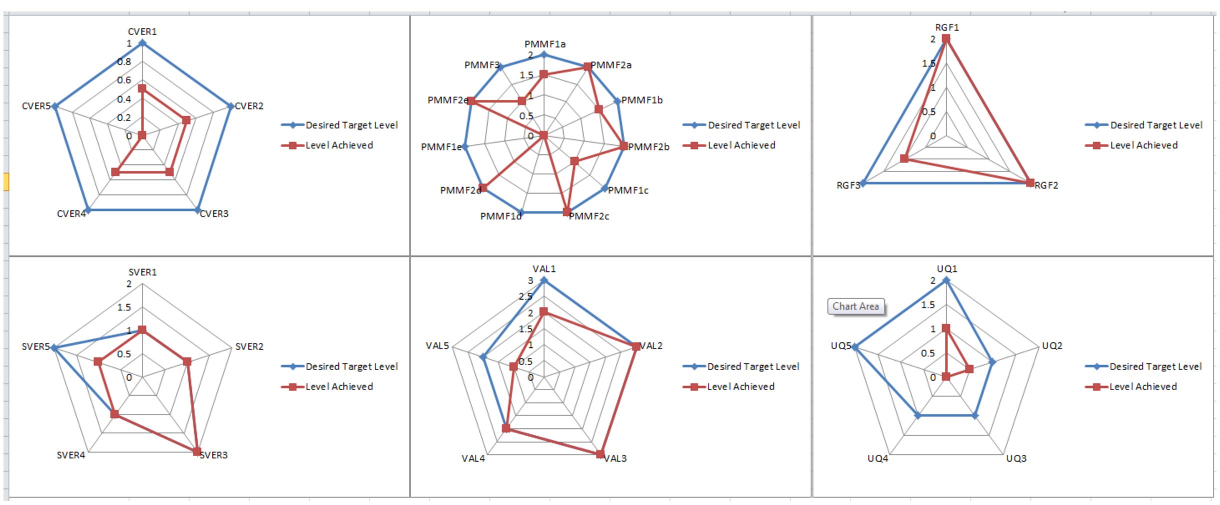


Figure 2.7: Radar plot used in PCMM [32]

Table 2.4: NASA Credibility assessment level [4]

Level	Verification	Validation	Input Pedigree	Results Uncertainty	Results Robustness	Use History	M & S Management	People Qualification
4	Numerical errors small for all important features	Results agree with real-world data	Input data agree with real-world data	Non-deterministic & numerical analysis	Sensitivity known for most parameters; key uncertainties identified	De facto Standard	Continual process improvement	Extensive experience in and use of recommended practices for the particular M & S
3	Formal numerical error estimation	Results agree with experimental data for problem of interest	Input data agree with experimental data for problem of interest	Non-deterministic analysis	Sensitivity known for many parameters	Previous predictions were later validated by mission data	Predictable process	Advanced degree or extensive M & S experience, and recommended practice knowledge
2	Unit and regression testing of key features	Results agree with experimental data or other M & S on unit problem	Input data traceable to formal documentation	Deterministic analysis of expert opinion	Sensitivity known for few parameters	Used before for critical decision	Established process	Formal M & S training and experience, and recommended practice training
1	Conceptual and mathematical models verified	Conceptual and mathematical models agree with simple referents	Input data traceable to informal documentation	Qualitative estimates	Qualitative estimates	Passes simple tests	Managed process	Engineering or science degree
0	Insufficient evidence	Insufficient evidence	Insufficient evidence	Insufficient evidence	Insufficient evidence	Insufficient evidence	Insufficient evidence	Insufficient evidence
	M & S Development		**M & S Operations**			**Supporting Evidence**		

A summary of different standards and methodologies for credibility assessment and their important features/elements is provided in Table 1.

Table 2.5: Comparison of different standards for assessing credibility of M & S tool

Standard	M & S tool	Intended use	Important features/ elements
CSAU	System codes	Characterization of safety margin for DBA	• Phenomena resolution (Specification of scenario, FOM, PIRT process) • Code scale-up analysis • Validation metrics (SETs, IETs) • Combined bias and uncertainty estimation (impact of scale effects, input parameters, nodalization), and Sensitivity analysis • Response surface method for estimation of total uncertainty in FOM
EMDAP	Any evaluation model (no specific code)	Transient and accident scenario in NPP	• Phenomena resolution (Specification of scenario, FOM, PIRT process) • Detailed Hierarchical system analysis for phenomena, model, and data • Requires estimation of IET distortion and SET scale up capability • Model/code scalability analysis • Combined bias and uncertainty estimation (impact of scale effects, input parameters, nodalization), and Sensitivity analysis
PCMM	M & S tools CASL codes	Nuclear Weapon application CASL CPs	• Assessment based on VVUQ, representation and geometric fidelity, physics and material model fidelity • Decision model, evaluation criteria based on application consequence • Spreadsheet tool and radar plot
NASA Credibility Assessment scale	NASA specific M & S	NASA space flight program	• Assessment based on VVUQ and Supporting evidence (past use, M & S management, and people qualification) • Compliance metric (checklist for technical review) and radar plots • M & S Influence-Decision Consequence Risk matrix
AIAA V & V guide	CFD tools	Aerospace Application	• Assessment based on VVUQ • Validation pyramid
ICME V & V guide	CFD tools	Aerospace application	• Assessment based on VVUQ • Decision model for assessing maturity -Tool maturity level (TML) • Use of checklists for assessing different elements of code's verification and validation

2.4. Decision analysis and decision process

The confidence in the code's V&V process is consolidated by following standard procedure and guideline (e.g., EMDAP and CSAU) provided by the code regulation authority. As regulation comes into the picture, the whole process of confidence assessment needs to be explored in the context of decision analysis. Therefore, it is important to understand the decision process. Holztman describes the methodology of using formal methods for decision analysis using a closed loop decision process [33]. The four major stages of decision process are [33],

- Formulation: Create a "formal model" of the given decision. A formal model consists of a network of decision-making elements.
- Evaluation: The next step in the decision process is the Evaluation of the formal model. Evaluation provides recommendation using the formal model based on the decision situation.
- Interpretation/Appraisal: This stage provides interpretation of recommendation provided by the evaluation stage.
- Refinement: Implement changes and observe their implication on the decision model.

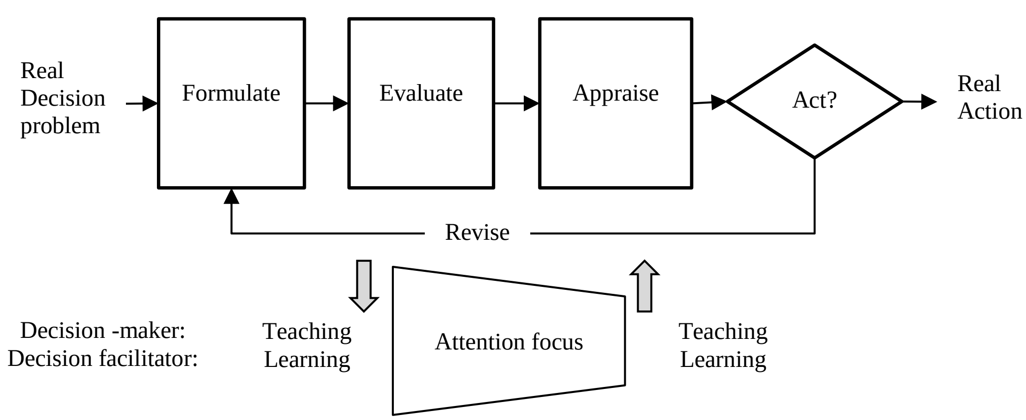


Figure 2.8: A closed loop decision process [33]

As described by Holtzman [33], the closed loop decision process can be illustrated as a "blueprint for conversation" which involves both "decision-facilitator" and "decision-maker." The closed loop decision process explains how there is a continuous exchange of knowledge and information between the two participating entities (i.e., decision-facilitator and the decision-maker) [33]. In the context code's regulation and licensing process, decision-makers could be the code regulation authority while decision-facilitator could be the people associated with decision analysis and other people involved in various activities of code's verification and validation like modeling, experimentation, phenomenon identification, etc.

The hierarchical approach provides an important technique for resolving complex decision-making problems. One of the most popular techniques for decision analysis using the hierarchical approach was presented by Saaty through the Analytic Hierarchy Process (AHP) [34]. The major steps involved in the Analytic Hierarchy Process as presented by Zahedi [35] are shown below [24, 35].

- The decision hierarchy is constructed by breaking down the decision problem into a hierarchy of interrelated decision elements. Major objectives or goal comes at the top of the hierarchy, while subsequent levels in the hierarchy are formed by attributes and sub-attributes that impact the quality of the decision. The bottommost level of the hierarchy is formed by the available alternatives or choices (see Figure 2.9).
- Once the hierarchical structure is complete, pairwise comparisons of the decision elements is performed at all the levels in the hierarchy. If a hierarchical level consists of n elements then the matrix of pairwise comparison is given by,

$$A = \begin{bmatrix} {}^{w_1}/_{w_1} & {}^{w_1}/_{w_2} & {}^{w_1}/_{w_3} & \cdots & {}^{w_1}/_{w_n} \\ {}^{w_2}/_{w_1} & {}^{w_2}/_{w_2} & {}^{w_2}/_{w_3} & \cdots & {}^{w_2}/_{w_n} \\ {}^{w_3}/_{w_1} & {}^{w_3}/_{w_2} & {}^{w_3}/_{w_3} & \cdots & {}^{w_3}/_{w_n} \\ \vdots & \vdots & \vdots & & \vdots \\ {}^{w_n}/_{w_1} & {}^{w_n}/_{w_2} & {}^{w_n}/_{w_3} & \cdots & {}^{w_n}/_{w_n} \end{bmatrix} \qquad (2.1)$$

here, $w_1, w_2, \ldots . w_n$ represent the weights of the n elements. The rank of matrix A is 1 and we have,

$$A.W = n.W \qquad (2.2)$$

where, $W = (w_1, w_2, \ldots . w_n)^T$ is the vector of the actual relative weights. In Eq. (2.2), n represents the eigenvalue and W represents the right eigenvector of matrix A. AHP assumes that the evaluator does not know W, due to which pairwise relative weights of matrix A cannot be determined accurately. Hence, the observed matrix A exhibit inconsistencies. The estimate of W (denoted by $\widehat{W}$) is be obtained by,

$$\hat{A}.\widehat{W} = \lambda_{max}.\widehat{W} \qquad (2.3)$$

here, $\hat{A}$ represents the observed matrix of pairwise comparison, λ_{max} represents the largest eigenvalue of $\hat{A}$ and $\widehat{W}$ represents the right eigenvector. λ_{max} is considered as the estimate of n. The observed values of $\hat{A}$ are more consistent if λ_{max} is closer to n. This consistency can be evaluated by the consistency index (CI) defined by,

$$CI = (\lambda_{max} - n)/(n - 1) \qquad (2.4)$$

Using this "eigenvalue" method the relative weights of decision elements at all the hierarchical levels is obtained.

- The last step in AHP requires synthesis of the relative weights obtained at different levels of the hierarchy to obtain a vector of composite weight that provides ratings for the decision alternatives. The composite relative weight vector of elements at the level k with respect to the level 1 is given by [34, 35],

$$C[1,k] = \prod_{i=2}^{k} B_i \tag{2.5}$$

here, B_i is the n_{i-1} by n_i matrix whose rows consists of estimated $\widehat{W}$ vectors. n_i represents the number of elements at level i.

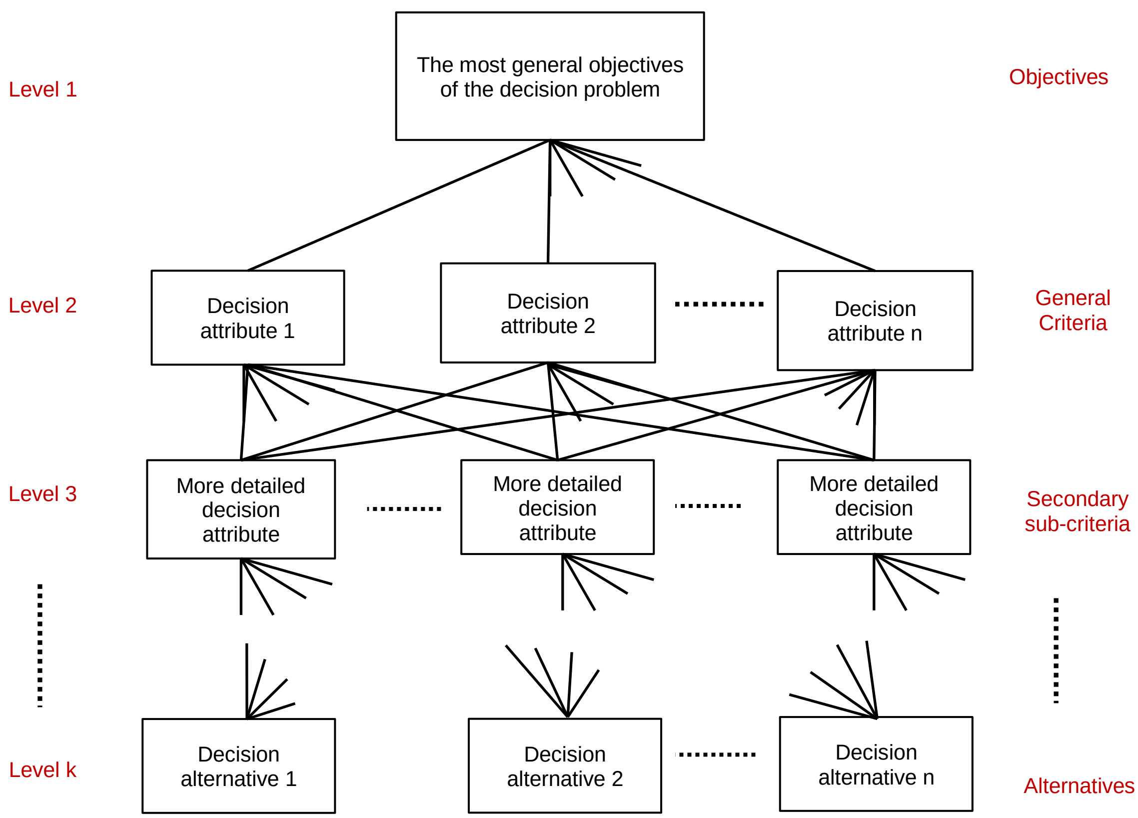


Figure 2.9: Standard form of decision schema in AHP [35]

An alternative means of ranking choices in a decision-making situation is provided by the Utility theory. Utility theory provides the criteria for selecting an alternative based on the uncertainties associated with the monetary consequences (profit or loss) of that alternative. The concept of utility function was originated in economics where it is used to express the preference of a customer in consumption of different goods. In the nuclear regulatory guide (NUREG/CR-6833), the utility is described *"as a figure of merit for a decision alternative that reflects how successfully the decision maker's values and preferences will be addressed by implementing that alternative"* [36]. According to the principle of maximum expected utility, a rational decision-maker should choose alternatives that maximize the expected utility [36]. Expected utility theory is based on three main principles [37]:

- "*Expectation: The maximum expected utility of a prospect is given by:*

  $$U(x_1, p_1; \dots . x_i, p_i; \dots . . x_n, p_n) \quad (2.6)$$
  $$= p_1 u(x_1) + \cdots + p_1 u(x_i) + \cdots + p_n u(p_n)$$

  here, p_i *represents the probability of the* i^{th} *outcome* x_i *and* $u(x_i)$*is the utility of the outcome* x_i.

- *Asset integration:* $(x_1, p_1; \dots . x_i, p_i; \dots . . x_n, p_n)$ *is acceptable at asset position w iff,*

  $$U(w + x_1, p_1; \dots . ; w + x_i, p_i; \dots . . ; w + x_n, p_n) > u(w) \quad (2.7)$$

  That is, a prospect is acceptable if the utility resulting from integrating the prospect with one's assets exceeds the utility of those assets alone.

- *Risk aversion: u is concave (*$u^{''} < 0$*)*" [37].

Another theory that provides analysis of decision under risky scenarios is the Prospect theory proposed by Kahneman and Tverskey [37]. Prospects theory is a critique of expected utility theory and states that the value of an outcome is governed by the changes in the wealth (i.e., gains or losses) instead of the final assets [37]. In prospect theory, probabilities are replaced by decision weights.

2.5. CASL codes and activities

The Consortium for Advanced Simulation of Light water reactor is a U.S. Department of Energy (DOE) sponsored Energy Innovation Hub (EIH) for M & S of nuclear reactor applications. The primary objective of CASL is to develop modeling and simulation capabilities to support decisions regarding safe and efficient operation of commercial nuclear power reactors. As described by Kothe (see Table 2.6), there are three critical elements that govern the integration of M & S into decisions related to operation and safety of NPP. The rigor, depth, and quality of M & S tools developed by CASL are dependent on these elements.

Table 2.6: Critical elements governing M & S integration into the decision (as presented by Kothe [38])

Acceptance by user community	• Address real problems in a manner that are more cost-effective than current technology • Meet needs of utility owner-operators, reactor vendors, fuel suppliers, engineering providers, and national laboratories
Acceptance by regulatory authority	• Address issues that could impact public safety • Deliver accurate and verifiable results
Acceptance of outcome by public	• Provide outcomes that ensure high levels of plant safety and performance

CASL identified different key issues or challenge problems (CPs) that needs to be addressed using advanced M & S tools to resolve the problems surfacing safe and efficient operation of the current fleet of nuclear reactors. The CASL challenge problems involve complex multi-physics and multi-scale interactions. CASL identified different simulation codes with the capability to model different physics - CTF [39] (and CFD simulation) for thermal-hydraulics, MPACT [40] for neutronics, BISON [41] for fuel performance and MAMBA [42] for coolant-chemistry. Although the majority of these codes are mature with respect to their domain of individual physics, their extension to multiphysics and multiscale CASL challenge problems has been subject to further functionalities development and extensive Verification & Validation work. Approximate nature of models and data makes it difficult to ascertain the confidence in the predictive capability of computation tools. CASL adopted the Predictive Capability Maturity Model (PCMM) to assess the predictive capability of individual and coupled simulation codes for different challenge problems. Code assessment using PCMM is based on different elements that encompass - software quality assurance, verification, calibration, validation and uncertainty quantification of codes.

CASL M & S work is focused on high fidelity simulation of CPs. Consequently, validation is one of the most challenging elements of CASL M & S activities. This challenge primarily arises due to the shortage of data to match the high level of modeling details in CASL codes. CASL adopted the validation pyramid approach to counter these challenges. However, multiphysics and multiscale nature of CASL challenge problems limit the use of AIAA validation pyramid for CASL CPs. CASL developed a modified validation pyramid for CPs using the Component Identification and Ranking Process [26].

CASL validation pyramid [26] consists of four levels (see Figure 2.10) where the quantity of interest (challenge problem) under full-system condition lies at the top of the pyramid. Decomposition is performed top down with respect to the QOI. Scaled prototype forms the second level of the pyramid. At this level, scaling argument is used to establish the applicability of data for predicting full-scale plant scenarios. Multiphysics components and subsystems form the third level of the pyramid and assess the coupled calibration and validation of different codes. Finally, at the bottom of the pyramid, we have single physics component where validation is focused on individual codes.

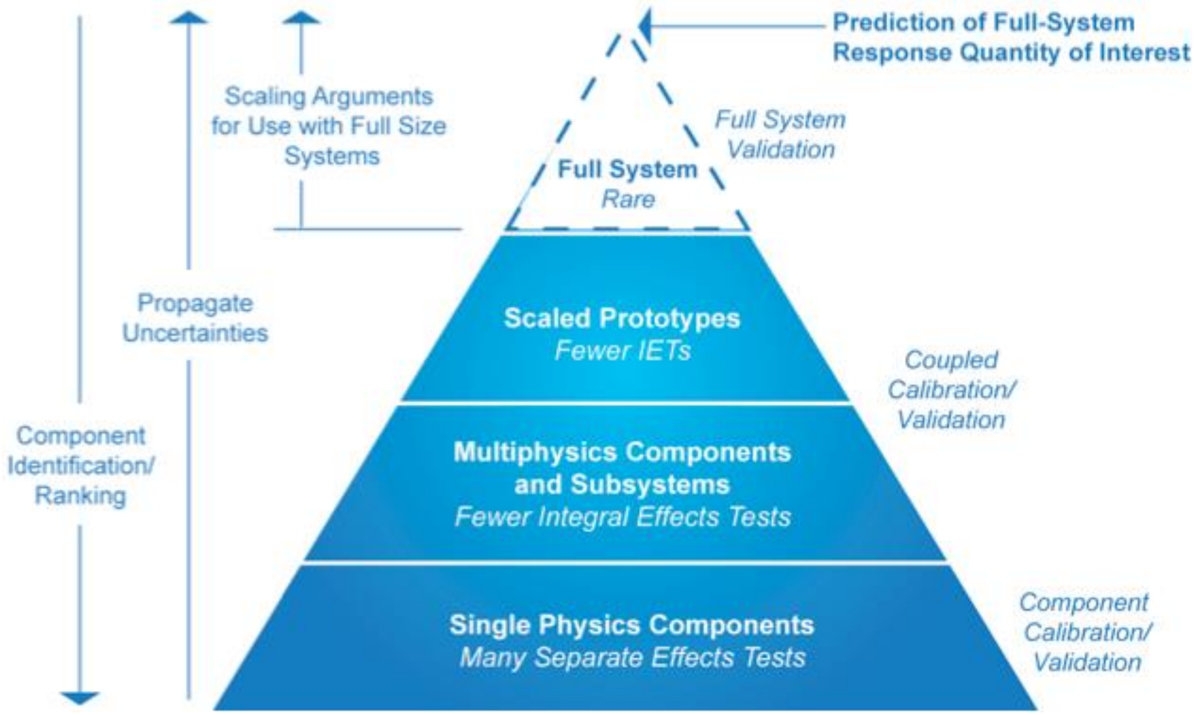


Figure 2.10: CASL Validation Pyramid [26]

As described in the CASL report [31], the CASL validation pyramid capture three set of information in the validation pyramid:

- Information related to our understanding of physical phenomena that occur in the system (Px)
- Information related to component/subsystem in the plant system (Sx)
- Information related to experiments used for validation at each level (Ex)

Given the level of complexity of the CASL challenge problem and the nature of available information (in terms phenomena, model, and data), it becomes difficult to perform the assessment using a single pyramid for validation. Therefore, in the recent CASL V & V plan [31], the process of validation is described through three individual pyramids corresponding to identified phenomena (Phenomenological Pyramid), relevant experiments (data pyramid) and relevant models (model or code pyramid). Phenomenology pyramid (PP) serves as the guiding structure for the formation of data pyramid (PE) and model pyramid (PM).

2.6. Scaling techniques

Scaling analysis is an integral part of the code validation process. This section presents an overview of different scaling methodologies/techniques that have been developed to perform similarity analysis of nuclear reactor systems at different scales.

The earliest methodology for determining similarity at different scales was based on a simple dimensionless analysis. Buckingham formalized this methodology into a theorem called Pi-theorem. The Pi theorem is useful in the analysis of simple models with few parameters. However, as the number of parameters and complexity of model increases, a more structured way for similarity analysis needs to be adopted.

Scaling analysis is the process of assessing similarity at different scales using the mathematical model of the system of interest. The dimensionless groups that represent the ratio of different forces or physical processes are used to determine similarity at different scales. Depending on the complexity of the system different criteria for similarity has been identified. We start our discussion with the illustration of scaling analysis for hydraulic systems and then discuss scaling methodologies for thermal-hydraulic systems in nuclear reactors.

Heller describes three criteria for mechanical similarity in a hydraulic system [43],

- Geometric similarity: Geometric similarity is determined based on the similarity in shape. It requires all length dimensions to be scaled by a constant factor in the experiment.

- Kinematic similarity: Kinematic similarity requires geometric similarity along with similarity of motion of particles in both experiment and application. This criterion implies a constant ratio of time, velocity, acceleration and discharge at all the time in both experiment and application.
- Dynamic similarity: Requirement of dynamic similarity includes geometric similarity, kinematic similarity and an identical ratio of all the forces in both experiment and application.

One method of performing scaling analysis of a hydraulic system is based on the dimensionless groups that represent the ratio of different forces or physical processes in the system. Table 2.7 lists important force ratios in fluid dynamics. It is evident from these force ratios that all the requirements for dynamic similarity cannot be fulfilled simultaneously in one experiment, e.g., to preserve Re ratio vL should be constant and to preserve F, $v/\sqrt{L}$ should remain same (as gravitational acceleration g and kinematic viscosity ν is constant if identical fluid is considered). Such type of discrepancies leads to scaling distortions in the similarity analysis of a complex phenomenon.

Table 2.7: Important force ratios in fluid dynamics [43]

Force ratio	Expression	Symbolic representation
Euler number (E)	Pressure force/inertial force	$P/\rho v^2$
Reynolds number (R)	Inertial force/viscous force	Lv/ν
Prandlt number (Pr)	momentum diffusivity/thermal diffusivity	ν/α
Froude number (F)	(inertial force/gravity force)$^{1/2}$	$v/(gL)^{1/2}$

For a nuclear reactor thermal-hydraulic process, additional similarity criteria are required that take into account heat transfer and energy dissipation during different processes (Energy scaling). Material scaling is another issue that arises due to cost limitation of the experimental facility, e.g., Freon instead water is employed as a coolant to reduce pressure and electrically heated rods are employed in place of fuel rods.

Different methodologies for the scaling analysis of nuclear reactor thermal-hydraulic system has been developed. These methodologies perform similarity analysis using the conservation equation, boundary and initial conditions of the process under consideration. Before discussing these methodologies, let us first take a brief look at the major objectives of scaling analysis in the context of a nuclear reactor [44, 45]:

- Guide the design of new test facility.
- Identify dimensionless groups which provide compact representation and correlation of experimental results applicable to both scaled experiment and full-scale plant (local scaling).
- Rank phenomena based on the importance of underlying processes in a transient scenario (PIRT process).
- Provide quantitative estimation of various scaling distortions.
- To determine similarity criteria for global scaling analysis, considering the interaction of different components within the system.

From the perspective of simulation code, two objectives of scaling analysis can be identified:

- Development of empirical correlation or constitutive laws for modeling meso-scale or micro-scale processes in a system code.

- Assess the applicability of data from existing facility for validation of existing code or calibration of new codes.

Scale effects are also present in code in the form of empirical models that are obtained from different separate effect tests. The range of validity of these constitutive relations is not completely specified (e.g., pressure and flow rate ranges are specified, but void fraction or slip ratio ranges may not be specified). Furthermore, relationships are often used outside their range of validity. As the physical size of a nuclear power plant is much larger than scaled test facility, nodalization becomes another issue that needs to be addressed by scaling analysis [46].

All scaling methodology adopts a hierarchical approach for scaling analysis to perform similarity assessment of the complete system. Ishii scaling employs three levels in the hierarchy[13]:

- Integral system scaling: Integral system scaling is performed by introducing small perturbations in the system conservation equation in the transient scenario. The solution of the perturbed conservation equation gives various transfer functions relating different variables (pressure, inlet flow, enthalpy and void fraction). These transfer functions are non-dimensionalized to identify the similarity between experiment and reactor application.
- Control volume and boundary flow scaling: At the second level similarity analysis are performed by non-dimensionalizing the balance equation of mass momentum and energy of the control volumes. Integral system and control volume together provide the criteria for dynamic similarity of the system responses.
- Local phenomenon scaling: At the third level, scaling is performs based on the similarity analysis of various local phenomena. The ratios of dimensionless groups

representing different physical phenomena are used to perform similarity analysis at this level. Scaling of these local phenomena is described as the major sources of scaling distortion by Ishii [43].

The Hierarchical Two-Tiered Scaling Analysis (H2TS) [14] and Fractional Scaling Analysis (FSA) [15] proposed by Zuber also adopts hierarchical system decomposition for similarity analysis of the complete system. H2TS methodology specifies detailed hierarchical decomposition which consists of a system, subsystems, modules, constituents, phases, geometric configurations, fields, and processes at different levels in the hierarchy. Figure 2.11 shows the system decomposition and hierarchy described by H2TS methodology. FSA consist of only three levels in the hierarchy[15]:

- System (Macro-scale),
- Components (Meso-scale),
- Processes (Micro-scale).

Both FSA and H2TS were primarily developed for scaling analysis of transient or accident scenario. Therefore, time and length scales are crucial in the estimation of scale distortion in these methodologies. In H2TS, scaling distortion is assessed based on the time ratios of the dominant process. Characteristic time ratio in H2TS is expressed as [14],

$$\pi = \omega\tau \qquad (2.8)$$

here, ω represents the frequency and τ represents the residence time. π provides a measure of relevance of a process by combining residence time in the control volume with the characteristic frequency of the process. In FSA, scale distortion is estimated on the basis of fractional change in the state variable due to an agent of change.

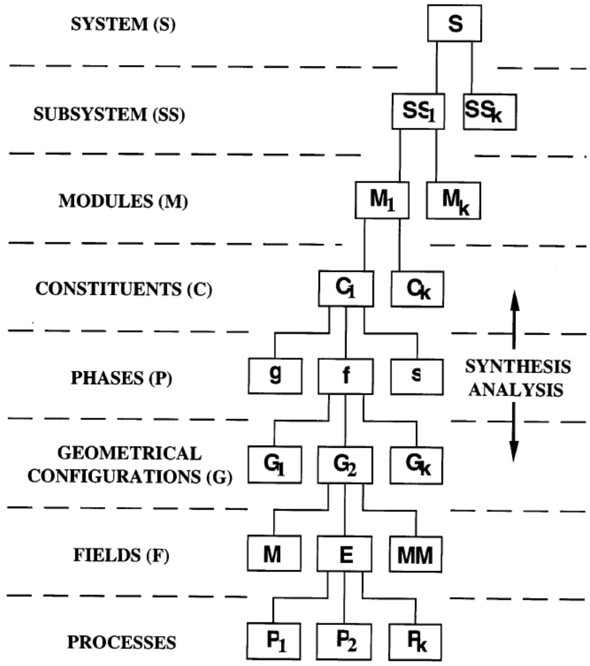


Figure 2.11: System decomposition and hierarchy in H2TS [14]

Recently, another approach for scaling analysis called Dynamical system scaling (DSS) methodology was developed by Reyes[47]. It has been developed to assess process scale distortion over the entire duration of a process [47].

Even though different scaling techniques have been developed over the past three decades, scaling assessment remains a daunting task with a limited demonstration of practical scenarios.

2.7. Safety case and argumentation

A safety case is analogous to code prediction and validation in several ways, particularly from the perspective of "nature of the problem." A safety case is a structured argument, supported by evidence, which intends to justify that a system is acceptably safe. Similarly, code validation can be described as the "confidence argument" supported by evidence (model and data) that justifies the claim that code provides reliable prediction in the extrapolation domain. Therefore, it is important to explore and understand the philosophy of safety case and related concepts. This section presents a brief review of development in the field of safety case, argumentation and evidence theory.

2.7.1. Safety case, arguments, and evidence

As described by Talus, a safety case is a document produced by the operator of a facility which [48]:

- "*Identifies the hazards and risks,*
- *Describe how the risks are controlled,*
- *Describes the safety management system in place to ensure that the controls are effectively and consistently applied*" [48].

The safety case and supporting safety assessment is submitted to the regulatory body for approval. A safety case provides a structured framework for documenting and presenting all the safety-related information in a systematic and consolidated manner. Safety assessment is the main component of the safety case and involves assessment of a number of components, as shown in Figure 2.12. Safety assessment is performed by determining limits, controls, and conditions for the safety problem under study [49].

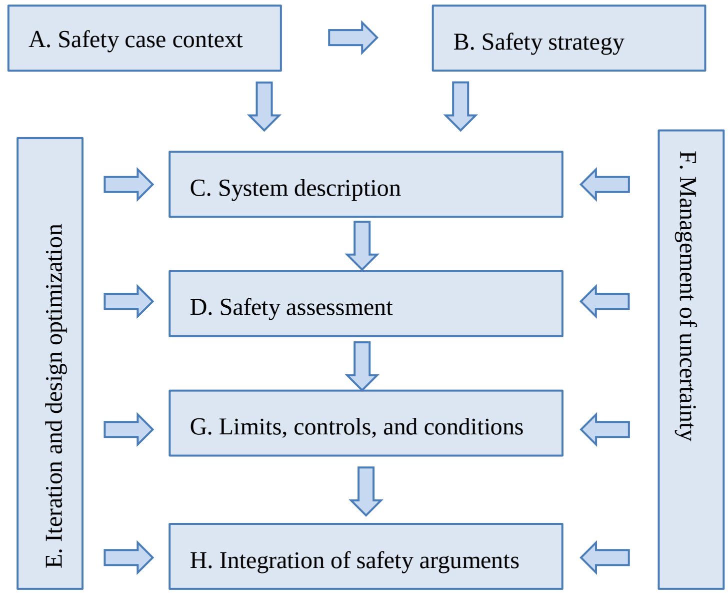


Figure 2.12: Components of a safety case [49]

The U.K. Defense Standard 00-56 describe the safety case as, "a structured argument, supported by a body of evidence that provides a compelling, comprehensive and valid case that a system is safe for a given application in a given environment"[12]. In the context of a safety case, an argument is defined as, "a set of claims that a person puts forward in an attempt to show that some further claim is rationally acceptable"[50]. Evidence in the context of safety case is defined as, "the information that serves as the grounds and starting-point of (safety) arguments, based on which the degree of truth of the claims in arguments can be established, challenged and contextualized" [51]. It is evident from these definitions that "Argument" and "evidence" play a key role in the formation of a safety case. Nair et al. [52] explains that argument and evidence follow a mutual dependency relation in the representation of a safety case and neither is complete without the other. An argument needs to be supported by convincing evidence to make it rationally acceptable. Similarly, evidence needs to be accompanied by argument to clarify its significance in the context of the objectives of a safety case[52].

Sun [51] describes how evidence can be categorized as direct evidence, backing evidence and counter-evidence based on their association with the confidence in the safety case. Direct evidence supports confidence assessment based on the product of the safety case processes while backing evidence supports confidence assessment based on the safety case processes itself [51]. Counter-evidence are evidence that undermines the confidence in the arguments presented in the safety case. Sun and Kelly [51, 53] also describe the classification of evidence as, analytical evidence, empirical evidence, adherence evidence and engineering judgment, based on their form and origin (see Table 2.8 for examples).

Table 2.8: Types of evidence [51, 53]

Evidence	Example
Analytical evidence	model simulation, hazard analysis, cause analysis, consequence analysis, behavior modeling
Empirical evidence	observation and measurement of behaviors from various types of testing, historical operation, or real practice
Adherence evidence	adherence to standards, guidance, design rules, prescribed process, accepted best practice
Engineering judgement	inspection, review, or expert opinion based on personal knowledge, engineering experience, and creative thoughts

Argumentation theory provides the basis for the formulation of safety cases. Therefore, it is important to understand how the nature of the problem and its solution impact the depth and rigor of the evidence and arguments. The U.K. Defense Standard 00-56 [12] introduced “The McDermid Square” (see Figure 2.13) to illustrate this relationship. It shows that minimum argumentation and evidence is needed for situations where both problem and solution are familiar (top left quadrant in the McDermid square). On the other hand, situations where both problem and solutions are unfamiliar (bottom right quadrant in the McDermid Square, situations with high uncertainty and risk) extensive argumentation, evidence and scrutiny are required to establish confidence in the safety claim.

		Solution	
		Familiar	Unfamiliar
Problem	Familiar	Minimal Argument and standard evidence from the domain, e.g., stability certificate	Focused argument on reasons for novel solution, plus the appropriate evidence
	Unfamiliar	Minimal Argument and standard evidence from another domain, e.g., railway safety case	Extensive argument and evidence, with substantial independent scrutiny

Figure 2.13: The McDermid Square as presented in the U.K. Defense Standard 00-56 [12]

Figure 2.14 shows the Toulmin's argument model [54] to explain the process of argumentation. Elements of Toulmin's argument model consist of different classifiers like claim, data/ground, warrant, backing, qualifier, and rebuttal. Description of these classifiers as presented by Toulmin [54] with an example related to code adequacy assessment is shown in Table 2.9. Toulmin's argument illustrates how supporting a claim with explicit pieces of information can enhance clarity and assurance in a specific claim.

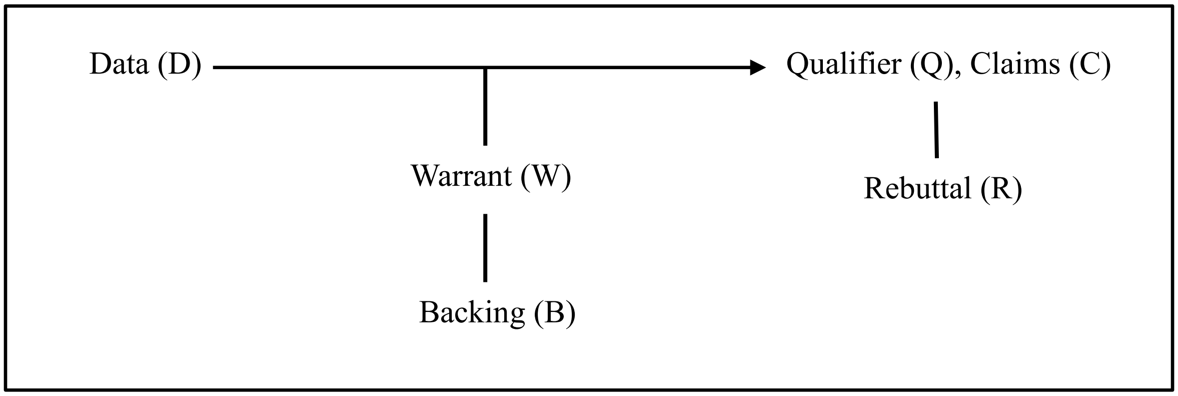


Figure 2.14: Toulmin's Argument model [54]

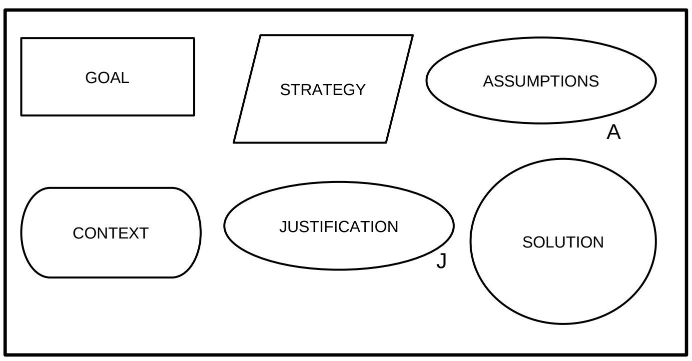
Table 2.9: Elements of Toulmin's Argument model [54]

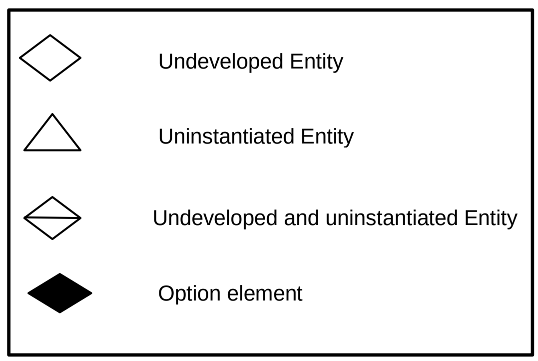
Claims	The statement we wish to justify	e.g., Claims related to the adequacy of a code for an intended reactor application
Data	The fact we appeal to, the grounds or information on which our claim is based	e.g., Evidence related to code verification, validation and uncertainty quantification
Warrant	A statement authorizing the step from data to claim is true; an inference rule	e.g., Scaling argument authorizing data applicability in the extrapolation regime
Backing	A reason for trusting the warrant	e.g., Result of scaling analysis
Qualifie r	A term or phrase reflecting the degree to which the data support the claims, e.g. generally, probably	e.g., "Adequate" relevant database is available for validation
Rebuttal	Specific circumstances in which the argument will fail to support the claims as exceptions	e.g., Insufficient evidence (lack of validation data)

Structuring information using explicit classifiers enhance clarity in the representation of a safety case. There are two approaches that are used for formalizing the safety case: Claim, argument and evidence notation (CAE) [55] and Goal Structuring Notation (GSN) [6]. The Goal Structural Notation is perhaps the most popular technique used for structured, graphical representation of assurance arguments for confidence assessment in safety case.

2.7.2. Goal structuring notation (GSN)

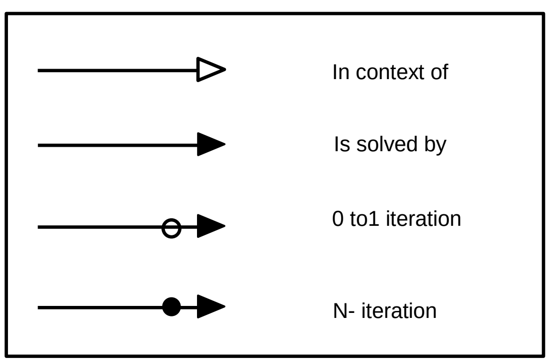
Goal structuring notation has gained widespread popularity over the past decade as a technique for logical representation and formalization of safety cases. Technically, GSN provides a reason-based conceptual approach for graphical representation of arguments. In this approach, goals are broken down into sub-goals until they can be directly supported by direct evidence; meanwhile, the strategy for decomposition, justification and the assumption made during the process, and the context of each step is clearly specified [6]. GSN consist of six elementary blocks: *goal, strategy, assumption, justification, context, and solution* (see Figure 2.15 (a))





(a) Basic GSN blocks

(b) Status indicators in GSN



(c) Links in GSN

Figure 2.15: Elementary blocks of GSN

A *Goal* block contains a claim or statement related to the objective of the problem of interest. Every GSN tree consists of a top goal that specifies the overall objective of the problem or safety case. A *goal* may be accompanied by a set of contextual information and assumptions to further clarify and elaborate the claim or statement in the goal block. *Context* and *Assumption* blocks define the basis on which a goal is stated and specify the conditions under which the claim (or statement in the goal block) is assumed to be valid. A *Strategy* block contains reasoning information that illustrates the "*nature of inference that exists between the goal and its supporting sub-goals,"[56]*. It contains an argument that asserts the approach to decompose goals into sub-goals. A *strategy* is often accompanied by a *Justification* block to clarify the rationale and provide backing to the argument in the strategy block *[56]*. *Solution* blocks act as the termination points in the GSN network. They containb the reference to evidence or facts that support different claims and goals in GSN network. The techniques to develop a GSN network can be summarized using the six-step method proposed by Kelly [57]:

1. Identify the *goals (i.e. claims)* to be supported;
2. Define the basis on which the *goals* are stated (*context, justification,* and *assumption*);
3. Identify the *strategy* used to support the *goals*;
4. Define the basis on which the *strategy* is stated (*context, justification*, and *assumption*);
5. Elaborate the *strategy* (and proceed to identify new *goals* – back to step 1), or step 6;
6. Identify the basic *solution (i.e., evidence).*

GSN has different indicators like "undeveloped entity," "uninstantiated entity" and "option element" that can be used throughout the network (see Figure 2.15(b) for symbols). Indicator for "undeveloped entity" is used whenever a specified line of argument needs further development. It can be used to indicate an undeveloped goal or strategy. "Uninstantiated entity" indicator is used when an abstract entity is required to be restored with a substantial instance at some later stage. It can be used with any GSN block. “Option element” is used when several alternatives are available to make a choice [56]. Blocks in GSN are connected using different types of links. These links are described in Figure 2.15 (c). An example of using GSN (from GSN community standard [56]) is shown in Figure 2.16.

Modular GSN extension has been developed to enable management of large safety case using modular architecture. GSN community standard documents all the elements of GSN and can be referred for further reference [56].

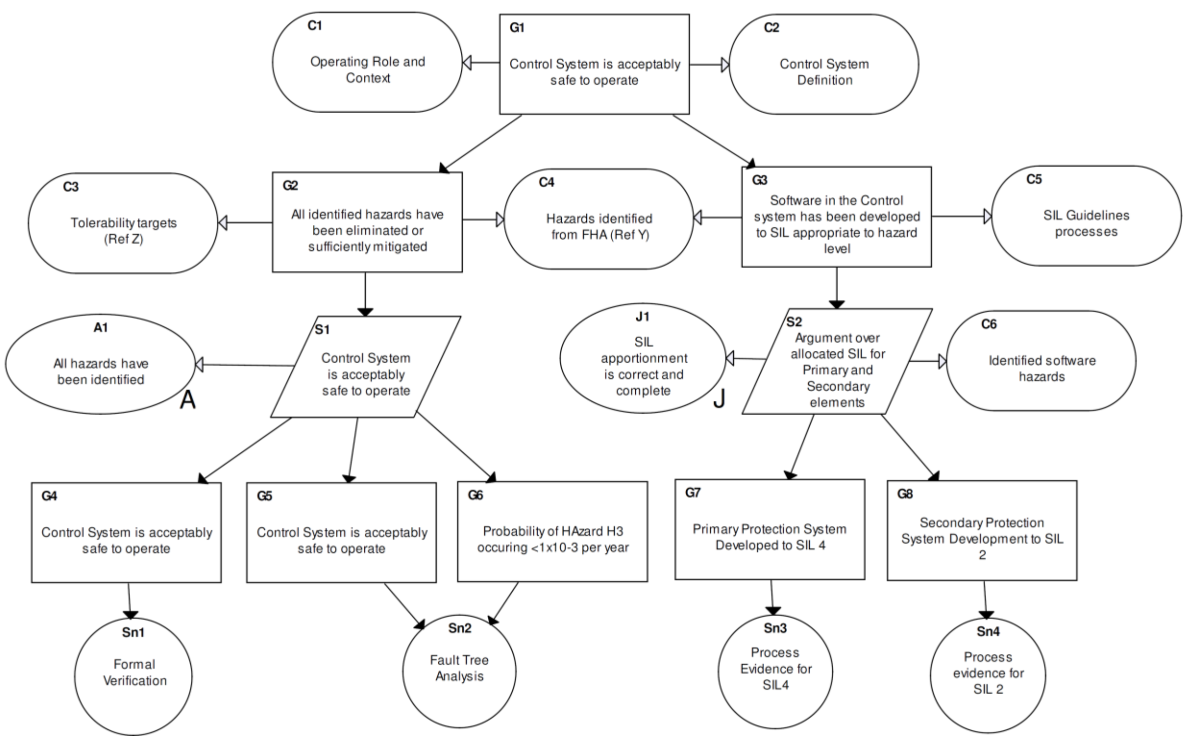


Figure 2.16: An example of using GSN from GSN community standard [56]

2.8. Candidate tools/techniques for maturity quantification

Maturity quantification requires a technique that can fuse subjective information based on expert's knowledge and judgment (criteria for evidence assessment) with objective information related to the evidence of VVUQ. This section presents review and illustration of different techniques that have been explored for the formulation of maturity quantification in the proposed research. Three techniques are discussed in this section:

- Evidential reasoning,
- Fuzzy logic and fuzzy inferences system,
- Bayesian Networks.

2.8.1. Evidential reasoning

Evidential reasoning approach was proposed by Yang et al. [10] to resolve problems involving *multiple attribute decision-making situation under uncertainty[10]*. It is based on the Dempster-Shafer theory and provides a framework to rank, assess and quantify qualitative attributes in a decision problem. In particular, it addresses situations where multiple factors need to be assessed simultaneously by uncertain, subjective judgment[10]. The important elements of this approach as presented by Yang & Singh [10] and Yang & Xu [58]are discussed below:

The ER approach is based on the hierarchical process where attributes are identified at each level in the hierarchy and assessment is performed based on the degree of confidence in each attribute. As this approach follows a hierarchical structure, each higher-level attribute (y) is subdivided into a set lower level sub-attributes or basic attributes ($e_1, e_1, \dots . e_i, \dots e_L$). This set is defined by,

$$E = \{e_1, e_2, \dots . e_i, \dots e_L\} \tag{2.9}$$

Furthermore, each sub-attribute can be assigned a weight depending on its relative importance. Based on expert judgment all lowest-level attributes are graded using fuzzy quantifiers that are expressed using a Likert-scale that consist of N grades $H = \langle H_1, H_2, \dots, H_n., H_N \rangle$ where H_1 corresponding to the lowest grade like "very poor" and H_N corresponds to the highest grade like "Excellent". All lowest-level attributes are assessed by expert judgment using "Belief function". Belief functions help the assessor to incorporate any uncertainty that they might have in a particular grade. For example, an assessor may have only 70% confidence that the experiment used in the validation has "very high" relevance to the application and 30 % confidence that the experiment has "high" relevance to the application. The expert overall assessment of an attribute e_i, can be expressed as

$$S(e_i) = \{(H_n, \beta_{n,i}), n = 1, \dots . N\} \qquad (2.10)$$

here, β_{ni} represents expert confidence (or degree of belief) that attribute e_i achieves grade H_n. For a given attribute e_i, we have,

$$\sum_{n=1}^{N} \beta_{n,i} \leq 1 \qquad (2.11)$$

The belief in an attribute does not require to sum to one and can be smaller than 1. Sum equal to zero implies no confidence while sum equal to 1 implies complete confidence. The distribution of belief at the lowest level nodes are combined with the belief function of their adjoining nodes and propagated up in the hierarchy using the ER algorithm. In this way, confidence is obtained at the top most level in the hierarchal structure. If E as described in previous equation consists of all the basic attribute that are needed to describe a general attribute y (higer level attribute) then axioms are followed for the propagation of belief function from lower to higher level attribute (y) in the hierarchy (Nair et al. [59] and Yang & Xu [58]):

- *"If none of the basic attributes for y is assessed at a grade H_n , then $\beta_{n,y} = 0$.*
- *If all the basic attributes are assessed to a grade H_n, then $\beta_{n,y} = 1$.*
- *If all the basic attributes are completely assessed to a subset of evaluation grades, then y should be completely assessed to the same subset of grades.*
- *If an assessment of a basic attribute in E is incomplete, then the assessment of E should also be incomplete to a certain degree"* (Nair et al. [59] and Yang & Xu [58]).

Major steps in the ER algorithm can be described as:

- Weights assignment and weight normalization: In ER weight is assigned to each attribute in the hierarchy. As the weight assignment is an important part of the ER approach, the authors recommend using rating methods or pairwise comparison as proposed by Saaty[7], in the Analytic Hierarchy Process.

$$\omega = \{\omega_1, \omega_2, \dots \omega_i \dots . \omega_L\} \tag{2.12}$$

ω_i is the relative weight of the i^{th} basic attribute (e_i) with $0 \le \omega_i \le 1$.

In general, normalization of weights is performed by normalizing all the weights with respect to the most important weight. Yang et al. [10] adopted a different approach for weight normalization where normalized weights ($\overline{\omega}_i$) are obtained by,

$$\overline{\omega}_i = \alpha \frac{\omega_i}{\max_i\{\omega_i, i = 1, \dots . L\}} \tag{2.13}$$

The constant α in the above equation is obtained by,

$$\prod_{i=1}^{L} (1 - \overline{\omega}_i) \le \delta \tag{2.14}$$

here, δ is constant which represents the degree of approximation in aggregation.

- Basic probability assignment: Basic probability assignment is the next step in the ER approach. It involves calculation of probability masses, $m_{n,i}$ and $m_{H,i}$. $m_{n,i}$ is the basic probability mass that represents the extent to which e_i supports the hypothesis that the general attribute y is assessed to the grade H_n. It is given by,

$$m_{n,i} = \bar{\omega}_i \beta_{n,i} \qquad n = 1,2 \ldots . N \tag{2.15}$$

$m_{H,i}$ is the residual probability mass which is obtained by,

$$m_{H,i} = 1 - \sum_{n=1}^{N} m_{n,i} = 1 - \bar{\omega}_i \sum_{n=1}^{N} \beta_{n,i} \tag{2.16}$$

- Attribute aggregation: If $E_{I(i)}$ is the subset of the first i basic attributes then,

$$E_{I(i)} = \{e_1, e_1, \ldots . e_i\} \tag{2.17}$$

$m_{n,I(i)}$ is the probability mass that represents the extent to which all the attributes in $E_{I(i)}$ support hypothesis that y is assessed to a grade H_n. $m_{H,I(i)}$ is the corresponding residual probability mass. $m_{n,I(i)}$ and $m_{H,I(i)}$ are obtained by combining all basic probability masses $m_{n,j}$ and $m_{H,j}$ for $n = 1, \ldots . , N$ and $j = 1, \ldots . i$ using the recursive ER algorithm:

$$m_{n,I(i+1)} = K_{I(i+1)}\big(m_{n,I(i)} m_{n,i+1} + m_{n,I(i)} m_{H,i+1} + m_{H,I(i)} m_{n,i+1}\big) \tag{2.18}$$

$$m_{H,I(i+1)} = K_{I(i+1)} m_{H,I(i)} m_{H,i+1} \tag{2.19}$$

where, $K_{I(i+1)}$ is the normalizing factor given by,

$$K_{I(i+1)} = \left[1 - \sum_{t=1}^{N} \sum_{\substack{j=1 \\ j \neq t}}^{N} m_{t,I(i)} m_{j,i+1}\right]^{-1} \qquad i = 1,2, \ldots . . L-1 \tag{2.20}$$

- Estimation of Combined degree of belief: Combined degree of belief in ER approach is given by,

$$\beta_n = m_{n,I(L)} \qquad n = 1, \ldots, N \tag{2.21}$$

$$\beta_H = m_{H,I(L)} = 1 - \sum_{n=1}^{N} \beta_n \tag{2.22}$$

here, β_H is the degree of belief that has not been assigned to any individual grade after all the L basic attributes (or sub-attributes) has been assessed.

Nair et. al. [59] have illustrated the use of Evidential reasoning approach in conjunction with Goal structuring notation for assessment of safety cases.

2.8.2. Fuzzy logic and fuzzy inference system

Since its conception, the fuzzy set theory has found wide application in different streams of science like biology, medicine, controls [60-70], etc. Fuzzy logic is well known for its ability to capture expert knowledge; it has been applied to resolve several decision-making problems in different areas of engineering and science [71-76]. The strength of fuzzy logic lies in the membership function. Membership function (MF) provides a unique methodology to express fuzzy quantifier and fuzzy probability using a mathematical function. Membership functions also help in weighing all the evidence based on the degree of confidence in their "truth value". Another important feature of fuzzy logic is the ability to deal with heterogeneous data. Availability of a large number of mathematical operation makes it easier to combine different type of information. Fuzzy logic provides an efficient methodology to codify expert knowledge using membership function. These qualities of fuzzy logic make them a candidate tool/technique for quantitative maturity assessment.

In the following sections, we provide a brief review of different concepts related to fuzzy logic and fuzzy sets for further reference and understanding.

2.8.2.1. Definition of fuzzy set

A fuzzy set can be defined as a set of fuzzy boundaries. According to Zadeh [8], "Fuzzy logic is determined as a set of mathematical principle for knowledge representation based on the degree of membership rather than on crisp membership of classical binary logic." Fuzzy logic is a theory of sets that calibrate obscurity or uncertainty [74]. Unlike Boolean algebra that works on binary logic (Truth or false, 0 or 1), fuzzy logic characterizes the data using the degree of membership. It provides a precise way of representing approximate reasoning and imprecise information. A fuzzy set "A" of the universe of discourse X, with elements represented by x, is defined by its membership function $\mu_A(x)$ as,

$$\mu_A(x): X \to [0,1] \quad (2.23)$$

where,

$\mu_A(x) = 0$, if x does not belong to A,

$\mu_A(x) = 1$, if x completely belongs to A,

$0 < \mu_A(x) < 1$, if x partially belongs A.

Fuzzy set for the class of middle-aged man is shown in Figure 2.17. The range of x for which $\mu_A(x) \neq 0$ is called the "support of the fuzzy set" and the range of x for which $\mu_A(x) = 1$ is called the "core of the fuzzy set." "Universe of discourse" consist of all possible value of the variable x and is usually decided by the expert judgment.

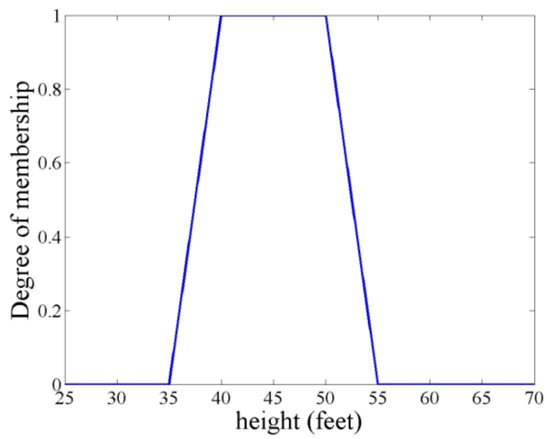


Figure 2.17: Fuzzy set for middle aged man

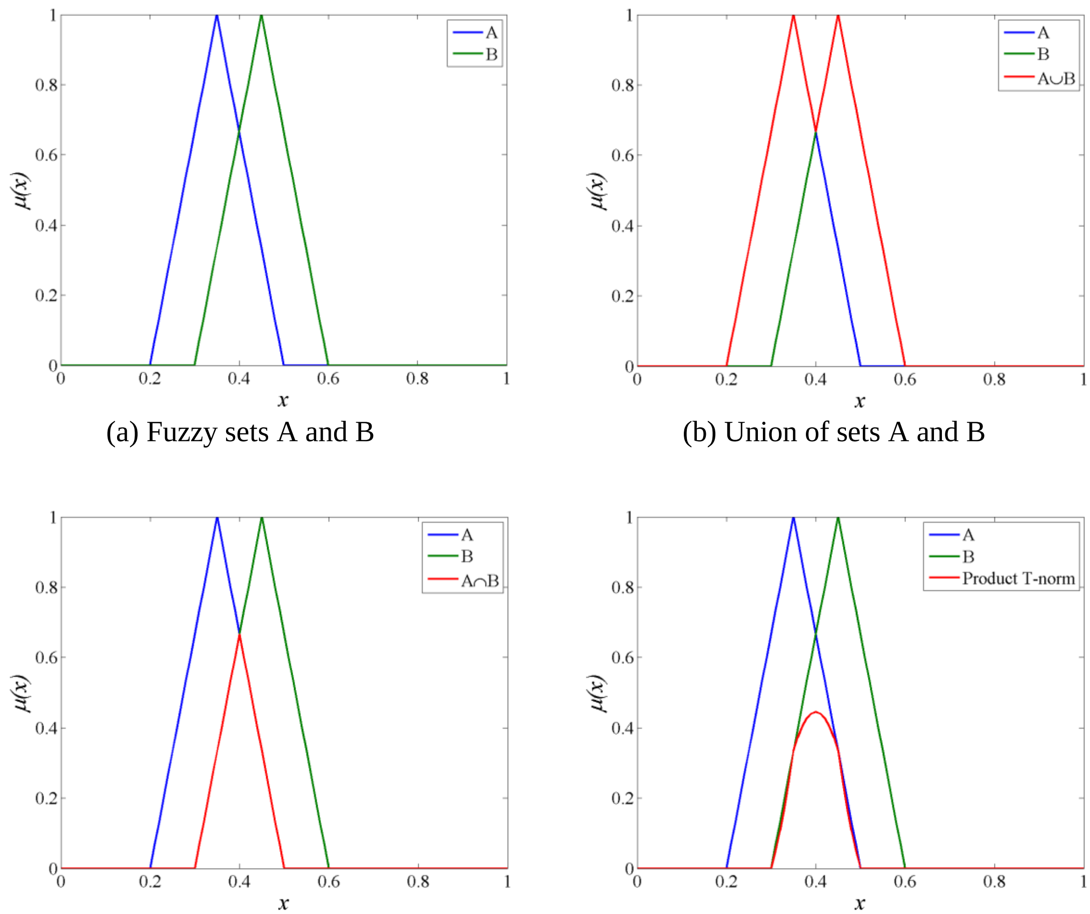
2.8.2.2. Operations on fuzzy sets

Fundamental operations that are defined on classical sets like Complement, Union, Intersection, and Containment are also applicable to Fuzzy sets. Additionally, there are several other operations described in the literature [73]. In this section, some of these operations are described. Let us consider a set of finite elements $S = \{x_1, x_2, \dots \dots, x_n\}$. The fuzzy sets A and B are defined as A⊆S and B⊆S. Standard fuzzy operations on these sets are defined in Table 6.

Table 2.10: Standard fuzzy operations

Operation	Symbol	Definition
Union	$A \cup B$	$\mu_{A\cup B}(x) = \max[\mu_A(x), \mu_B(x)]$
Intersection	$A \cap B$	$\mu_{A\cap B}(x) = \min[\mu_A(x), \mu_B(x)]$
Complement	A^c	$\mu_{A^c}(x) = 1 - \mu_A(x)$
Inclusion	$A \subseteq B$	$\mu_A(x) \leq \mu_B(x), \forall x \in S$
Equality	$A = B$	$\mu_A(x) = \mu_B(x), \forall x \in S$

Figure 2.18 shows the application of some of these standard fuzzy operations on fuzzy sets. Apart from these standard operations triangular norm (T-norm), triangular conforms (T-conorm), averaging and several other operations can be performed on the fuzzy sets.



(a) Fuzzy sets A and B

(b) Union of sets A and B

(c) Intersection of sets A and B

(d) Product T-norm on sets A and B

Figure 2.18: Different operations on fuzzy sets

2.8.2.3. Fuzzy rules

Fuzzy rules are conditional statements that can be used to perform fuzzy operations on linguistic variables. A simple fuzzy rule can be described as,

IF x is A

AND y is B

THEN z is C

here, x, y, and z are linguistic variables, and A, B and C are linguistic values obtained from the fuzzy sets with the universe of discourses X, Y, and Z, respectively.

2.8.2.4. Fuzzy inference systems

Fuzzy inference system is a system that maps a given input to output using the theory of fuzzy sets. The basic structure of a typical fuzzy inference system is shown in Figure 2.19 [77]. Specifying the crisp inputs is the first step in the design of a fuzzy inference system. A crisp input consists of some measured quantity, observation or any other direct evidence that is available as a numerical value. Sometimes, when a direct evidence/ measurement/observation is not available, these crisp inputs can be estimated based on the expert knowledge. Crisp input lies within the universe of discourse of the input variable. The range of the universe of discourse is usually determined based on the expert judgment. Each crisp input is applied to its corresponding fuzzy set to obtain its membership value which represents our degree of confidence in its *truth* value. Rule base consists of a set of rules that tell the inference system how different input quantities can be combined to reach the final inference. Fuzzy operation (min, max, prod, or, average, etc.) are used to formalize these rules. Fuzzy operation on the fuzzified input variables gives a fuzzified output set. Defuzzification of this fuzzified output set is performed to obtain the crisp output which represents our quantity of interest.

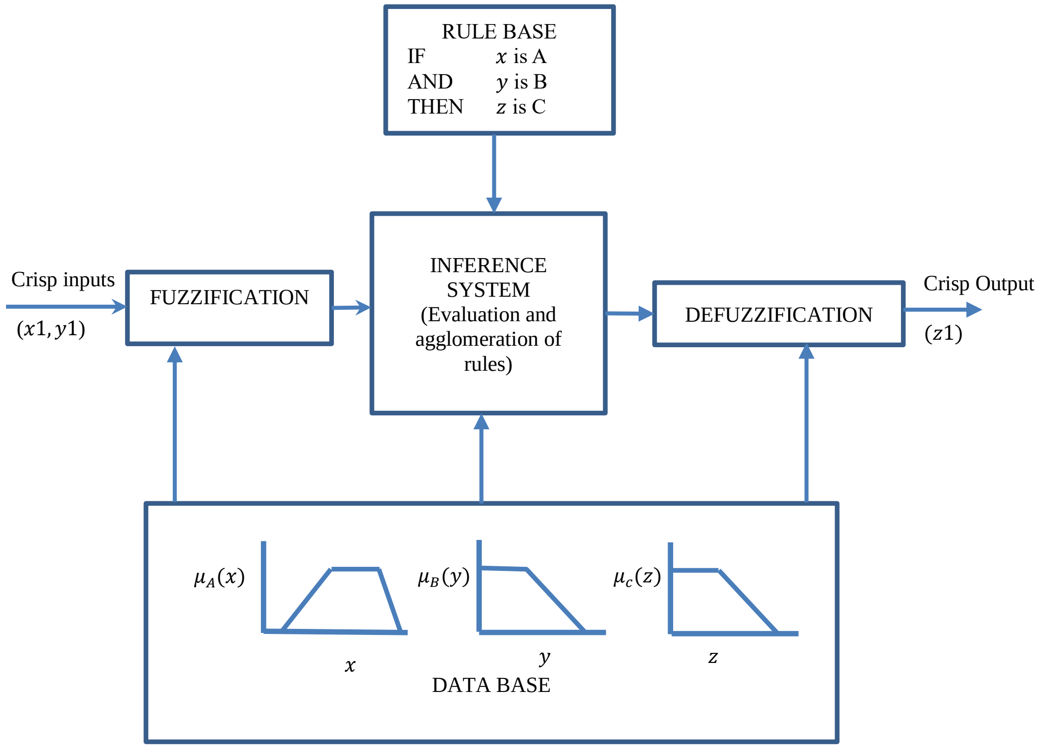


Figure 2.19: Fuzzy Inference system [77]

To illustrate how a fuzzy inference system works, we present a simple two inputs, one output problem which consists of two rules. The set of rules and linguistic definition for this problem are given in Table 2.11.

Table 2.11: Example of fuzzy set

Rule # 1	
Linguistic definition	Fuzzy Definition
IF Rain is less AND Temperature is cold THEN Crop_produce is bad	IF x is A1 AND y is B1 THEN z is C1
Rule # 2	
Linguistic definition	Fuzzy Definition
IF Rain is more AND Temperature is warm THEN Crop_produce is good	IF x is A2 AND y is B2 THEN z is C2

x, y, and z (rain, temperature, and crop produce) are linguistic variables; A1 and A2 (less and more) are linguistic values obtained from fuzzy set on the universe of discourse X (rain); B1 and B2 (cold and warm) are linguistic values that are obtained from fuzzy set on the universe of discourse Y (temperature); C1 and C2 are linguistic values that are obtained from fuzzy set on the universe of discourse Z (crop produce). It should be noted that we have not considered all the possible scenarios in this problem. To keep the problem simple and easy to understand, only two scenarios (represented by rule 1 and rule 2) have been considered. Figure 2.20shows the structure of the fuzzy inference system for this problem. Depending on the problem any shape of membership function can be chosen for the input and output variables. Triangular, trapezoidal, Gaussian are some of the commonly used membership functions. We have chosen a triangular membership function for all the variables in this problem. Mamdani-style inference technique [74] is used to solve this problem. Different steps used in this fuzzy inference system are described below:

- **Fuzzification:** The crisp input for this problem are assumed to come from a weather forecast report. The universe of discourse for the fuzzy set of rain consists of all possible values of the amount of rain during the monsoon season (starting from 0 for "no rain" to a maximum value X). Similarly, the universe of discourse for temperature and crop produce consist of all possible values of these variables. The range of universe of discourses is decided on the basis of the past weather reports consisting of temperature and rain data from past years. The weather forecast report provides the crisp input values x1 and y1 for the set of rain and temperature, respectively. The crisp input x1 gives a membership value $\mu_{A1}(x) = 0.2$ and $\mu_{A2}(x) = 0.6$ using the membership function A1 and A2, respectively. Similarly, for the crisp input y1, we obtain membership values $\mu_{B1}(y) = 0.4$ and $\mu_{B2}(y) = 0.4$ using the membership function B1 and B2, respectively. In this way, both input variables are fuzzified over membership functions used by their respective fuzzy sets.

- **Rule evaluation:** Each fuzzy rule in our problem has two antecedents corresponding to the two input variables x and y, respectively. To combine these antecedents, we apply the fuzzy intersection operator (min function) on the fuzzified inputs obtained in the previous step.

$$\mu_{A1 \cap B1}(x, y) = \min[\mu_{A1}(x), \mu_{B1}(y)] = 0.2 \tag{2.24}$$

$$\mu_{A2 \cap B2}(x, y) = \min[\mu_{A2}(x), \mu_{B2}(y)] = 0.4 \tag{2.25}$$

Next, the result of these antecedents evaluation is applied to the consequent's membership functions C1 and C2, respectively. The output membership functions C1 and C2 are scaled to the truth value of their respective rule antecedent. This method of correlating the truth value of the rule antecedent with the rule consequent is called clipping or correlation minimum.

- **Agglomeration of rule's consequent:** In this step, the clipped membership function for all the rule consequents are combined to obtain a single fuzzy set ($\mu_C(z)$) for the output variable.

- **Defuzzification:** This is the last step of the fuzzy inference process and involves evaluation of the final quantity of interest (i.e., the crisp output) from the fuzzy set obtained in the previous step. Different methodologies for defuzzification is available in the literature. The centroid technique is one of the most popular techniques. It finds a point (often referred as Centre of gravity COG due to its analogy with gravitational center of gravity) inside the universe of discourse where a vertical line would divide the output Fuzzy set into two equal parts. The crisp output (z1) for this example is obtained by,

$$\mathrm{COG(z1)} = \frac{\int_0^Z z\, \mu_C(z)\, dz}{\int_0^Z \mu_C(z)\, dz} \tag{2.26}$$

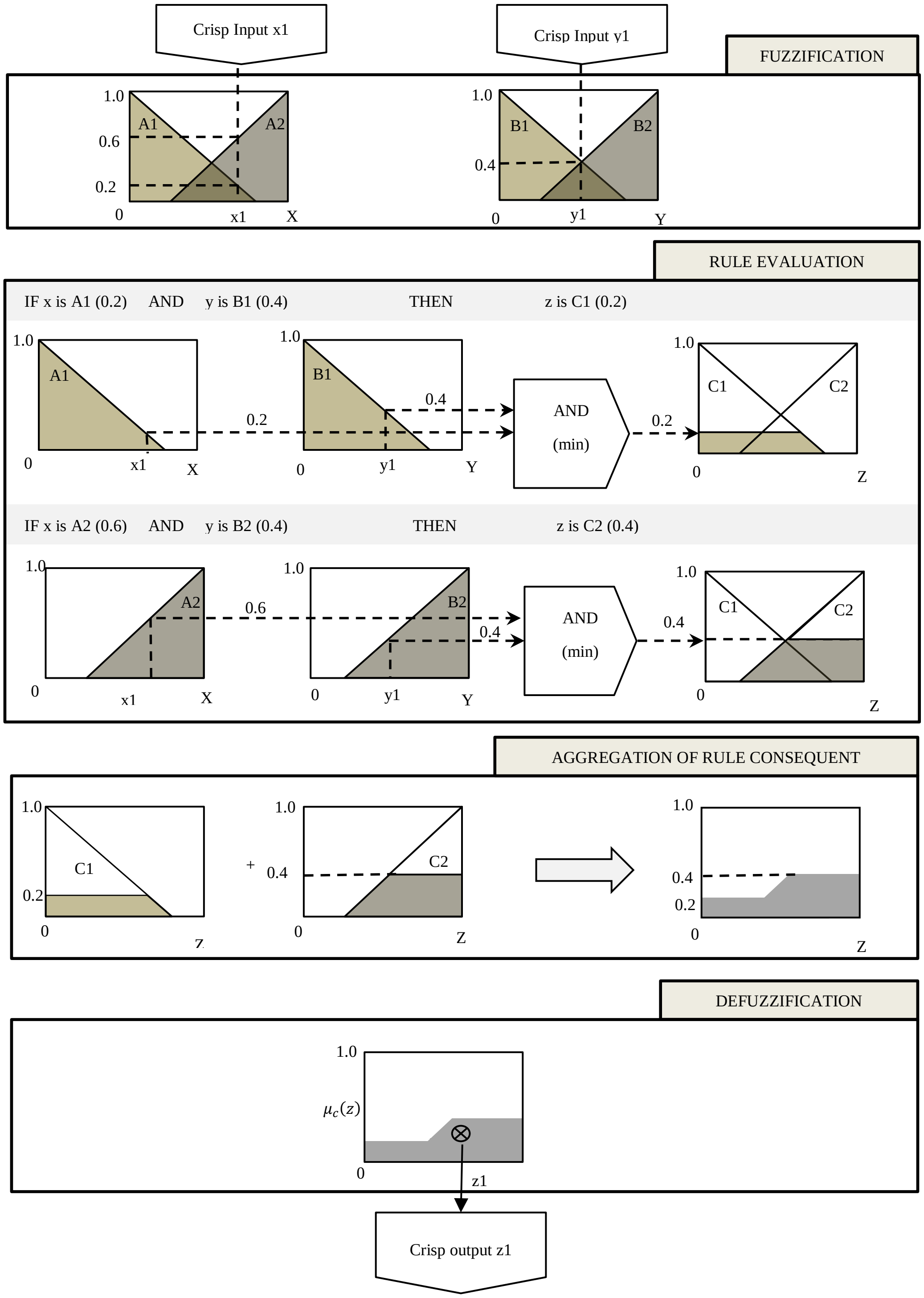


Figure 2.20: A Fuzzy inference system (based on Mamdani-type inference

The Mamdani-style fuzzy inference has found wide application due to its ability to capture expert knowledge. However, the defuzzification process for this inference technique is computationally less efficient. To counter this problem, Michio Sugeno[78] introduced another inference technique which is like Mamdani inference except for the rule consequent. This inference technique is called Sugeno inference technique. It can have a single spike called singleton as the membership function. For a Mamdani-style fuzzy inference we have the rules of the following form:

Mamdani style

IF x is A

AND y is B

THEN z is C

while for a Sugeno-style fuzzy inference we have the rules of the following form:

Sugeno-style

IF x is A

AND y is B

THEN z is $f(x, y)$

here, x, y and z are linguistic variables; A, B and C are fuzzy sets on the universe of discourse X, Y and Z, respectively; and $f(x, y)$ is a mathematical function of x and y, e.g. for a zero-order Sugeno model we can have: $f(x, y) = k$, were k is constant. For a 1st order Sugeno model we can have $f(x, y) = k_0 + k_1 x + k_2 y$, here k_0, $k_1$and k_2 are constants called consequent parameters.

Some example to illustrate the use of fuzzy logic for maturity quantification are provided in Appendix A.

2.8.3. Bayesian networks

Bayesian network (also known as Bayes network or Bayesian belief network) are directed acyclic graphs (DAG) which employ probabilistic reasoning and Bayes' theorem to model relationship between a set of random variables. The concept of Bayesian networks was introduced by Pearl [9] in 1985. Thereafter, the Bayesian networks have been extensively used to model belief in biological science, medicine, forensic science, law, decision system, risk analysis, [9, 79-91], etc. There are two important elements of a Bayesian network: (1) Directed acyclic graph, (2) Conditional probability distribution.

A directed acyclic graph is formed by a network of nodes and arcs. Nodes represent random variables (binary, discrete or continues) and arcs are connection links between nodes that reflect probabilistic dependence between the nodes. Each random variable has a set of mutually exclusive states. Figure 2.21 shows an example of DAG with node description.

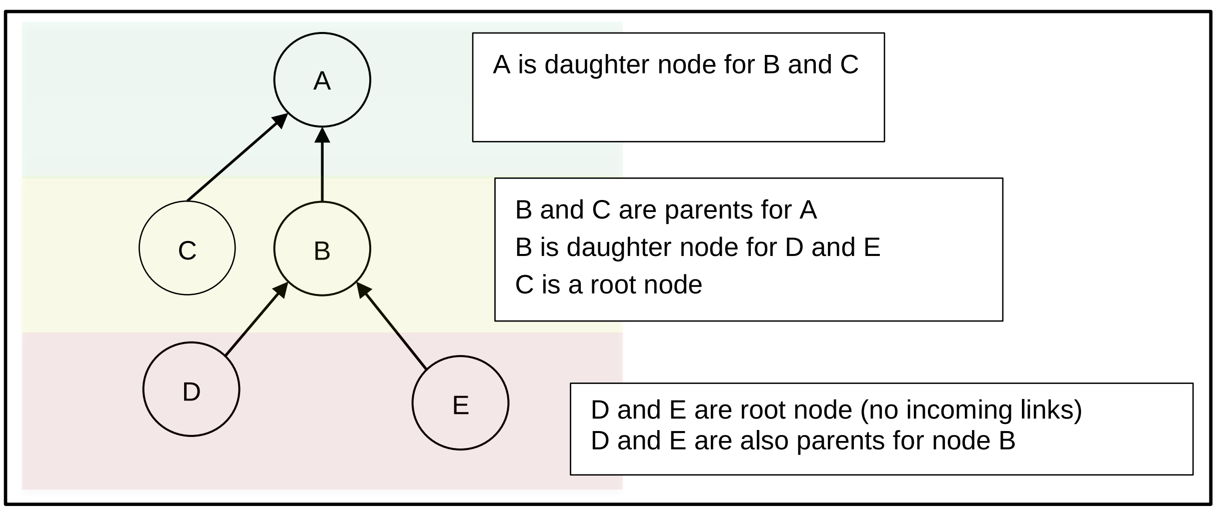


Figure 2.21: Example of DAG with node description

A Conditional Probability Distribution (CPD) provides the relation between a node and its parent nodes. It is often represented in tabular form in the Bayesian network and called as Conditional Probability Table (CPT). CPD at a node x_i can be represented as $P\left({x_i}/{\gamma_i}\right)$, where γ_i is a set of all parent node of x_i. As root nodes, do not have any parents, γ_i is a null set for them. Therefore, $P\left({x_i}/{\gamma_i}\right)$ is determined from the priors, i.e. $P\left({x_i}/{\gamma_i}\right) = P(x_i)$.

A BN follows Markov property, which implies that a node is conditionally independent of all other nodes given its parents, descendant (children) and descendants' parents. Using the probability laws, a joint distribution with n variable can be broken down as a product of n-1 conditional distribution and a marginal distribution [88],

$$P(x_1, x_2, \dots . x_n) = \left[\prod_{i=2}^{n} P\left({x_i}/{x_1, x_2, \dots . x_{i-1}}\right)\right] P(x_1) \tag{2.27}$$

This decomposition forms the basis of the chain rule in BN and facilitates computation in the BN.

As described by Taroni et. al.[88], Bayesian networks provide a built-in computational architecture that helps in determining the effect of the evidence on the state of the variable. This architecture [88],

- "*Updates probabilities of the states of the variable on learning new evidence.*
- *Utilizes probabilistic independence relationships, both explicitly and implicitly represented in the graphical model, to make computation more efficient*" [88].

Let us consider a simple example to further illustrate the above-mentioned points related to the Bayesian networks. We consider an example concerning a crime-scene investigation. In this example, two factors are considered to determine if person X committed the crime or not: (1)

Person X was present at a nearby location just after the crime, and (2) Fingerprints of person X were found at the site of the crime. The BN for this problem is shown in Figure 2.22. We have three nodes A, B and C in BN that are described as:

A: Person X committed the crime

B: Person X was present at a nearby location just after the crime

C: Fingerprints of person X were found at the sight of the crime

Each node is associated with two mutually exclusive states: True or False. Based on the relation of node A with node B and C, a conditional probability table (CPT) is constructed (see Table 2.12). Conditional probability table is usually constructed using subjective probabilities that are based on expert knowledge and judgment about the strength of different cause and effect relations in the network.

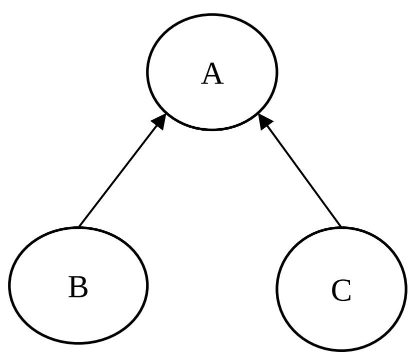


Figure 2.22: BN for the Crime Investigation problem

Table 2.12: Conditional probability table for node A

Node C	True		False	
Node B	True	False	True	False
True	0.9	0.6	0.4	0.1
False	0.1	0.4	0.6	0.9

The marginal probability that person X committed the crime is calculated using the CPT by,

$$P(A_{True}) = \sum_i \sum_j P\left(A_{True}|B_iC_j\right)P(B_i)\,P\left(C_j\right) \tag{2.28}$$

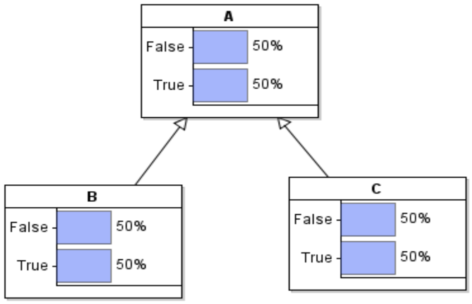
and

$$P(A_{False}) = \sum_i \sum_j P\left(A_{False}|B_iC_j\right)P(B_i)\,P\left(C_j\right) \tag{2.29}$$

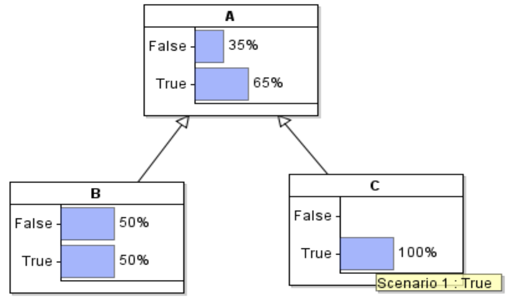
here, $i = True, False$ and $j = True, False$

Initially, there was no evidence to support the claims in node B and C. Therefore, initial assessment gives $P(A_{True}) = 0.5$. When the evidence associated with node C is obtained, the probability associated with A is updated. If C is True and we obtain $P(A_{True}) = 0.65$. If the evidence associated with node B also indicate that B is True, then the probability of A being true becomes very high. In this case, we obtain, $P(A_{True}) = 0.9$. These three cases are shown in Figure 2.23.

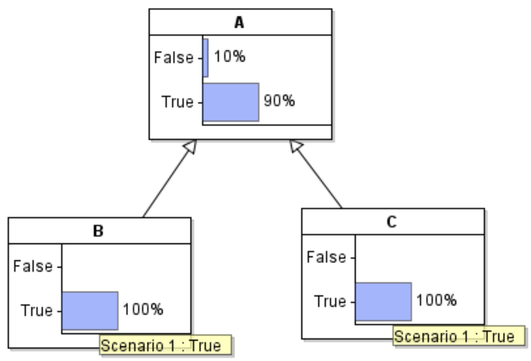
Denney et al. [92] and Guiochet et al. [93] have illustrated the use of Bayesian network in conjunction with Goal structuring notation for confidence assessment in safety cases.



(a) No evidence to support B and C



(b) Evidence supports, “C is True”



(c) Evidence supports, “both B and C are True”

Figure 2.23: Probability for the node A under three different condition: (a) No evidence to support B and C, (b) Evidence supports, “C is True”, and (C) Evidence supports, “both B and C are True”

In this section, we reviewed three techniques: (1) Evidential reasoning (ER), (2) Fuzzy logic (FL) & fuzzy inference system (FIS), and (3) Bayesian networks (BN). Comparison of the three techniques discussed in this section is presented in Table 2.13. Apart from these techniques, decision trees and influence diagram are two other techniques that could be potential tool for maturity quantification in the proposed research. However, in this work we restrict our focus to Bayesian Network. As Bayesian network are based on Bayes’ theorem they provide strong mathematical basis for evidence-based quantitative maturity assessment. Furthermore, they can handle large network and takes in to account correlation between variables.

Table 2.13: Comparison of BN, FL & FIS and ER

Property	Bayesian Network (BN)	Fuzzy logic and fuzzy inference system (FL & FIS)	Evidential reasoning (ER)
Mathematical base	Bayes' theorem	Fuzzy membership function, arithmetic and logical operations	Evaluation analysis model and Dempster-Shafer theory (evidence combination rule)
Basis for network architecture	Directed acyclic graphs (DAG)	Arithmetic and logical operation	Dempster-Shafer theory (Attribute aggregation and weighting process)
Ability to model Hierarchical system	Yes	Hierarchical structure is not straight forward but can be obtained by using multiple fuzzy inference systems in hierarchy	Yes
Incorporation of expert knowledge	Conditional probability table	Membership function	Belief function
Ability to integrate subjective data (expert opinion) and objective data (evidence)	Strong	Strong	Strong
Ease of implementation	Yes, easy to build but estimation of conditional probabilities could be challenging	Yes, easy to implement but proper choice of membership function is crucial for building an efficient framework	Yes, easy to implement
Ability to handle complex networks	Very strong	Fair	Fair
Application to decision-making situation	Yes, very popular	Popular	Not as popular as Bayesian network and Fuzzy logic & fuzzy inference system.

2.9. Transforming goal structuring notation (GSN) to computable network

Goal structuring notation facilitates structural knowledge representation and provides basic architecture for confidence representation. However, a GSN by itself does not provide a computable network for quantitative confidence assessment. Therefore, it needs to be used in conjunction with other methodologies that facilitate quantitative confidence assessment. This section provides a brief review of techniques used to transform a GSN representation to a computable network for quantitative estimation of confidence in safety cases.

Transformation of GSN to other computable networks is facilitated by identifying the basic sources of uncertainty in the argument model (i.e., the GSN representation). Guiochet et al. [93] provides a systematic technique to transform GSN to the Bayesian network. They address two type of uncertainties in the GSN argument model. These uncertainties are associated with "appropriateness" and "trustworthiness" of the evidence [94]. In Figure 2.24, "Uncertainty in B supports A" is related to the appropriateness of the evidence while "Uncertainty in solution B" is related to the trustworthiness of the evidence B [93]. Confidence network for this problem would consist of a simple BN with two nodes A and B, and a directed link from B to A.

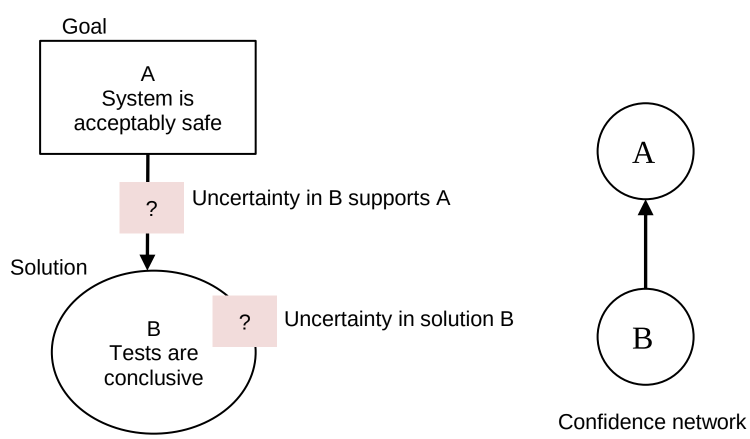


Figure 2.24: Transformation of a simple argument to confidence network as illustrated by Guiochet et al. [93]

Another illustration by Guiochet et al. [93] for transforming an alternative argument to confidence network is shown in Figure 2.24.

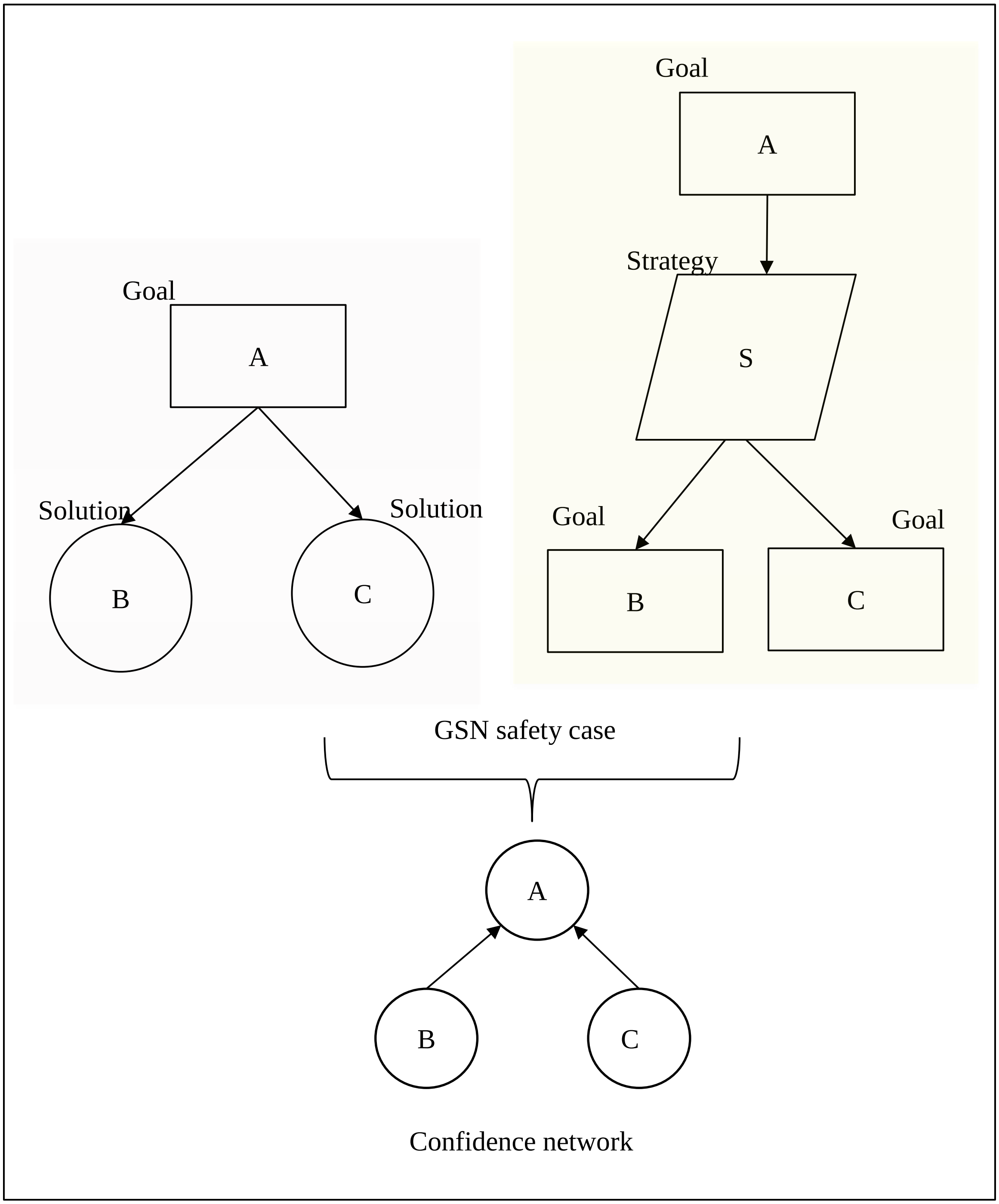


Figure 2.25: Transforming GSN network to Confidence network [93]

Nair et al. [59] uses assurance claim point (ACP) to identify the key uncertainty points for quantitative confidence assessment using the Evidential Reasoning (ER) approach. ACP are a graphical notation in GSN that are used to link the assertion in the safety argument to the confidence argument. Figure 2.26 shows assurance claim points in a GSN representation as illustrated by Hawkins et. al. [94].

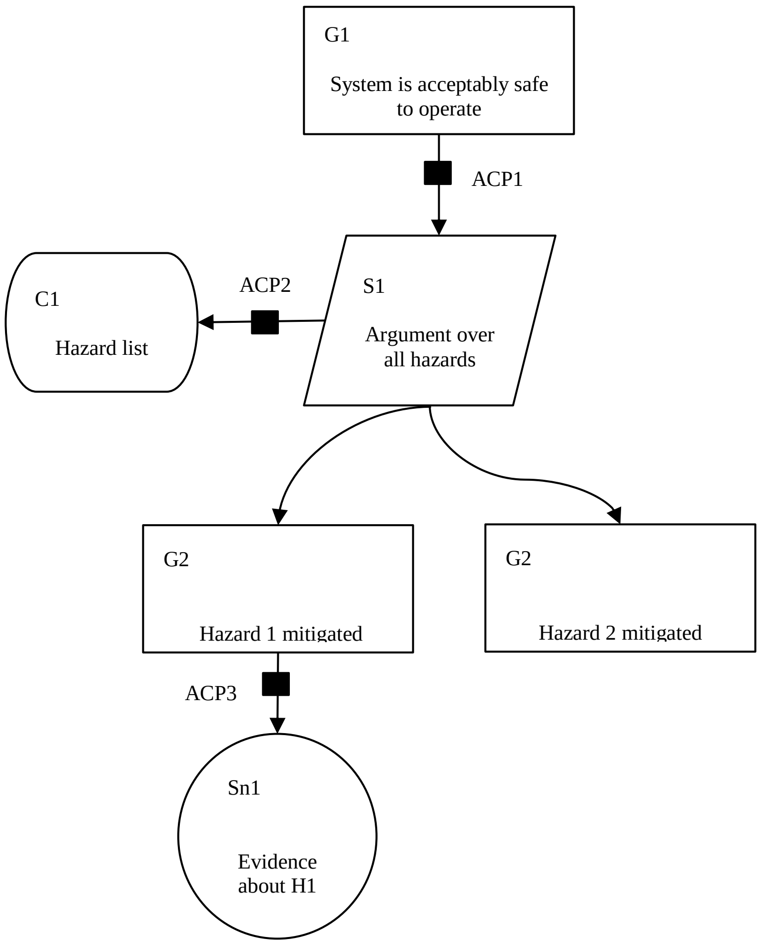


Figure 2.26: Illustration of ACP as presented by Hawkins et. al. [94]

2.10. Summary

This chapter presents a comprehensive review of different standards, methodologies, and techniques that provide necessary background and fundamental support for the development of the proposed framework. As complexity resolution is the most important step in the development and assessment of a computational tool for a real engineering application, one section of this chapter was dedicated to a review and discussion of techniques for complexity resolution. Next, different standards and methodologies for credibility assessment that guides the development of the proposed framework were discussed and compared. A strong emphasis was placed on maturity assessment methodologies and decision process as they form the basis for the formulation of the proposed framework. Next, Goal structuring notation which is an argument modeling technique used in the formulation of the framework was described along with related concepts like, safety case, argument, and evidence.

As quantitative maturity assessment is an important part of the proposed framework, different techniques that can be employed for this purpose were illustrated and compared with each other. In the last section, current techniques for transforming GSN into a computable network were discussed.

CHAPTER 3: FORMALIZING THE ASSESSMENT PROCESS

3.1. Introduction

This chapter is divided into four parts. The first part discusses the code development process and different sources of uncertainty that impacts the code prediction. In the second part, an overview of code verification and validation process is presented. The third section describes the research approach adopted for the formulation of the proposed framework. The last section describes the formulation and illustration of different elements of the framework.

3.2. Process of code development and sources of uncertainty

A nuclear reactor is a complex system that involves innumerable physical processes occurring in conjunction with each other at different ranges of scale. These physical processes continuously interact with each other and govern the operation and performance of the reactor system at any given time. Computational tools are employed to simulate these processes, in order to support decisions regarding design, operation and safety analysis of the reactor system. Ideal simulation of the reactor systems is not possible due to lack of knowledge, modeling limitation, computational and experimental overhead. Consequently, code prediction becomes highly uncertain. Different sources of uncertainty become eminent as we go through the process of development of an M & S tool. Figure 3.1 illustrates the basic steps in the development of a Thermal-hydraulic code. The architecture and complexity of different M & S tools differ based on their intended use and domain of physics. However, in a broad sense the development of any (nuclear engineering) M & S tool can be described by three major phases:

(1) Problem Specification and model conception

(2) Model formulation

(3) Numerical simulation

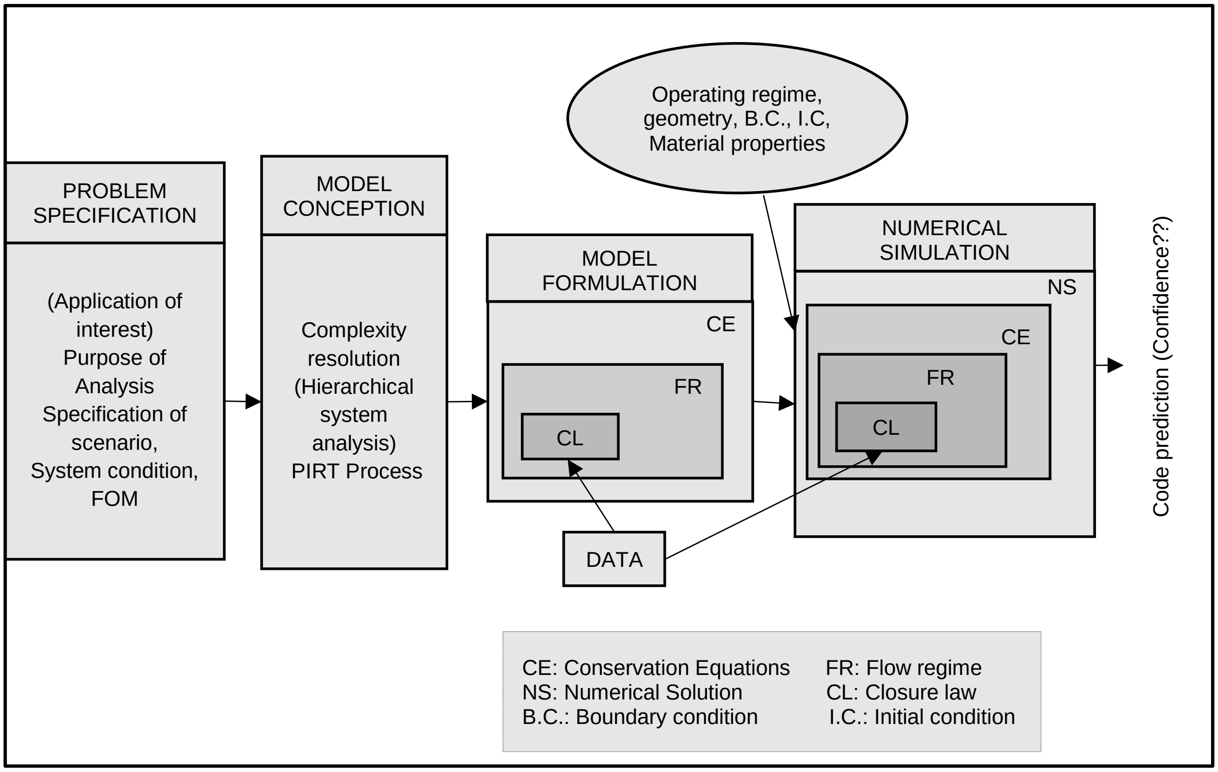


Figure 3.1: Illustration of the process of M & S (for a Thermal-hydraulic code)

(1) Problem Specification and model conception: This phase involves specification of the purpose of analysis and problem definition. Based on the problem definition, system condition and figure of merit are identified. Next important step is the complexity resolution. In nuclear engineering, complexity resolution is performed by the PIRT process. The rationale behind the PIRT process is that all the processes and phenomena are not important and does not contribute equally to the figure of merit. Therefore, processes and phenomena that are important with respect to the FOM are identified and ranked using the PIRT process. For CASL challenge problems, the system decomposition is performed

with respect to governing physics (Neutronics, Fuel performance, Coolant chemistry and thermal hydraulics) and scale (micro-scale, meso-scale and macro-scale) of the underlying phenomena. Hence, scale separation and physics decoupling are the two elementary principles that guide complexity resolution for CASL Challenge problems. The outcome of PIRT process is governed by the expert's knowledge and understanding about the problem of interest. Therefore, this step is a major source of epistemic uncertainty.

(2) Model formulation: The process of the model formulation can be attributed to large model uncertainty due to two factors. First, due to uncertainty in the selection of appropriate model form for the solution of the problem (i.e., model form uncertainty). Second, due to uncertainty in the parameters of the selected model (i.e., model parameter uncertainty). Incomplete knowledge about initial and boundary condition are additional sources of epistemic uncertainty during model formulation. Specifically, in thermal hydraulics codes lot of empirical/semi-empirical correlations are employed to fill in the gap created by missing physics. These correlations are developed from small-scale experiment often employing different fluid, geometries, and are developed under steady-state conditions. Hence, these correlations are another source of uncertainty in model formation. As a discussed in section 2.3.1, nodalization is a major source uncertainty and scale distortion in system analysis codes. Material properties and other input quantities are also contributors to uncertainty in response prediction.

(3) Numerical solution: The third phase involves the numerical solution of the model. Major sources of uncertainty in this phase are associated with discretization and numerical solution schemes.

Model calibration is employed to reduce uncertainty in the model parameter by calibrating the model using the experimental data. As discussed earlier, the code contains different sources of uncertainty. It is difficult to isolate parameter uncertainty from other sources of uncertainty. Consequently, the process of reducing uncertainty in model parameters using model calibration often overcompensates for other sources of uncertainty. This effect leads to an unknown impact on the code prediction in the extrapolation regime.

3.3. Code verification and validation overview

The process of V & V helps in determining the reliability of code prediction. Figure 3.2 depicts the V & V process and different sources of error in the prediction of the response quantity of interest based on the illustration by Oberkampf et al. [3]. The original illustration by Oberkampf et al. [3] does not consider scaling analysis. However, as we extend this illustration to nuclear engineering codes, scaling analysis becomes very important. Due to cost constraints and safety implications of the accident scenarios, validation data from full-scale reactor application is rarely available. Therefore, scaling analysis becomes essential to determine the applicability of data from reduced scale test facility to full-scale reactor systems.

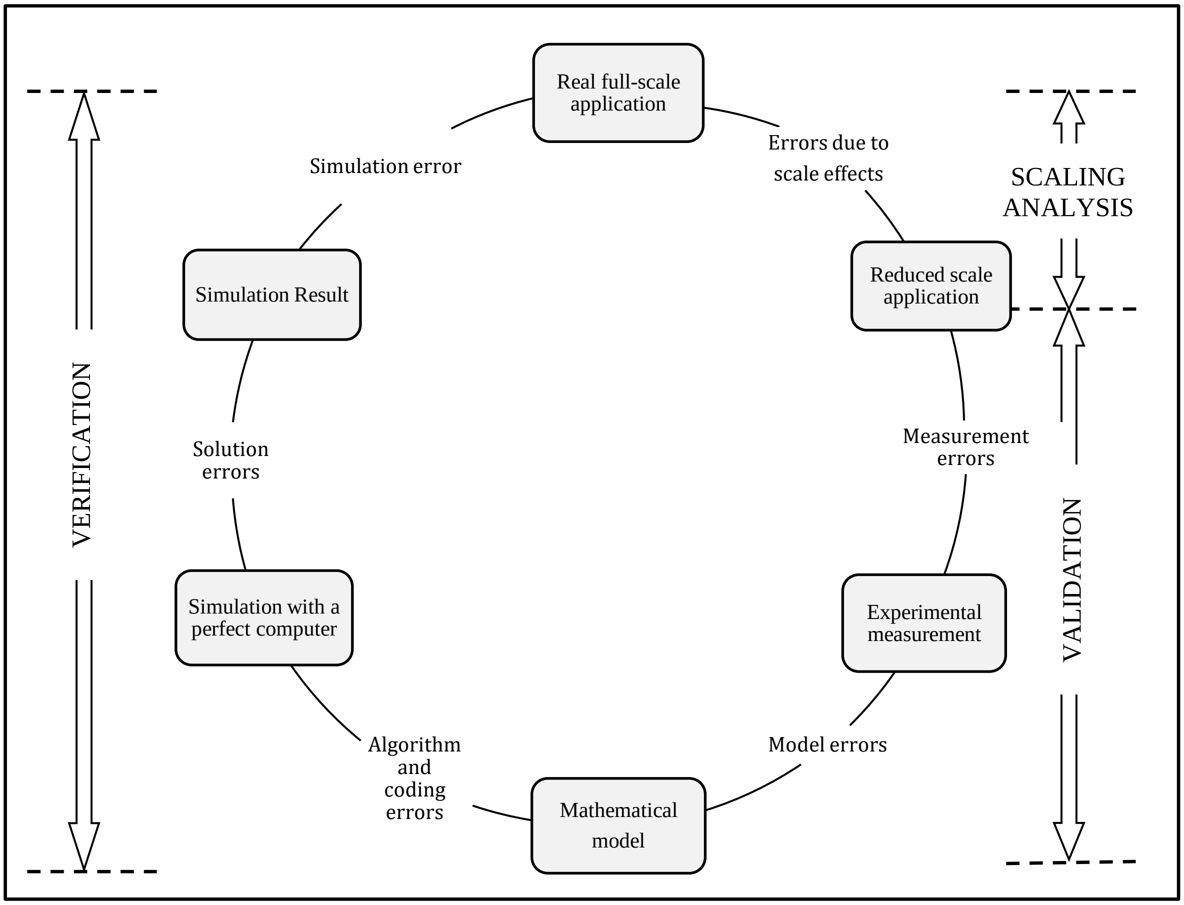


Figure 3.2: Illustration of verification, validation, sources of error and scaling

Verification can be divided into three parts,

- Software quality assurance,
- Code verification
- Solution verification.

(1) Software quality assurance is performed on the basis of three types of tests: Unit testing, Regression testing, and Benchmarking. These tests are described below:

- Unit Testing: Units test involve simple test problems to check if small parts or units of the code are working correctly.

- Regression testing: It is a type of software quality check which verifies that the code did not undergo any unintended change due to any modification in the source code.
- Benchmarking: It is also part of software quality check. Benchmarking is performed by code-to-code comparison. It involves comparison of simulation of an identical problem on different simulation codes.

(2) Code Verification: Code verification can be described as the process of agglomeration of evidence to evaluate the assertion (or claim) that the numerical algorithms are implemented correctly inside the code [11]. Code verification is focused on,

- Debugging the source code
- Eliminating errors in the numerical algorithm.

Code verification encompass discretization error quantification, convergence study, and order-of-accuracy tests.

(3) Solution verification: Solution verification can be described as the process of agglomeration of evidence to evaluate the assertion (or claim) that the solution to the mathematical functions represented in the simulation is correct (or correct enough) when compared with the true solution of those same functions [11].

Validation is a process of agglomerating the evidence to evaluate the assertion (or claim) that the numerical simulation of the mathematical function can predict a real physical quantity [11].

Code's V & V (for nuclear reactor applications) can be described as a confidence-building process. It is an iterative process that requires continuous exploration, learning, and assessment. A successful V & V process should address all sources of uncertainty and provide sufficient evidence for reliable and robust decision making.

The V & V process involves different activities, which includes the PIRT process, data collection, and data applicability analysis, pyramid formulation, model testing, evaluation, etc. Figure 3.3 provides an illustration of the validation process and related activities using a series of hierarchical pyramids for phenomena, model, and data. The phenomenology pyramid helps in aligning the code and data pyramid for validation assessment. Formulation of pyramid involves different supporting activities, like PIRT, evidence (model and data) collection, classification and characterization, database management, etc.

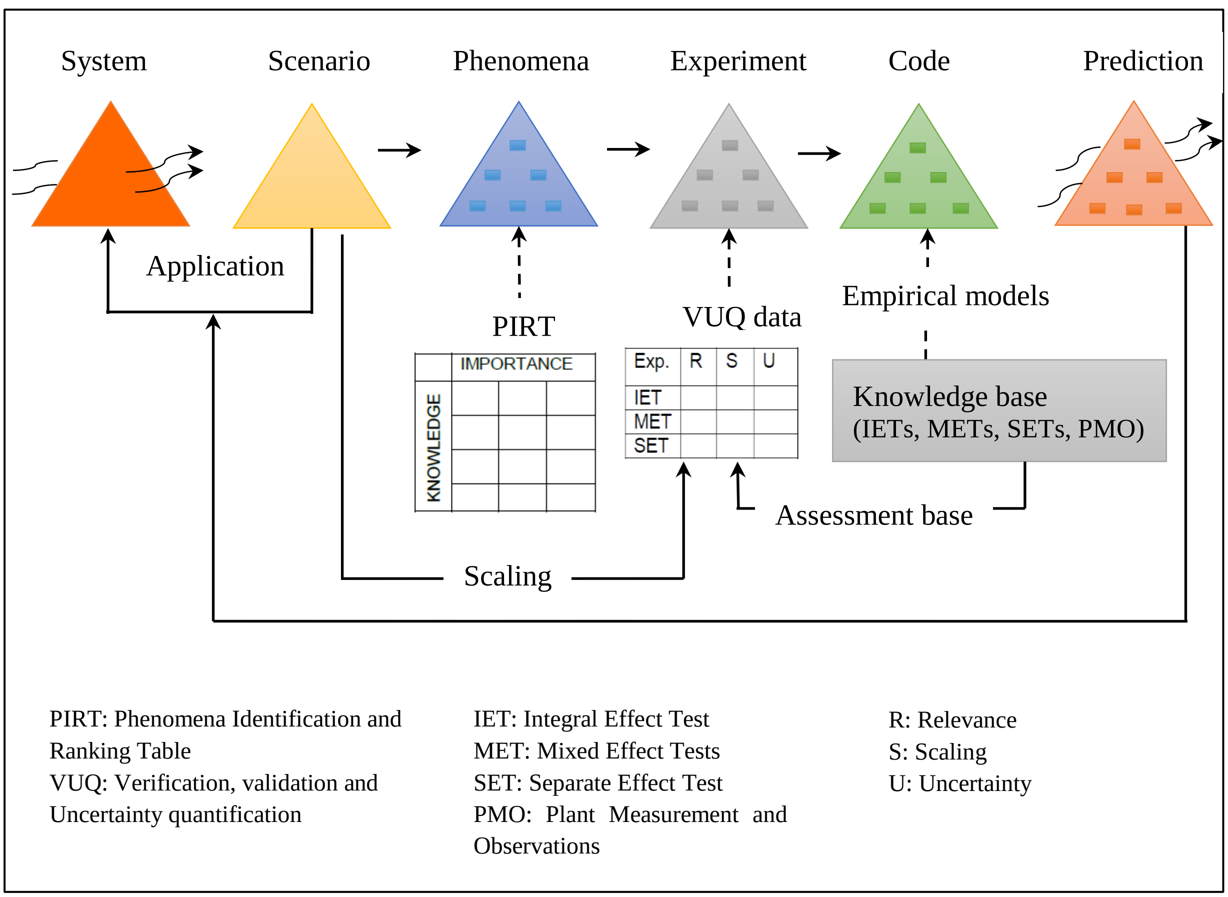


Figure 3.3: Code validation process and related activities

Adopting a similar approach, the process of verification of a code can be represented using individual pyramids for solution feature, code verification (includes SQA) and solution verification (see Figure 3.4 for illustration). Pyramid for solution feature is based on the solution type identification and ranking table (STIRT). STIRT consists of a list of distinctive features of the code that needs to be examined by different verification tests. This list is prepared based on the opinion of the SME. This pyramid serves as a guiding structure for code verification and solution verification. The process of verification of code and associated activities using individual pyramids is shown in Figure 3.4.

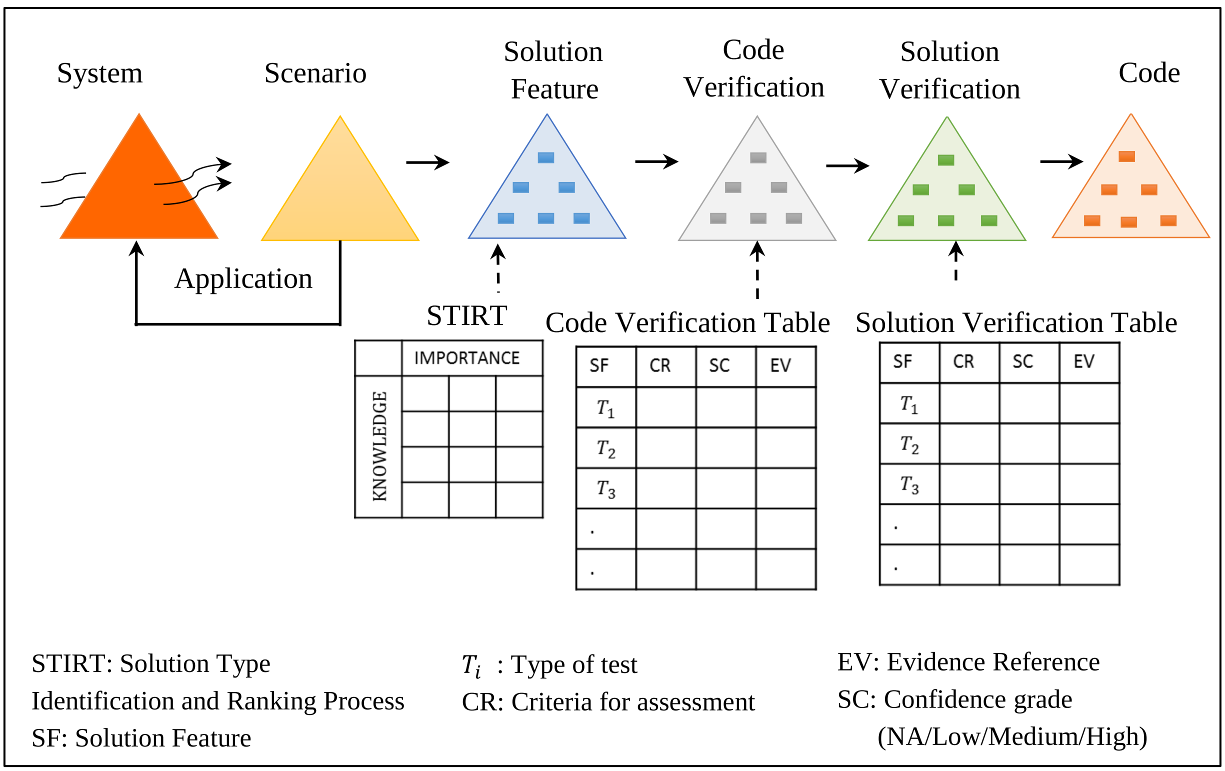


Figure 3.4: Code verification process and related activities

Along with the evaluation of verification and validation results, assessment of rigor and quality of the supporting processes and activities becomes very important. Therefore, a systematic process for maturity assessment is required. The formalization of maturity assessment process not only helps in assessment of code but also helps in streamlining the V & V activities and enhances clarity and traceability of information.

3.4. Research approach

PCMM is a decision model in which different attributes related to VVUQ of code are assessed using different maturity assessment sets. The maturity assessment set is formulated based on the application's consequence. The Analytic hierarchy process adopts the hierarchical approach to decision making. The decision schema for development of the decision model in the proposed framework is based on the architecture of PCMM (for maturity set and assessment criteria) and Analytic hierarchy process (for decision hierarchy).

Figure 3.5 illustrates the research approach for formalizing the decision model for predictive capability maturity assessment. Formulation of the decision model is performed using the argumentation techniques (Goal structuring notation). Each decision attribute/sub-attribute is formulated as a claim, where the degree of validity of the claim (attribute's assessment) is expressed using different maturity levels: <decision attribute, maturity level>. The strategy for decomposition (reasoning step or argument) for breaking down attributes (claims) into sub-attributes (sub-claims) is provided based on the CSAU/EMDAP process.

The argument model for decision is designed using GSN and transformed into a computable network (Bayesian network) to support evidence-based quantitative maturity assessment of all the attributes and sub-attributes in the decision model. The evidence to support the claims and sub-claims in the decision model are provided by the objective data obtained from different V & V activities. Subjective information based on expert's input is assimilated as subjective probability of assessment grades and conditional probabilities of the attributes in the Bayesian network. Depending on the consequence of the decision, maturity level for the target is decided. Evaluation is performed by comparing the target level for each decision attribute with their achieved level based on the available evidence. The formulation of the framework for PCMA based on the closed loop decision process [33] (discussed in section 2.4) is presented in the subsequent section.

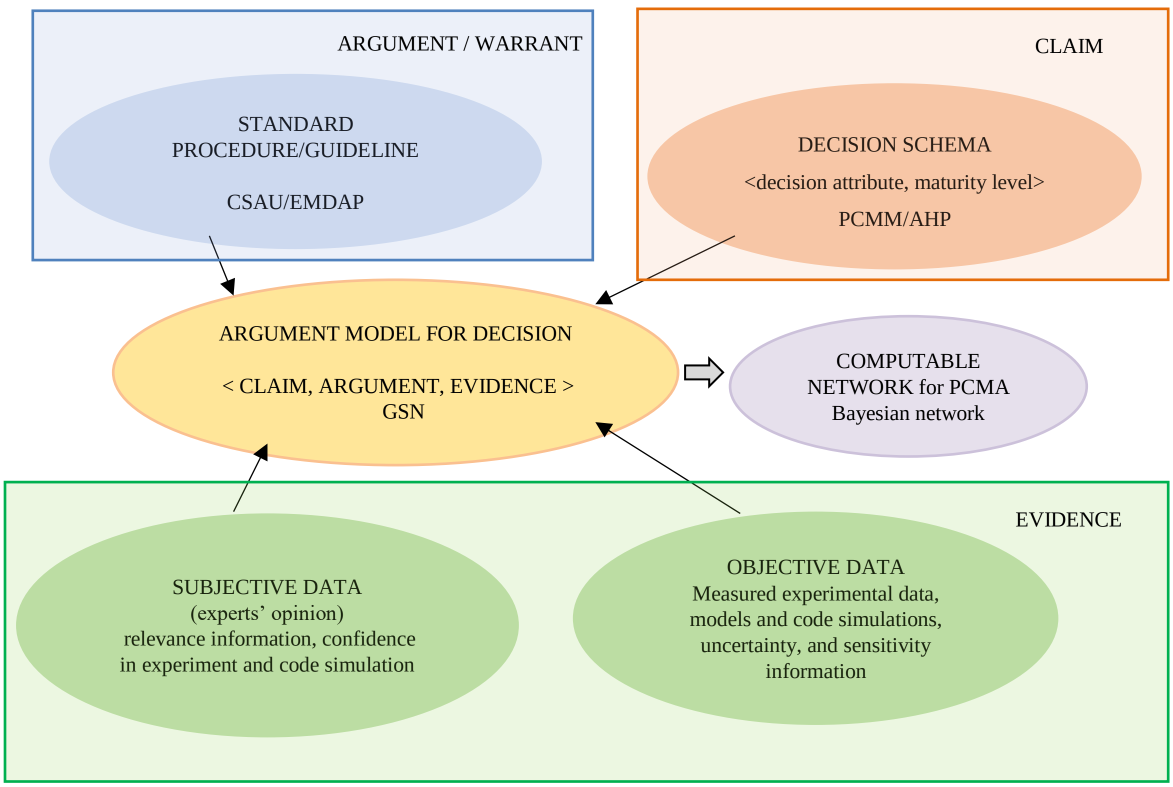


Figure 3.5: Formalizing the decision model for code prediction

3.5. Formulation of the assessment framework

This section illustrates the framework for predictive capability maturity assessment. The primary objective of this framework is to provide structural knowledge representation, detailed evidence incorporation and maturity assessment of a simulation code for an intended application. As validation is the most important attribute for a code's maturity assessment, the framework is particularly focused on validation assessment of code. The conceptual schematic of the framework is shown in Figure 3.7.

The proposed framework is divided into six sections, shown below (see Figure 3.7 for illustration):

I. Preprocessing for the framework development

II. Structural Knowledge representation

III. Classification and characterization of evidence

IV. Formulation of the decision model

V. Evaluation and interpretation of results

VI. Refinement

3.5.1. Preprocessing for the framework development

This section describes the preprocessing requirement for the development of the framework for predictive capability maturity assessment. Figure 3.6 illustrates the steps involved in this process. These steps are based on the Evaluation model development and assessment process (EMDAP) [2].

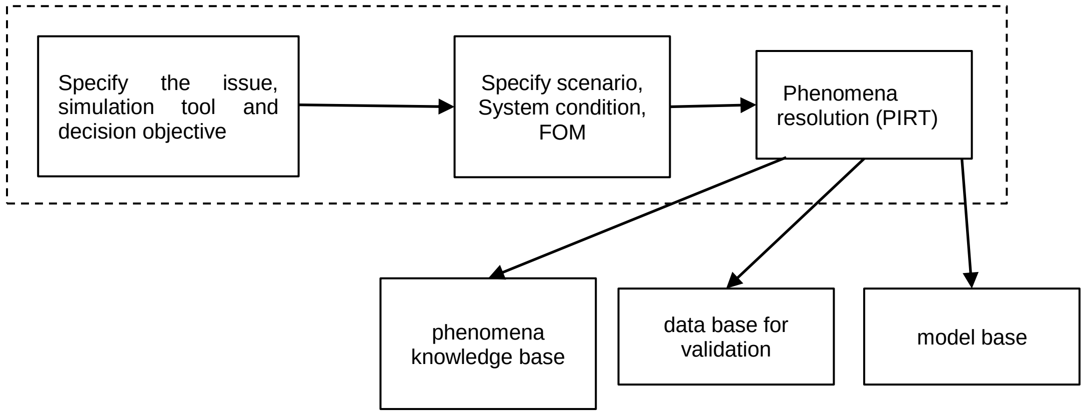


Figure 3.6: Preprocessing requirement for development of the framework for PCMQ

The first step is the specification of the issue (i.e. description of the problem of interest) and decision objectives. Specification of the decision objective is very important at the beginning of the code development process. As discussed before, NE codes are used to support different decisions related to design, operation, performance, and safety analysis of nuclear reactor system. The architecture and complexity of the M & S tool, and the rigor and depth of the code assessment process is dependent on these decision objectives. In PCMM also, the criteria for code assessment is based on the intended use or decision objective of the M & S tool (see section 2.3.4)

The second step involves identification and specification of scenario (transient or steady state), system condition and FOM (see definition of FOM in section 1.3). Depending on the nature of problem., multiple FOM may be specified. In the case of transient or accident scenario, the scenario is decomposed into time phases based on the dominant mechanism or process. Relevant phenomena are identified and ranked using the PIRT process in step 3. Based on the PIRT process, a knowledge base is created where the document and excel files related to PIRT process are maintained. PIRT also helps in identifying the data for validation. Following the PIRT process a database for validation of code and model base with a collection of models (closure models) is created.

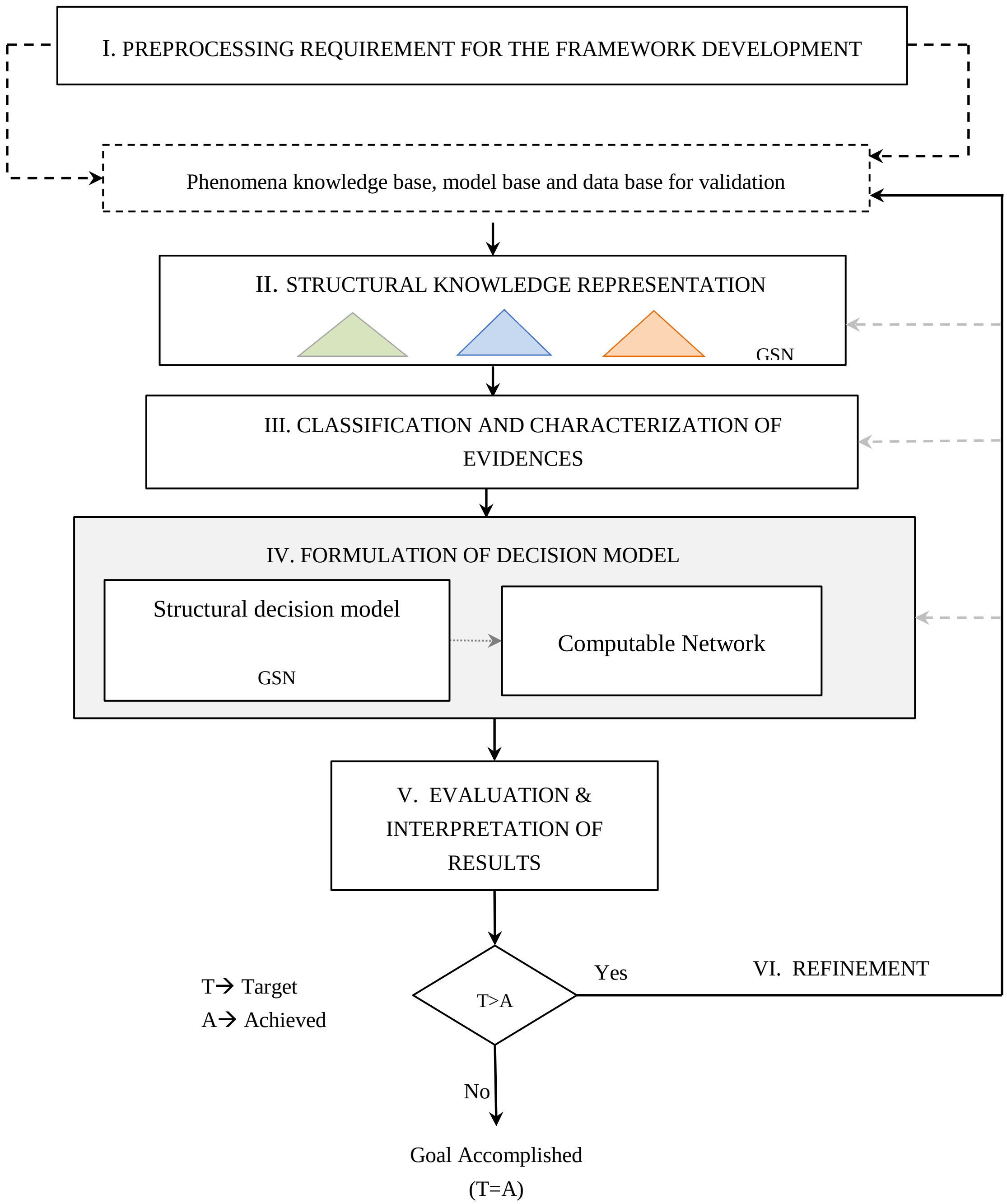


Figure 3.7: Conceptual Outline of the framework for PCMQ

3.5.2. Structural knowledge representation

Due to the synergistic effects of complex physical interactions and multiple scales, it becomes important to adopt a hierarchical approach for structural organization of information for code assessment. EMDAP also emphasizes the use of hierarchical system decomposition (see Figure 2.1) in the development of evaluation model for nuclear reactor applications and treats information related to phenomena, data, and model separately in its three elements:

- Element 1 (Establish requirement for evaluation model capability) → Phenomena hierarchy
- Element 2 (Develop assessment base) → data hierarchy
- Element 3 (Develop evaluation model) → model hierarchy

The importance of hierarchy is clearly visible in all the four elements of the EMDAP process (see Figure 2.1). Hierarchical representation adopted in this framework is based on the three pyramids approach proposed in the CASL V & V plan [83]. It consists of three pyramids (see Figure 3.8 for illustration):

- PIRT based Phenomenology pyramid (PP)
- Code system-based model pyramid (PM)
- Validation experiment-based data pyramid (DP).

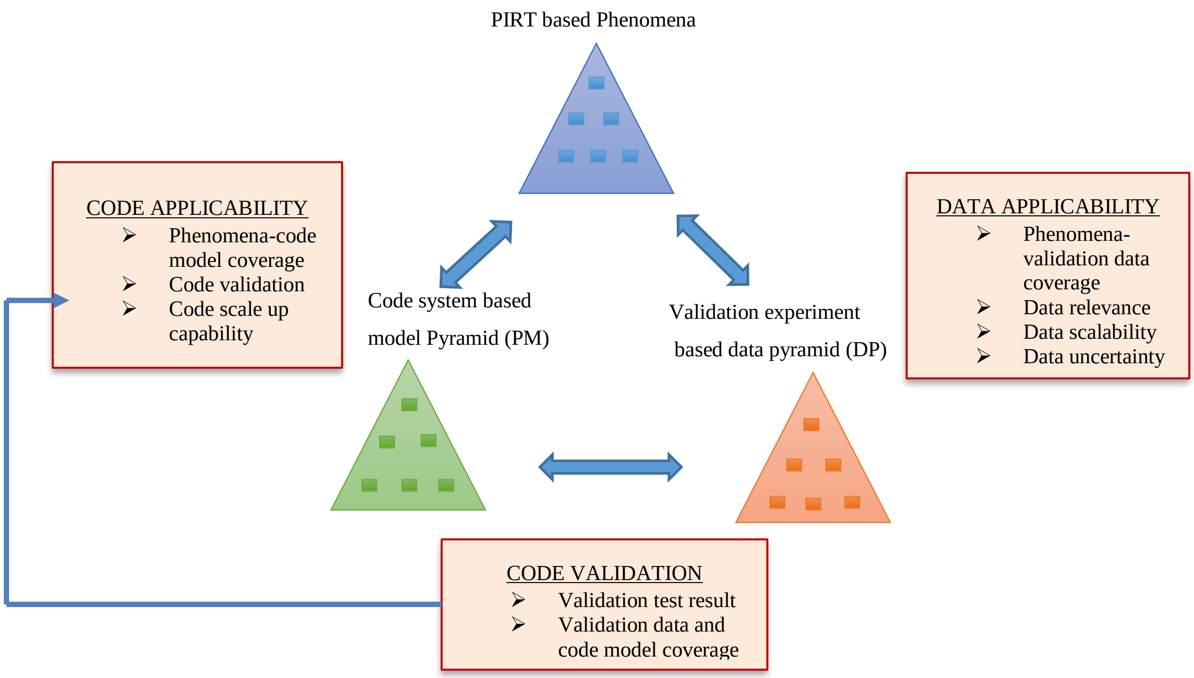


Figure 3.8 : Illustration of three pyramid approach for code validation

Phenomenology pyramid serves as a blueprint for data and model pyramid. It guides the formation of data pyramid (PE) and model pyramid (PM) based on different phenomena in the phenomenology pyramid (PP). The pyramid consists of J phenomena, where the j^{th}phenomenon is given by $\boldsymbol{P}_j,\ j = 1,2,3 \dots J$. Each j^{th} phenomenon, $\boldsymbol{P}_j$ in the pyramid is characterized by [31],

- a set of the quantity of interest represented by $[\boldsymbol{QOI}_{jk}]$, $k = 1,2, \dots K$
    - a set of system condition $[\boldsymbol{SysCond}_{jkn}]$, $n = 1,2, \dots . N$, where N is the number of system condition, each condition is characterized by,
        - a set of $\boldsymbol{M}$-dimensional and/or non-dimensionalized parameters, each has an operating range for the given application (scenario)

$$\boldsymbol{Par}_{ijkm,n} = \left(\boldsymbol{Par}_{jkn,m}^{min}, \boldsymbol{Par}_{jkn,m}^{max}\right),\ m = 1,2 \dots . . M$$

Structured knowledge representation of pyramids and other entities in the framework is obtained by using Goal structuring notation. We use Adelard's Assurance and Safety Case Environment (ASCE 4.2) to build the GSN network. As discussed in section 2.7, GSN is an argument modeling technique used for graphical representation of assurance argument in safety cases. There are several analogies between safety case and code prediction and validation (see Table 3.1), particularly from the perspective of "nature of problem". A safety case is a structured argument, supported by evidence, which intends to justify that a system is acceptably safe. Similarly, code validation can be described as the "confidence argument" supported by evidence (model and data) that justifies the claim that code provides reliable prediction in the extrapolation domain.

GSN can be described as a goal-oriented technique for decomposing and structuring complex problems. It serves as an ideal tool for representing the hierarchical pyramids (Phenomenon pyramid, model pyramid, and data pyramid) and structuring the decision model for code validation. It can be used at any stage of analysis and suits well for iterative verification and validation process.

As discussed in section 2.2, the "Phenomena" in the PIRT process is treated as a general terminology and can be anything that impacts the FOM. It equivocally includes mathematical or engineering approximations, system conditions, physical processes, reactor parameter as phenomena in the PIRT process [23]. GSN provides a formal structure to PIRT by structuring information using explicit classifiers (like assumption, justification, context, solution, etc.). In this way, it provides a formal structure to the PIRT process. The "Importance" and "Knowledge" information in the PIRT process is incorporated in the GSN trees using the indicators shown in

Figure 3.9. Additionally, we can use priority indicators shown in Figure 3.10. These indicators are facilitated into the GSN using the node property dialog box in the ASCE.

Table 3.1: Analogy between safety case and code prediction and validation.

Attribute	Safety case	Code Prediction and Validation
Nature of "Problem"	Safety is not observable	Extrapolation to operating regime beyond validation domain
Objective	Consolidate reliability of safety statement	Consolidate reliability of code prediction
Framework	Methodology for decomposition and integration GSN, CAE networks	Predictive Capability Maturity Model (PCMM), CSAU, EMDAP
Evidence	Body of evidence Select set of evidence (Hazard log, PRA, fault tree, event tree, etc.)	Code calibration and validation Benchmark against experiments
Quantity of interest	Safety margin	Quantification of uncertainty due to modeling error, input parameter uncertainty, scaling, measurement error, etc.
Knowledge base, representation, and management	Code validation, code applications (scenarios, Experiments, PMO)	SET, MET, and IET Numerical experiments

GSN offers structure, clarity, and traceability to the maturity assessment process and facilitates systematic evidence incorporation and integration for confidence assessment.

Figure 3.11 depicts a simple example of code validation in GSN using the above-mentioned indicators and functionalities. It should be noted that prior to code validation assessment all the necessary verification activities on the code have been completed. The top goal contains the claim – "Code X is suitable for predicting the application XX." It clearly states the objective of the process. Strategy block contains the argument that states how this goal can be resolved. Therefore, the argument in the strategy block is based on the validation results and data applicability analysis. In this example, data applicability analysis has not been completed. Therefore, undeveloped and uninstantiated entity indicator is used to indicate an incomplete goal. Completion of this goal is important to support the top goal. Therefore it is marked with high importance, and high priority indicator flags.

It is important to check the consistency of the GSN network in order to conform with the rules of creating a GSN network. Table 3.2 contains different rules for checking the consistency of the network. ASCE 4.2 facilitate automatic consistency check of the GSN network. It enlists all the error with node number and severity of the error.

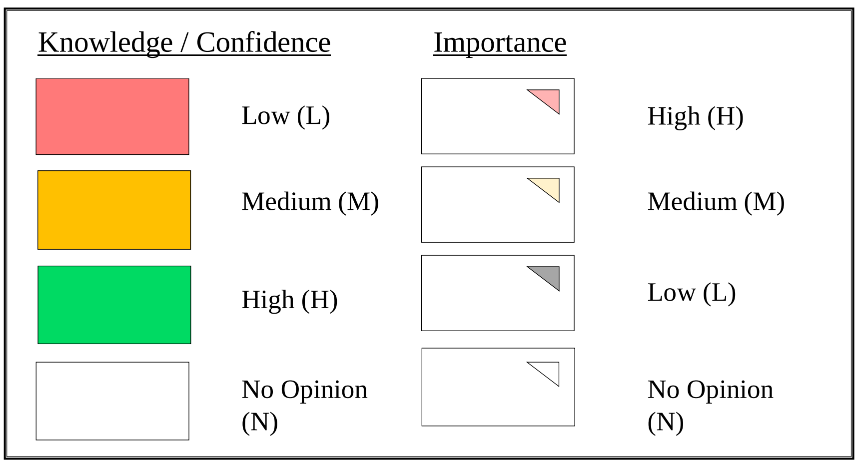
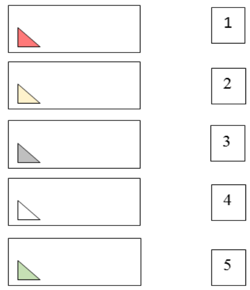


Figure 3.9: Knowledge and Importance indicator used in GSN

Priority



Importance	Knowledge	Priority score
H	L	1
H	M	2
H	H	
M	L	3
M	M	
M	H	4
L	L	
L	M	5
L	H	

Figure 3.10: Priority Indicator used in GSN

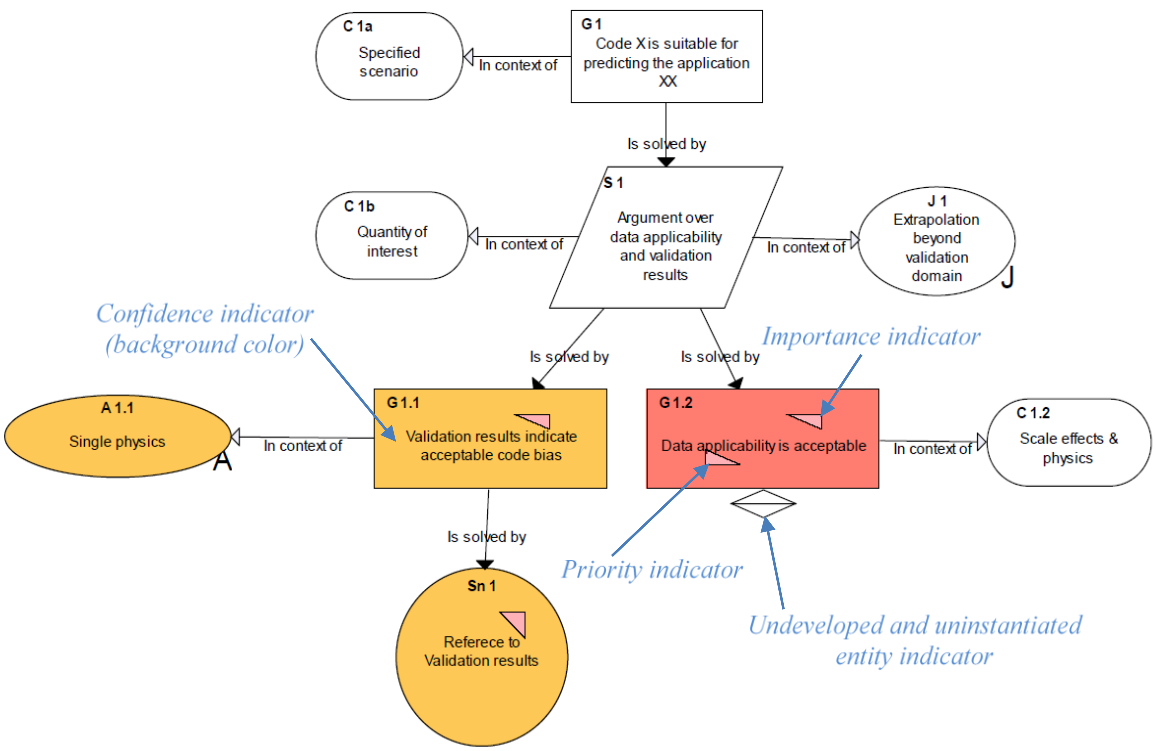


Figure 3.11: Example of using GSN for Validation

Table 3.2: Rules for checking consistency of GSN network in ASCE [95]

S. No.	Description	Severity
1	All node should eventually have status complete	1
2	Strategies must be solved by atleast one sub-goal	2
3	Goal must be solved by atleast one sub-goal, strategy or solution	2
4	Eventually, all option nodes should be removed	2
5	Eventually, all n-iteration and 0/1 choice links should be removed	2
6	Solutions, Assumptions, Justifications, and contexts should only have incoming links	2
7	Solutions must not be solved by anything	4
8	There should be only one top level node (excluding notes)	4

The GSN supports modular architecture. Modular architecture helps in managing individual GSN networks into separate modules. The GSN community standard [56] describes all the elements of modular GSN extension. Figure 3.12 shows a simple example of modular GSN. In this example, the main goal or claim is decomposed into three sub-goals. While the first sub-goal is resolved in the main module, the second and third sub-goals are resolved in different modules. These sub-goals are represented as *Away goal* to indicate that these goals are resolved in different GSN modules. *"Away goal:* 1.2*"* and *"Away goal:* 1.3*"* are expanded into separate GSN network in *"module 1.axml#" and "module 2.axml"(.axml#* is the file extension used by ASCE). The modules in ASCE are connected via hyperlinks, thereby supporting the formulation of layered architecture in GSN.

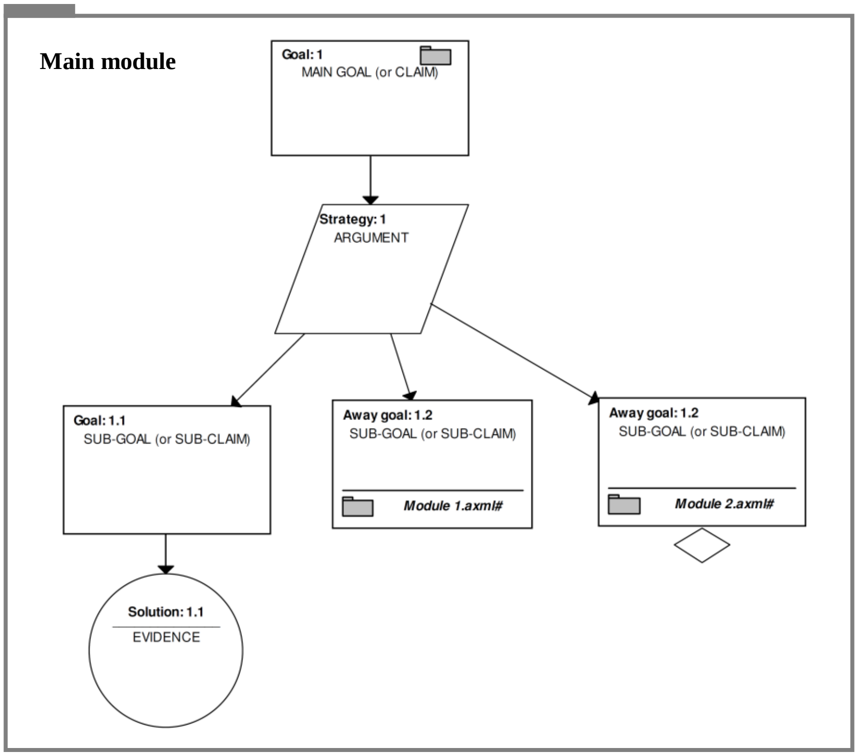


Figure 3.12: Modular GSN extension

The process of developing a Phenomenology pyramid involves several levels of information abstraction (see Figure 3.13). Therefore, it becomes important to maintain clarity and traceability as we integrate information in the GSN network for phenomenology pyramid. Traceability of information is maintained using the hyperlink and node description dialog box embedded with each node in the GSN network (e.g. see Figure 3.14). Excerpt from relevant excel sheets, PDF documents or word files can be captured inside these dialog boxes. We have provided extracted information or hyperlink to detailed documents inside each dialog box to maintain traceability of literature associated with a node. In this way, all the information corresponding to different levels of abstraction can be embedded inside the GSN network.

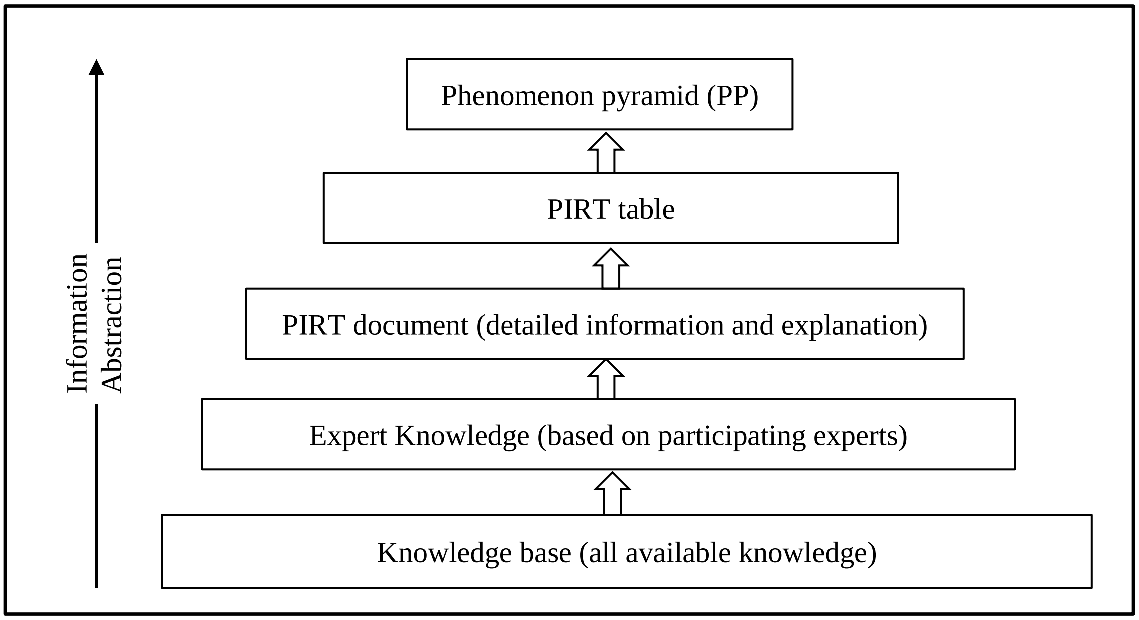


Figure 3.13: Illustration of information abstraction in PIRT/ phenomenology pyramid (PP)

Development of phenomenology pyramid is a crucial step in the development of the framework for V & V process, and phenomena resolution using PIRT is the key step in the evolution of the phenomenology pyramid (PP). All phenomena identified in the PIRT process comes from expert opinion. However, we can always trace back the relevant literature (analytical and experimental data) that forms the basis of expert opinion and judgment. These references provide evidence for phenomena presented in the PIRT process.

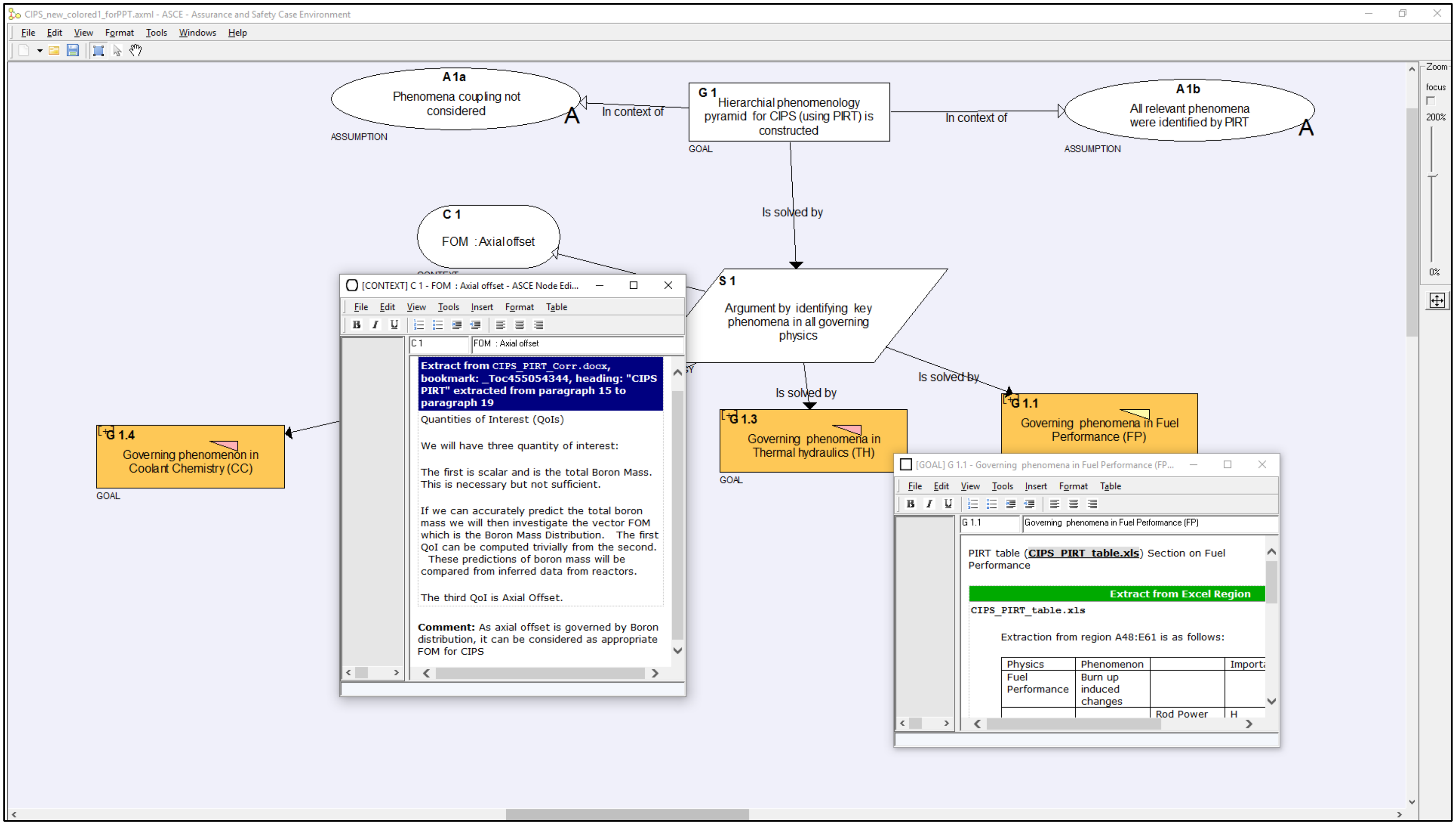


Figure 3.14: Document extraction and hyperlink functionality in ASCE 4

3.5.3. Classification and characterization of evidence

The decision model for the assessment of the predictive capability of a code is based on two types of attributes: (i) Direct maturity evaluation attributes, (ii) Process quality assurance factors. The classification and characterization of evidence is based on these two types attributes.

3.5.3.1. Direct maturity assessment attributes

Direct maturity evaluation attributes are those attributes that directly impact the decision regarding the adequacy of an M & S tool. Direct maturity evaluation attribute in our framework are based on the PCMM attributes:

- Representation and geometric fidelity (RGF)
- Physics and material model fidelity
- Verification
    - Software quality assurance
    - Code verification
    - Solution verification
- Model Validation
    - Separate effect test validation
    - Integral effect test validation
- Uncertainty quantification and sensitivity analysis.

These attributes may be further divided into sub-attributes depending on the depth and rigor of the maturity assessment process. Any evidence that supports the direct maturity evaluation attributes is regarded as the direct evidence. The direct evidence for all direct maturity evaluation attributes are assessed by using capability grades (or maturity levels).

Validation is considered as an important element in the predictive capability maturity assessment of codes. Therefore, we need detailed and in-depth assessment of validation (which includes separate effects model validation and Integral effect model validation). The direct evidence for validation assessment are categorized based on two sub-attributes: (1) Validation results, (2) Data applicability

(1) Validation result: Evidence associated with validation result can be characterized based on two sub-attributes: (a) Coverage, (b) Validation test result.

(a) Coverage [C]: Based on the phenomena in the phenomenology pyramid (PP) an experiment-based data pyramid (DP) and code-based model pyramid (MP) is constructed. Coverage information is obtained by comparing the range of parameters for each phenomenon in the phenomenology pyramid (PP) with the range of parameters for the corresponding model in the model pyramid (MP) and data in data pyramid (DP). In this way, coverage information has three components [31]:

- [CMP] - phenomenological coverage of phenomena [$\boldsymbol{P}_j$] in phenomenology pyramid (PP) by models in the code-based model pyramid (MP),
- [CME] - coverage of models [M_x] in the code-based model pyramid (MP) by data [E_x] in experiment-based data pyramid (DP),
- [CEP] - phenomenological coverage of phenomena [P_j] in phenomenology pyramid (PP) by data [E_x] in experiment-based data pyramid (DP).

(b) Validation test results (VTR): Starting from the mathematical formulation to the numerical solution of equations, there are several approximations and uncertainty sources involved at each level of the code formulation. These sources of uncertainty include model-form uncertainty, model parameter uncertainty, uncertainty due to incomplete knowledge of initial and boundary condition, uncertainty due to approximations in numerical simulation and discretization error, etc. Together, these sources of uncertainty can produce a large deviation in the code prediction. This deviation can be characterized by using different validation metrics. Validation metrics need to be computed for the entire set of the quantity of interest $[QOI_{jk}]$, that characterizes the phenomena P_j. Different validation metrics can be adopted based on the nature of the code and data. Broadly the validation metric can be divided in two categories [96] :

- Deterministic validation metric (RMSE, bias, etc.)
- Probability-based validation metric (hypothesis tests, probability box, confidence interval, information theoretic measure)

Maupin et. al.[96] provide a detailed description of different metrics that can be employed as validation metrics.

(2) Data applicability: Evidence associated with data applicability can be characterized based on two sub-attributes: (a) Scaling and (b) Data uncertainty. These attributes are based on the R/S/U grading system proposed by Dinh [97], [98] for assessing the quality of experiment.

(a) Scaling [S]: Cost and safety are the two constraints that restrict new data acquisition and experimentation for validation. Consequently, relevant data for validation is carefully selected from the available databases. Scaling analysis helps us in characterizing the quality of data with respect to the application of interest. Scaling information reflects the degree of similarity between phenomena in phenomenology pyramid and experiment in data pyramid on the basis of geometric similarity, material scaling and physics scaling (Dynamic and kinematic similarity). We divide scaling information into two parts:

- Relevance [R]
- Physics scaling [PS]

Relevance is determined based on geometric similarity and material scaling. It determines the degree of applicability of data "*based on the preconceived view of phenomenology/process"[97]*. Physics scaling reflects the degree of similarity between phenomena in phenomenology pyramid and experiments in data pyramid on the basis of physics scaling (Dynamic and kinematic similarity). It determines the gap between the test facility and reactor behavior (phenomena at reactor conditions).

(b) Data uncertainty [U]: Data uncertainty consists of uncertainty in the measured data due to instrumentation errors and limited resolution of measurement instruments. It may also include the effect of data acquisition and data processing.

Assessment of validation attribute is based on a capability grade (or maturity levels) shown in Table 3.3.

Table 3.3: Description of capability grade for different validation attribute

Attribute	Capability grade (or maturity level)				
	4	3	2	1	0
Relevance [R]	Very High (direct)	High	Medium	Low	NA/TBA
Physics scaling [S]	Prototypic (full scale)	Adequately scaled	Medium	Inadequately scaled (large distortion)	NA/TBA
Uncertainty [U]	Well Characterized	Characterized	Medium	Poorly characterized	NA/TBA
Coverage [C]	Very High (more than 90% coverage)	High (between 60% to 90% coverage)	Medium (between 25% - 60 % coverage)	Low (less than 25% coverage)	NA/TBA
Validation test result [VTR]	Very High	High	Medium	Low	NA/TBA

3.5.3.2. Process quality assurance factors

Comprehensive confidence assessment requires consideration of not only the direct maturity assessment attributes but also several secondary factors related to process quality assurance (PQA). These factors indirectly affect the confidence assessment in VVUQ process. The evidence supporting the PQA factors are regarded as indirect evidence. NASA's credibility assessment scale also explicitly consider "use history," "M&S management" and "people qualification" in its assessment as secondary evidence [4]. Process quality assurance factors that impact confidence in VVUQ process are described below:

- Execution of standard procedure/guideline (EMDAP/CSAU) in the VVUQ process.
- The method of analysis, efficiency of tools and techniques.
- Breadth and depth of expert knowledge (i.e., domain knowledge, experience and expertise of personnel associated with various activities of code verification and validation)

We identify two types of PQA factors in our framework: (i) PQA factors related to PIRT/ phenomenology pyramid, and (ii) PQA factor related to evidence assessment process

(1) PQA factors related to PIRT/Phenomenology pyramid

Decomposition of PQA for the PIRT process (or phenomenology pyramid) is based on the three secondary factors described by Nair et al. [52]. These factors are based on process, personnel, and tools/techniques involved in the formation of PIRT/phenomenology pyramid (PP). Figure 3.15 shows different attributes that impact assessment of these factors. To perform confidence assessment for each node, a questionnaire can be created, and responses of SME can be recorded. A sample questionnaire corresponding to process quality assurance factor for phenomenology pyramid (PP) is presented in Table 3.4. The response can be documented using different grades like, “NA”, “Low”, “Medium”, “High.”.

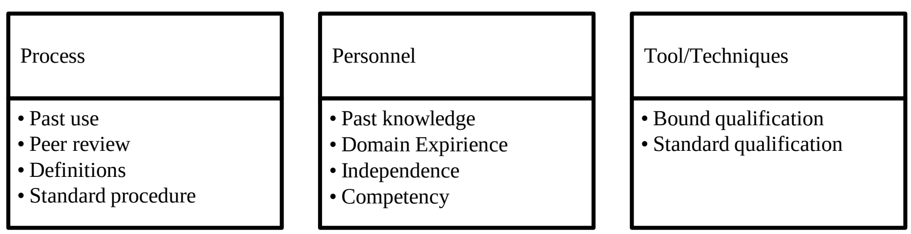


Figure 3.15: Process quality assurance factors related to PIRT/phenomenology pyramid(PP)

Table 3.4: PQA factors related to Phenomenology Pyramid

Process factors	
Past use	What is the confidence in the past use of the technique used for complexity resolution?
Peer review	What is the confidence that adequate peer review of the PIRT has been carried out? (NA/L/M/H)
Definition	What is the confidence that all phenomena (in the PIRT process) have been clearly defined/described in the context of the Application of interest? (NA/L/M/H)
Standard Procedure	What is the confidence that standard procedure as specified in the EMDAP process has been followed? (NA/L/M/H)
Personnel factors	
Past knowledge	What is the confidence in the past knowledge about the phenomena and processes identified by the PIRT process?
Domain experience	What is the confidence that the persons involved in PIRT process have adequate domain experience? (NA/L/M/H)
Independence	What is the confidence that the people involved in the PIRT process are independent (in the context of the domain of expertise)? (NA/L/M/H)
Competency	What is confidence in the competency of the personnel involved in the PIRT process? (NA/L/M/H)
Tool/Technique	
Bound qualification	What is the confidence in the technique (phenomenon pyramid and GSN) adopted for phenomenon decomposition for the application of interest? (NA/L/M/H)
Standard qualification	Does the pyramid-based technique comply with the regulatory standard (i.e. EMDAP)? (NA/L/M/H)

(2) PQA factor related to evidence assessment process

Evidence are the backbone of the decision model; therefore, PQA for evidence assessment process (EAP) becomes important. Process quality assurance factors related to evidence assessment process are based on three factors: (a) Level of detail of evidence, (b) Credibility of evidence, (c) Tools and techniques. These three factors for PQA of validation evidence assessment (VEA) process is described below.

(a) Level of detail of evidence: This factor is based on the level of detail and completeness of the evidence used for the validation assessment. Level of detail is determined based on the following four grades/levels:

- Gap (G)→Gap refers to an undeveloped entity (model needs to be developed or data do not exist).

- High-level composition (HLC) → HLC refers to global statement or any activity related to the VVUQ code.
- Medium-level composition (MLC) → MLC refers to specific task that support high level evidence.
- Low-level composition (LLC) → LLC refers to performance or test detail and results.

(b) Credibility of evidence: The credibility is based on the people who assessed the validation evidence. Credibility is assessed based on the following four grades:

- No assessment(NA)→ No assessment refers to the condition when evidence are not assessed.
- Initial author assessment (IA)→ Initial author assessment is based on the preliminary assessment by author.
- Specialist assessment (SA)→ Specialist assessment is based on a thorough assessment by subject matter expert.
- Peer-reviewed assessment (PA)→ Peer-reviewed assessment is based on independent peer review by a group of experts.

(c) Tools and techniques: This factor is based on the type of technique used for assessment of validation data and test results. This factor consists of two parts: (i) Scaling technique, (ii) Validation technique.

(i) Scaling technique: Scaling analysis is a crucial element in validation assessment as it warrants the applicability of experimental data for real reactor application (full-scale). Even though different methodologies for scaling assessment have been developed over past few decades, scaling analysis in practice is still a daunting task. Based on the quality of assessment techniques we have the following four levels to designate the type of scaling technique:

- No assessment (NA) → This level is used when no scaling assessment is performed.
- Observation (O)→ This level is used when scaling assessment is based on the observation only.
- Selective dimensionless group (SDG)→ This level is used when scaling assessment is based on the comparison of selective dimensionless group for test facility and real application.
- Scaling methodology (SM) → This level is used when scaling assessment is performed by using a proper scaling methodology (e.g. hierarchical two-tier scaling mythology[14], fractional scaling analysis [15], etc.).

(ii) Validation techniques: Different types of validation technique can be adopted for assessing the validation test result. Based on the type of assessment technique, we have four different levels to designate the validation technique.

- No assessment (NA) → This level is used when no validation assessment is performed.
- Point estimate (PE)→ This level is used when validation assessment is performed by the deterministic assessment based on point estimate of response quantity of interest.
- Deterministic and graphical assessment (DGA) → This level is used when validation assessment is based on deterministic validation metric and comparison of graphical results or patterns.

- Probabilistic validation metric (PVM)→This level is used when scaling assessment is based on probabilistic validation metrics (e.g. hypothesis tests, probability box, confidence interval, information theoretic measure, etc.).

Table 3.5 shows different factors (or attribute) which are used for PQA of VEA process, and their associated grades.

The grading scale for 'level of detail of evidence' and 'credibility of evidence' is same for PQA of evidence assessment process (EAP) for other attributes. However, grading scale for 'tool/techniques' depends on the attribute.

Table 3.5: Process quality assurance factors related to validation evidence assessment with associated grade

Attribute	Grade			
	3	2	1	0
Level of detail of evidence [D]	LLC	MLC	HLC	Gap
Credibility of evidence [C]	Peer reviewed (PR) assessment	Specialist assessment	Initial author assessment (IA)	NA
Scaling technique [ST]	Scaling methodology	Selective dimensionless groups	Observation	NA
Validation techniques [VT]	Probabilistic validation metric	Deterministic and graphical	Point estimate	NA

3.5.4. Formulation of the decision model

The fourth segment of the framework corresponds to the formulation of the decision model for maturity assessment. This segment can be described using three steps:

Step 1: Formulate the decision model using GSN

This step involves structural representation of the hierarchical decision model using GSN. It includes criteria, sub-criteria, evidence incorporation, and dependency relations for decision analysis. The higher-level decision attributes are formed by the elements (attributes) described in the PCMM. As we are focused on code validation, the decision attribute and sub-attribute are based on the different validation attribute discussed in the previous section. Each decision attribute is defined as a claim in the *Goal* blocks (or nodes) of the GSN network. The degree of validity of that claim is expressed using a maturity scale based on different grades (e.g., NA, Low, Medium, High). The concept of maturity scale in this framework is based on the confidence grade adopted in the assessment of safety cases [92], [52].

Step 2: Transforms the GSN based decision model into a confidence network

Transformation of GSN to a computable network is facilitated by identifying the basic sources of uncertainty in the argument model because uncertainty directly affects the degree of confidence. Guiochet et. al [93] explains the transformation of a GSN tree to a confidence network (Bayesian network) by annotating different nodes in the GSN tree based on the uncertainty associated with those nodes. We adopt a similar approach; however, the confidence network in our formulation is based on the *Goal* and *Sub-goal* nodes only. The *Goal* nodes in the GSN tree consist of claims and sub-claims. Each claim, sub-claim is associated with uncertainty. Therefore, the transformation of GSN to a confidence network is performed by transforming the GSN tree into a network that consists of only the goals and sub-goals. We term this network as the reduced GSN network. For each node, a set of maturity level is defined to evaluate the degree of validity of the claims/sub-claims contained in the *Goal/Sub-goal* nodes in the GSN tree. Weight factors are assigned to all goals and sub-goals in the GSN tree to formulate the dependency relation in the decision model based on the relative importance of the decision attributes. We call these weight factors as the decision parameters. These decision parameter needs to be carefully selected based on the expert input regarding the relative importance of different decision attributes. These weight factors or decision parameter are employed in the construction of conditional probability table for the Bayesian network in step 3. As the confidence network (or reduced GSN network) is based on the GSN tree, the end nodes directly correspond to the evidence (i.e., solution nodes in GSN tree). Therefore, we designate the end nodes as evidence node. The target level for each evidence node is decided based on the required degree of sufficiency and completeness of the evidence. The achieved level (or distribution) for each evidence node is decided depending on the available evidence. Based on the target level and achieved level (or distribution) of all the evidence nodes the target level and achieved level for all higher-level nodes in the entire network is computed in step 3.

Step 3: Perform quantitative confidence assessment

Decision model consists of two sets of information/data: (i) Subjective data based on expert opinion, and (ii) objective data based on evidence. We need the fusion of these two sets of information to provide quantitative confidence assessment. In this work, we use the Bayesian network (BN) for quantitative maturity assessment. The structure of the Bayesian network is based on the reduced GSN network. The nodes in the Bayesian network are same as the nodes in the reduced GSN network; however, the orientation of directed arrows is reversed because GSN follows top-down approach for decomposition while the Bayesian network is based on the bottom-up approach for computation. The Bayesian network incorporates expert opinion using casual relation and subjective probabilities. The probability distribution for the daughter nodes (i.e. higher-level node in the confidence network) is computed based on the distribution of the parent node and the conditional probability table (CPT) for the daughter nodes. Construction of conditional probability table (CPT) is a crucial part of the formulation of the decision model. CPT for a daughter node (i.e. decision attribute or goal) is constructed based on the weight factors assigned to its corresponding parent nodes (i.e. decision sub-attribute or sub-goal). The procedure for constructing the CPT is shown below.

3.5.4.1. Construction of CPT and probability estimation in the Bayesian network

Consider the Bayesian network shown in Figure 3.16. Let $N1,\ N2, \dots, Nj, \dots NJ$ represent the parent nodes, where J is the number of parent nodes. C represents the daughter node (see Figure 3.16). The weight factor for parent nodes are represented by $w_1, w_2, .. w_j, \dots w_J,$ such that,

$$w_1 + w_2 + .. + w_j + \cdots + w_J = 1 \qquad (3.1)$$

The probability distribution for the parent nodes are formulated using subjective probability based on different maturity levels. M_{ij} represents the i^{th} maturity level for the j^{th} parent node where, $i = 1, 2 \dots I$, indicates the index for the maturity levels of parent nodes and $j = 1,2, \dots .J$ represents the index for the parent nodes. D_k represent the k^{th} maturity level for the daughter nodes, where $k = 1,2, \dots .K$. For simplicity, we assume an equal number of maturity levels for all the nodes in the Bayesian network.

The maturity levels are designated by different grades. If four maturity levels or grades are used to form the set of maturity level for daughter and parent nodes, then the maturity levels can be defined by non-numeric or numeric representation as,

$D_k \in \{ NA, Low, Medium, High\}, K = 4$

$M_{ij} \in \{ NA, Low, Medium, High\}, I = 4$

or

$D_k = \{‘0’, ‘1’, ‘2’, ‘3’\}, K = 4$

$M_{ij} = \{‘0’, ‘1’, ‘2’, ‘3’\}, I = 4$

The rows in the CPT is based on the number of possible states or maturity levels for the daughter node. Therefore, the CPT has K rows. The columns in the CPT is based on different combination of states considering all the parent nodes. As there are I states (or maturity levels) and J parent nodes, the CPT has I^J columns. The conditional probability for the i^{th} element in a column is obtained by,

$$P\left(C_{D_k} \middle| N_{1_{M_{i1}}}, N2_{M_{i2}}, \dots Nj_{M_{ij}} \ \dots NJ_{M_{iJ}}\right) = \left(1 - \sum_{\substack{j=1 \\ D_k \neq M_{ij}}}^{J} w_j\right) \tag{3.2}$$

It should be noted that the columns of CPT represent conditional probability distribution corresponding to different states or maturity level of the parent node. Therefore, the sum of all elements in a column is always equal to 1.

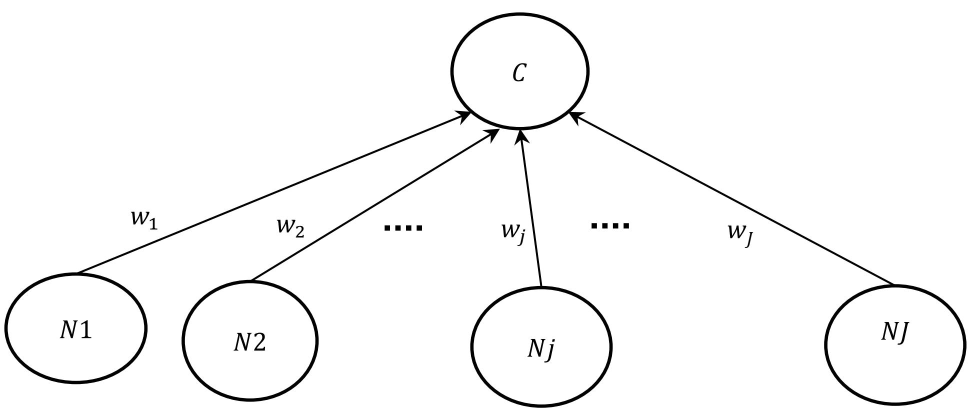


Figure 3.16: Bayesian network with J nodes to illustrate CPT computation

If $P(N1_{M_{i1}}), P(N1_{M_{i2}}), \dots P\left(N1_{M_{ij}}\right) \dots P\left(N1_{M_{iJ}}\right)$ represent the marginal probability for the parent nodes, then the probability distribution for the daughter node is obtained by,

$$P(C_{D_k}) = \sum_{M_{i1}} \sum_{M_{i2}} \dots \sum_{M_{ij}} \dots \sum_{M_{iJ}} \left(\begin{array}{c} P\left(C_{D_k} \middle| N1_{M_{i1}}, N2_{M_{i2}}, \dots Nj_{M_{ij}} \ \dots NJ_{M_{iJ}}\right) P(N1_{M_{i1}}) P(N2_{M_{i2}}) \dots \\ \dots P\left(Nj_{M_{ij}}\right) \dots P\left(NJ_{M_{iJ}}\right) \end{array} \right)$$

(3.3)

In a Bayesian network where multiple nodes are connected hierarchically, we can identify different sub-sets based on parent-daughter relation. The daughter nodes in lower level sub-sets becomes parent nodes in higher level sub-set. For each daughter node in a sub-set, the sum of assigned weights of its parent nodes should be equal to 1. For example, in Figure 3.17 the sum of weight factors for parent nodes in sub-sets 1, 2 and 3 should be equal to 1. The CPT and probability distribution for all the daughter nodes in different sub-set in the hierarchy can be computed using Eq. (3.2) and Eq.(3.3), respectively.

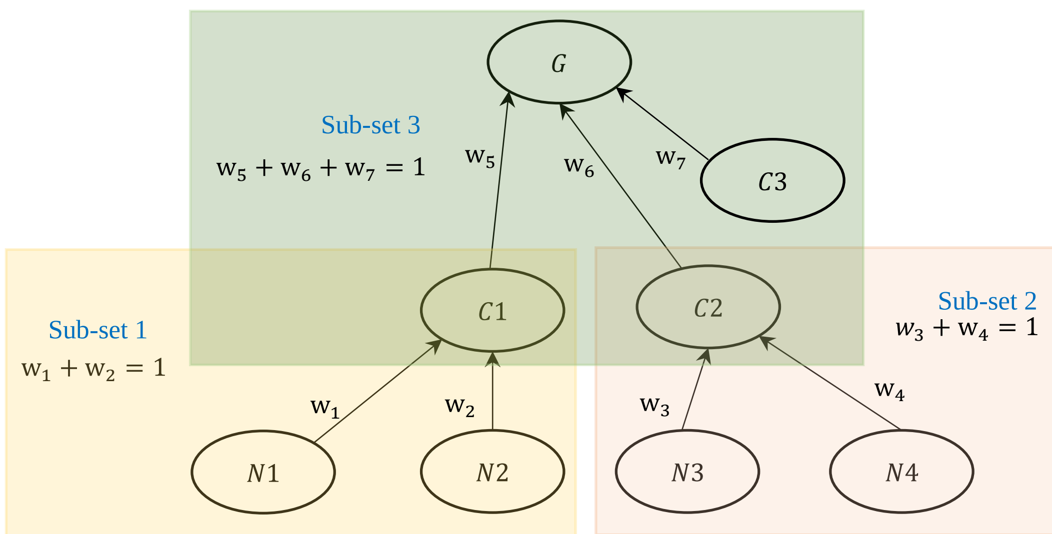


Figure 3.17: Bayesian network with multiple nodes in hierarchy

Further illustration of the formulation of the decision model is provided in Figure 3.18 based on the validation example shown in section 3.5.2. The main objective of the decision is to assess the code adequacy for an application. In this example maturity quantification implies a quantitative evaluation of the claim G1, i.e. "Code X is suitable for predicting the application XX." The GSN tree is transformed into a confidence network based on the goal (claim G1) and sub-goals (sub-claims G1.1 and G1.2). Node G 1.1 and G 1.2 are the evidence node in the confidence network. A set of four maturity level, which are labeled as {NA, Low, Medium, High}, is used to express the degree of validity of the claims and sub-claims in the reduced GSN network. The weight factor for the two sub-goals (G 1.1 and G1.2) is shown in Figure 3.18 (i.e. 40% for G1.1 and 60% for G1.2). The marginal probability distribution of evidence nodes (G1.1 and G1.2) are formulated using the maturity levels, based on the available evidence. The target level for all the nodes in the network is decided based on the required maturity level of the evidence nodes. The criteria for assessment is decided based on the bounds for validation result (VR) and data applicability (DA) shown in Table 3.6.

VR is assumed to be estimated by deterministic validation metric like, bias. Therefore, VR is obtained by,

$$VR = 1 - Bias \tag{3.4}$$

DA is assumed to be estimated by scaling methodology. Therefore, DA is given by,

$$DA = 1 - SD \tag{3.5}$$

here, SD represents scale distortion, which is a measure of dissimilarity between experiment and application. The assigned weight factors are used to obtain the conditional probability table (CPT) by following the technique described earlier in section 3.5.4.1. The CPT for CA is shown in Table 3.7

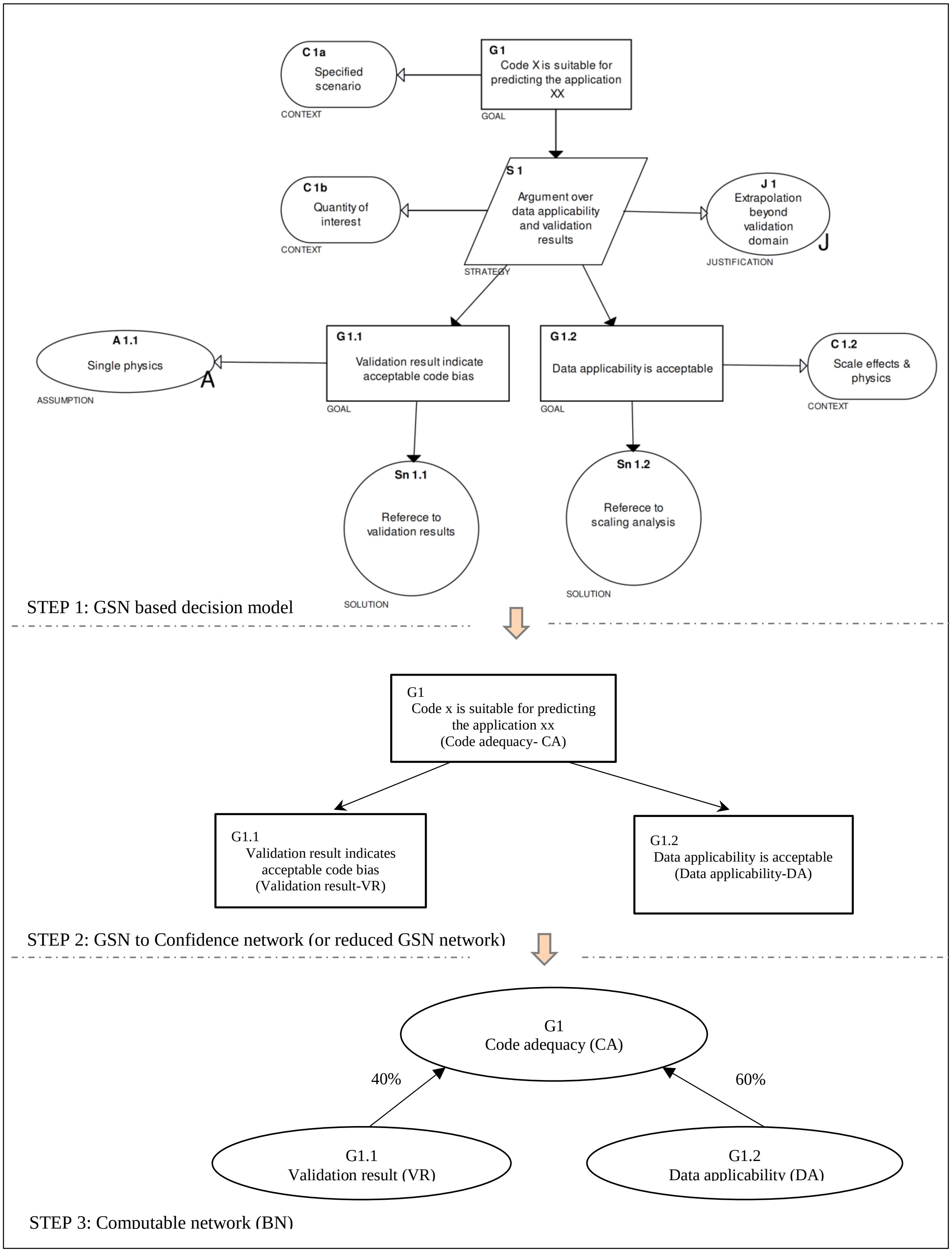


Figure 3.18: Transformation of GSN tree to Bayesian network

Table 3.6: Descriptors for states in the validation example

Validation results ($\boldsymbol{VR_{M_{i1}}}$) $M_{i1} = \{NA, L, M, H\}$	Data applicability ($\boldsymbol{DA_{M_{i2}}}$) $M_{i2} = \{NA, L, M, H\}$	Code adequacy ($\boldsymbol{CA_{D_k}}$) $D_k = \{NA, L, M, H\}$
$VR_{NA} \in [0,0.1)$	$DA_{NA} \in [0,0.1)$	$CA_{NA} \in [0,0.01)$
$VR_{L} \in [0.1,0.4)$	$DA_{L} \in [0.1,0.5)$	$CA_{L} \in [0.1,0.4)$
$VR_{M} \in [0.4,0.7)$	$DA_{M} \in [0.5,0.7)$	$CA_{M} \in [0.4,0.7)$
$VR_{H} \in [0.7,1]$	$DA_{H} \in [0.7,1]$	$CA_{H} \in [0.7,1]$

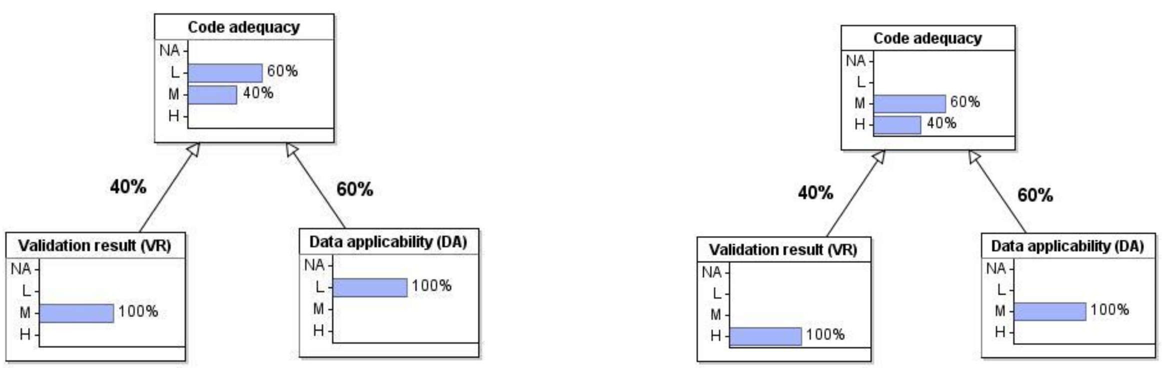
Table 3.7: Conditional probability table for the code adequacy (CA)

$\boldsymbol{VR_{M_{i1}}}$		Not available (NA) '0'				Low(L) '1'				Medium (M) '2'				High (H) '3'			
$\boldsymbol{DA_{M_{i2}}}$		NA '0'	L '1'	M '2'	H '3'	NA '0'	L '1'	M '2'	H '3'	NA '0'	L '1'	M '2'	H '3'	NA '0'	L '1'	M '2'	H '3'
$\boldsymbol{CA_{D_k}}$	NA	1	0.40	0.40	0.40	0.60	0	0	0	0.60	0	0	0	0.60	0	0	0
	L	0	0.60	0	0	0.40	1	0.40	0.40	0	0.60	0	0	0	0.60	0	0
	M	0	0	0.60	0	0	0	0.60	0	0.40	0.40	1	0.40	0	0	0.60	0
	H	0	0	0	0.60	0	0	0	0.6	0	0	0	0.60	0.40	0.40	0.40	1

Using Eq. (3.3), the probability distribution for the CA is estimated as,

$$P(CA_{D_k}) = \sum_{M_{i1}} \sum_{M_{i2}} P\big(CA_{D_k} | VR_{M_{i1}}, DA_{M_{i2}}\big) P\big(VR_{M_{i1}}\big) P\big(DA_{M_{i2}}\big) \tag{3.6}$$

If the target level for VR is High (H), and the target level for DA is Medium(L), then the target level for CA is as shown in Figure 3.19(b). Based on the available evidence, if VR is Medium (M) and DA is Low(L), then the achieved level for CA is as shown in Figure 3.19 (a).



(a) Achieved level for all nodes (b) Target level for all nodes

Figure 3.19: Illustration example for the Bayesian network with single evidence to support the attributes

3.5.4.2. Assessment with multiple evidence

In cases where single evidence is available to support an attribute, the grade can be decided with certainty based on the criteria of assessment. However, when multiple evidence with varying grades (or maturity levels) are available to support an attribute, it becomes difficult to decide the overall grade for the assessment of attribute. In such case, the marginal probability distribution for the evidence nodes is obtained by collaborating all the evidence by using information regarding frequency of grade and quality of evidence. The grade of an evidence reflects it quality. Evidence with higher level grade are considered as supporting evidence as they consolidate our confidence in the claim related to the attribute. Evidence with lower level grade are considered as counter-evidence as they challenge or refute our claim related to the attribute. As there is a general tendency for risk aversion (prospect theory), counter-evidence are weighted more compared to supporting evidence when multiple evidence are available. The steps for estimation of marginal probability for the evidence nodes with multiple evidence is shown in Figure 3.20.

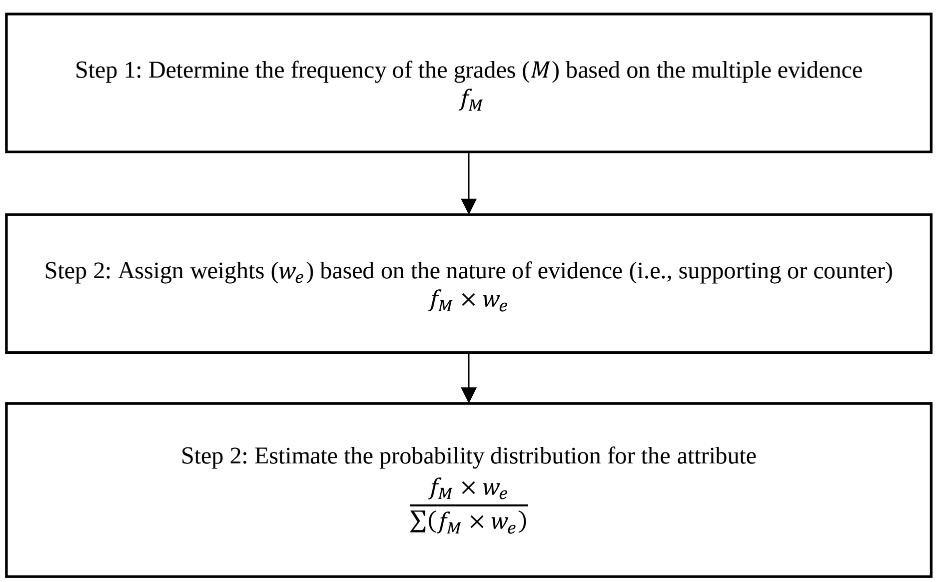


Figure 3.20: Estimation of marginal probability distribution for evidence node with multiple evidence

Let us assume that in the previous example for code adequacy assessment instead of one experiment we have 10 experiments with varying results from validation test and scaling analysis, then the probability distribution for the validation result and data applicability is obtained as shown in Table 3.8 and Table 3.9.

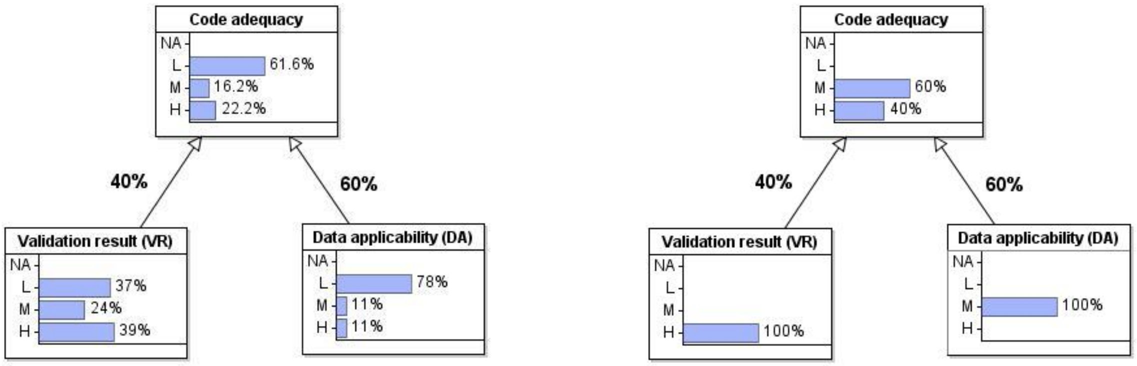
Table 3.8: Estimation of probability distribution (for validation result) based on multiple evidence assessment

Probability, $P(VR_{M_{i1}}) = \frac{f_{M_{i1}} \times w_e}{\Sigma(f_{M_{i1}} \times w_e)}$	0	0.37	0.24	0.39
Weight assignment$(f_{M_{i1}} \times w_e)$	0	2×70	3×30	5×30
Frequency of grade $(f_{M_{i1}})$	0	2	3	5
Grade (M_{i1})	NA	Low	Medium	High
Evidence weight (w_e)	70 % (counter evidence)		30 % (supporting evidence)	

Table 3.9: Estimation of probability distribution (for data applicability) based on multiple evidence assessment

Probability, $P(DA_{M_{i2}}) = \frac{f_{M_{i2}} \times w_e}{\Sigma(f_{M_{i2}} \times w_e)}$	0	0.78	0.11	0.11
Weight assignment	0	6×70	2×30	2×30
Frequency of grade $(f_{M_{i2}})$	0	6	2	2
Grade (M_{i2})	NA ‘0’	Low ‘1’	Medium ‘2’	High ‘3’
Evidence weight, w_e	70 % (counter evidence)		30 % (supporting evidence)	

Based on evidence from Table 3.8 and Table 3.9, the updated result of code adequacy assessment are shown in Figure 3.21.



(a) achieved (b) Target

Figure 3.21: Illustration example for the Bayesian network with multiple evidence to support the attributes

3.5.4.3. Estimation of weight factor for the Phenomenology pyramid

Validation is the most important attribute in the assessment of simulation code. Phenomenology pyramid (PP) serves as a guide for all activities related to validation of code. We assess different attribute related to code validation assessment based on the phenomena identified by the PIRT process. The importance factor for phenomena in the PIRT (provided by the expert elicitation process) helps in assigning the weight factor for different nodes in the phenomenology pyramid (PP). This step of weight assignment is important because the phenomenology pyramid (PP) forms the basis for the assessment of different validation attributes (discussed in section 3.5.3). The PIRT is formed by decomposing the system based on dominating physics, system conditions, sub-system components, etc. These demarcating conditions, physics or system components are regarded as the ancillary node in the reduced GSN network for the phenomenology pyramid (PP). Dominating processes or phenomena are identified and organized for each ancillary node. Figure 3.22(a) shows the reduced GSN network for a phenomenology pyramid corresponding to the PIRT in Table 3.10. $Ph1, Ph2, \ldots, Ph5$ are the phenomena identified by the PIRT (in Table 3.10). $S1$ and $S2$ are the ancillary nodes. $Ph5$ is a global phenomenon; therefore, it is not associated with an ancillary node. All phenomena nodes are evidence node, where evidence corresponding to relevant validation attribute are integrated in the network. For example, if we are making data relevance assessment based on the phenomenology pyramid (PP), then evidence related to relevant data will be integrated at the phenomena nodes.

The normalized importance factor (e.g. column 4 in Table 3.5) are used to assign weights to the bottom layer (evidence node) of the phenomenology pyramid (PP). The weight factors for the higher-level nodes are obtained by adding weight factors of the corresponding lower level nodes (see Figure 3.22 (a)). The weight factors assigned by this process need to be renormalized when the reduced GSN network is transformed into the Bayesian network. Weight renormalization is required because according to the condition in Eq. (3.1) the sum of weight factors for all parent nodes in a sub-set should be 1. In Figure 3.22(b), weight factor for parent nodes in sub-sets 1 and 2 are obtained by renormalizing the weight factors assigned to the reduced GSN network for the phenomenology pyramid (PP).

Table 3.10: PIRT table (example)

Sub-system or system condition	Phenomena (P_j)	Importance factor $I_j \in [0,3]$	Normalized importance factor $\left(w_j = \frac{I_j}{\sum I_j}\right)$
S1	Ph1	2	0.2
	Ph2	2	0.2
S2	Ph3	1	0.1
	Ph4	2	0.2
	Ph5	3	0.3

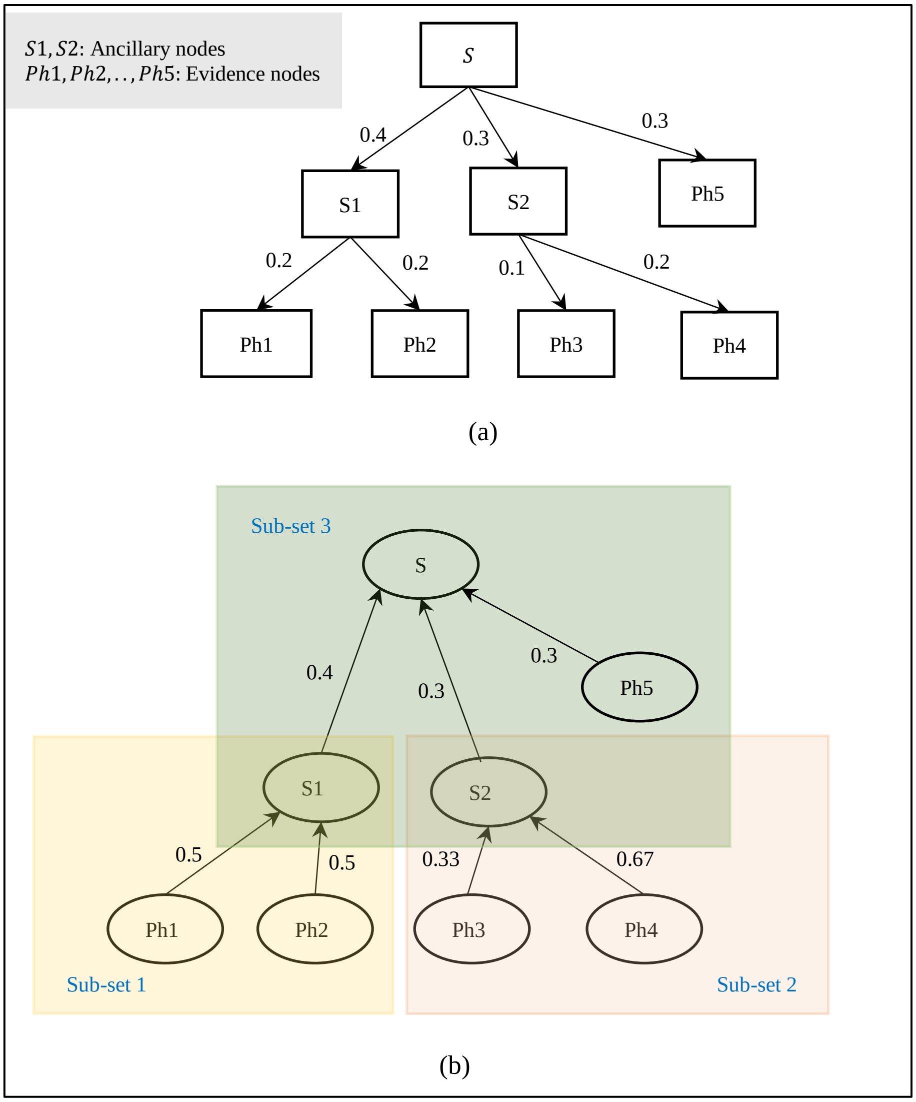


Figure 3.22: Reduced GSN network Bayesian network for phenomenology pyramid (PP)

3.5.5. Evaluation and interpretation of result

The formalized decision model is evaluated by comparing the target level and the achieved level of maturity of different attributes and sub-attributes based on their expected utility. The difference in the expected utility of target and achieved level for an attribute C (or sub-attribute or node) is given by,

$$\Delta E(C) = \left[\sum_{D_k} P\left(C_{D_k}\right)U(D_k)\right]_{Target} - \left[\sum_{D_k} P(C_{Dk})U(D_k)\right]_{achieved} \quad (3.7)$$

Here, $P\left(C_{D_k}\right)$ represents the probability corresponding to maturity level D_k and $U(D_k)$ represents the utility of the maturity level D_k. $\left[\sum_{D_k} P\left(C_{D_k}\right)U(D_k)\right]_{Target}$ represents expected utility of target level and $\left[\sum_{D_k} P(C_{Dk})U(D_k)\right]_{achieved}$ represents expected utility of achieved level. $\Delta E(C)$ is normalized to obtain a metric called expected distance metric that estimates the distance between the target and achieved level on a scale of 0 to 1. The Expected distance metric for an attribute (or sub-attribute or node) C is given by,

$$E_N(C) = \frac{\left[\sum_{D_k} P\left(C_{D_k}\right)U(D_k)\right]_{Target} - \left[\sum_{D_k} P(C_{Dk})U(D_k)\right]_{achieved}}{\left[\sum_{D_k} P(C_{Dk})U(D_k)\right]_{Target}} \quad (3.8)$$

Important properties related to $E_N(C)$ are given by,

(1) $0 \leq E_N(C) \leq 1$

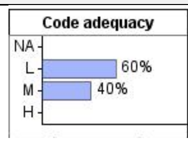
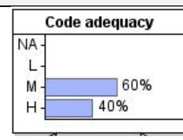
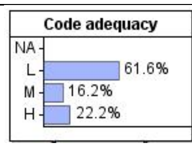
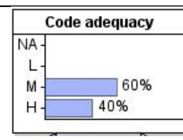
(2) $\left[\sum_{D_k} P\left(C_{D_k}\right)U(D_k)\right]_{Target} \geq \left[\sum_{D_k} P(C_{Dk})U(D_k)\right]_{achieved}$

(3) $U(D_{k+1}) > U(D_k)$

(4) $U(D_0) = 0$, as value of D_0 is 0$

$E_N(C)$ close to 0 implies achieved level is closer to the target level. $E_N(C)$ close to 1 implies that the attribute C is supported by insufficient and/or incomplete evidence. $E_N(C) = 1$, implies that credible evidence to support the attribute C are not available. The expected distance metric can be applied to any node in the Bayesian network to measure its current state of assessment. Table 3.11 shows the expected distance metric for the validation example shown in Figure 3.19 and Figure 3.21. It is evident from these results that the achieved level is closer to the target level for case 2 (as $E_N(CA\ )$ for case 2 is smaller than $E_N(CA)$ for case 1).

Table 3.11: Illustration of expected distance metric

Case	Utility for maturity levels {NA, L, M, H}	Probablity distribution for achieved level of code adequacy (CA)	Probablity distribution for target level of code adequacy (CA)	Expected distance metric, $E_N(CA\ )$
(1)	U(NA)=0 U(L)=2 U(M)=4 U(H)=6			0.42
(2)	U(NA)=0 U(L)=2 U(M)=4 U(H)=6			0.33

3.5.6. Refinement

Code's maturity assessment is a confidence-building process which may require several iterations. If the achieved level reaches the target level in the first iteration, the code is mature enough to predict the application of interest with the required degree of confidence. However, if the achieved level is less than the target level, refinement is required.

Refinement is performed based on the evaluation and interpretation of results obtained in the previous section of the framework (section 3.5.5.). Refinement section consists of action items that points to required modification and improvement in the decision process. The priority list for the action item is obtained by comparison of the Expected distance metric of different attributes. Items with expected distance metric close to 1 have higher priority while those with expected distance metric close to zero have lower priority. Action items for refinement can be categorized into four parts:

(1) Refinement of the decision model/framework: Refinement of decision model/ framework is focused on weight factor adjustment and any structural change in the model based on the input of SME. This step may be considered as the calibration of the decision model. Weight factor adjustment at a level in the hierarchy is performed by comparison of expected distance metric of the decision attributes at that level.

    Example: Direct validation factors and PQA factors are higher level attributes for code validation assessment. If we have higher confidence in PQA, i.e. $E_N(PQA) \approx 0$, then weight factor for PQA should be made smaller compared to weight factor for other attributes at this level. Similarly, if we have higher confidence in data quality then weigh factor for data applicability should be made smaller compared to the validation test results.

(2) Refinement of models: Refinement of models is performed based on the assessment of different phenomena simulated by the code. If we have high-quality data, (i.e.,$E_N(PSA)$, $E_N(DRA)$ and $E_N(DUA)$ is low), but validation result indicates high discrepancy (i.e., $E_N(VTR)$ is high) then the validation test results for the individual phenomena is examined and appropriate model in the code are modified.

(3) Refinement of data: Data refinement is based on the result of data applicability assessment. If we have low confidence in a data set (i.e., $E_N(PSA) > 0.5, E_N(DRA) > 0.5, E_N(DUA) > 0.5$) then the respective data set is discarded from the evidence database.

(4) Refinement of PQA factor: Refinement of PQA factor is performed based on the value of expected distance matric for different PQA factors included in the decision model.

3.6. Summary

This chapter describes the formulation of the proposed framework. The framework consists of different elements that encompass technique for structural knowledge representation, evidence classification and characterization, and quantitative maturity assessment. Structural knowledge representation in the framework is obtained using an argument modeling technique called Goal structuring notation (GSN) [6]. The PIRT based phenomenology pyramid is used to guide the classification and characterization of evidence for code validation assessment. The Pyramid is constructed using the GSN. The decision schema in the proposed framework is based on the PCMM [3] and the Analytic hierarchy process (AHP) [7]. The hierarchical decision model is constructed using the GSN. The number of levels in the hierarchy depends upon the required depth and rigor of the analysis. Each attribute and sub-attribute in the decision model is formulated as a claim (i.e. Goals nodes in the GSN tree) where the degree of validity of the claim (attribute's assessment) is defined by different maturity levels. Evidence are integrated across the lower level attribute in the decision model (using the solution nodes in the GSN tree). The GSN based decision model is transformed into a confidence network (Bayesian network) for quantitative maturity assessment. The Bayesian network enables the abstraction of maturity information from lower level attributes to higher level attributes. It helps in assessing the maturity based on the quality of evidence integrated in the decision model. Subjective data based on the expert opinion is incorporated into the decision model using condition probability table (CPT) and subjective probabilities based on the criteria of evaluation of the evidence. A metric based on the expected utility of the maturity levels is proposed to evaluate the distance between the target level and achieved level of maturity on a scale of 0 to 1 for each attribute and sub-attribute in the decision model.

CHAPTER 4: ASSESSMENT FRAMEWORK DEMONSTRATION - CASE STUDY I

The case study presented in this chapter is based on one of the CASL challenge problem called Departure from Nucleate Boiling (DNB). Departure from nucleate boiling is a boiling phenomenon which is observed at high heat flux (critical heat flux) conditions. Under this condition, bubble formation is so fast that a blanket of vapor is created on the boiling surface. Thus, heat transfer to bulk coolant is reduced and localized heat spots or surface dry out may lead to clad failure.

In this study, we focus on only one element of PCMM, i.e., Model Validation. Validation is one of the most challenging elements of CASL M & S activities. This challenge primarily arises due to a shortage of data to match the high level of modeling details in CASL codes. CASL adopted the validation pyramid approach to counter these challenges. However, multiphysics and multiscale nature of CASL challenge problems limit the use of AIAA validation pyramid for CASL CPs. CASL developed a modified validation pyramid for CPs using the Component Identification and Ranking Process [26].

Figure 4.1 shows the structural representation of decision model (using GSN) for validation assessment of multiphysics CASL codes (VERA) for DNB simulation based on the CASL validation pyramid (shown in Figure 2.10). It is evident from Figure 4.1 that the decision regarding validation assessment of VERA for DNB has different components. Each component is resolved in different GSN module represented as *Away goals* in Figure 4.1. The case study presented in this chapter is focused on one of the components, i.e. validation assessment of sub-channel thermal hydraulic code for DNB simulation (Away goal 1.1.1, module D1).

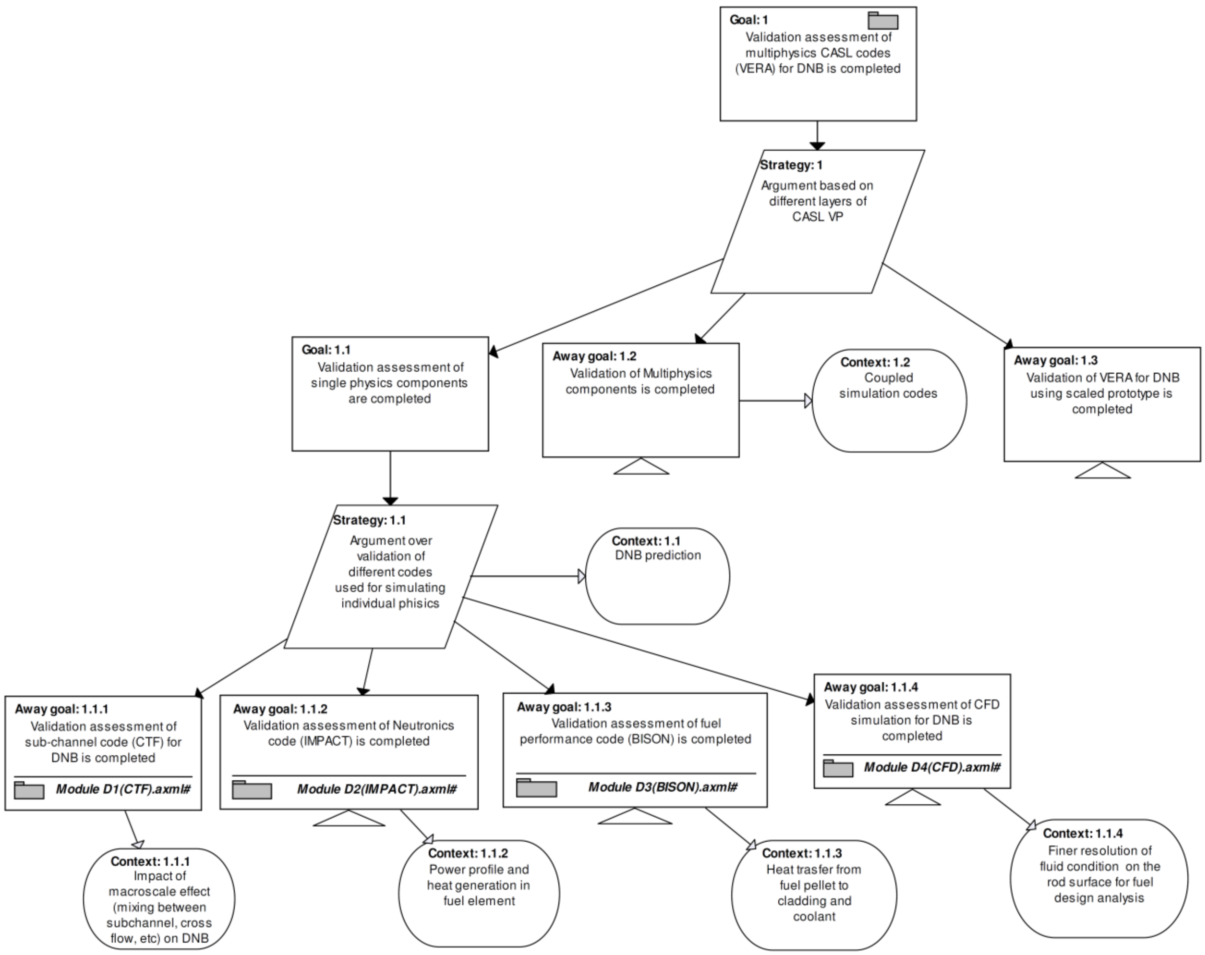


Figure 4.1: Decision model for validation assessment of Multiphysics CASL code (VERA) for DNB

4.1. Objective of the case study

Demonstrate formulation of different elements of the framework for validation assessment of a single physics code and test if the proposed framework can provide a significant improvement in the validation assessment of the selected code.

4.2. Demonstration of the framework

All the elements of the framework for this case study are discussed in the following sub-sections.

4.2.1. Preprocessing for the framework development

This section presents the preprocessing requirement for the development of the framework for validation assessment of CTF for DNB simulation. CORBA-TF or CTF stands for Coolant-Boiling in Rod Arrays-Two Fluids. It is a sub-channel thermal hydraulic code adopted by CASL (along with different simulation codes for different physics) to develop a high-fidelity multi-physics simulation capability for different challenge problem applications. CTF is used to assess the impact of macro-scale effects (~1 cm) on DNB prediction. These macro-scale effects encompass impact of mixing between sub-channels, cross-flow, turbulence, and grid-spacer effect on average flow parameters, like pressure (P), mass flux (G), and thermodynamic quality (X_{th}). Higher confidence in the prediction of the local fluid condition is important as they are used in the development of DNB correlation [99].The preprocessing requirement for development of framework consists of three main steps (see section 3.5.1). These steps for validation assessment of CTF are described in Table 4.1 to Table 4.3. The PIRT presented in Table 4.3 is based on the PIRT for DNB in the CASL V & V report for VERA [20].

Table 4.1: Specify the issue, simulation tool and decision objective (Step 1)

Issue	Assess impact of macroscale effects (~ 1cm) on DNB prediction Effect of mixing between sub channel, cross flow, turbulence and grid spacer on averaged flow parameter-Pressure (P), mass flux(G), thermodynamic quality (X_{th}) (DNB is empirically correlated at this scale as a function of P, G, Xth)
Simulation tool	COBRA-TF/CTF
Decision objective	Validation assessment of CTF for DNB

Table 4.2: Specify scenario, system condition, FOM (Step 2)

Specify scenario	Transient and accident scenario
System condition	PWR system condition during transient and accident scenarios→ Power excursion, loss of coolant flow, control rod malfunction events, etc.
FOM	DNBR ratio: Ratio of predicted critical heat flux to the local heat flux

Table 4.3: PIRT (Phenomena resolution)→ Important $I\epsilon[0,3]$; Knowledge $K\epsilon[0,3]$ (Step 3)

Phenomena	Description	$\boldsymbol{I}$	$\boldsymbol{I_N} = \left(\frac{I}{\sum I}\right)\%$	$\boldsymbol{K}$
Turbulent Mixing (TM)	Mixing associated with turbulence, usually near the spacer grid			2
• TM in SPF		1.5	7.8	
• TM in TPF		1.5	7.8	
Cross flow (CF)	The directed flow associated with mixing vanes commonly found on spacer grids			2
• CF in SPF		1.5	7.8	
• CF in TPF		1.5	7.8	
Nucleate boiling (NB)	Boling confined to the surface of the clad below the critical heat flux	3	15.78	2.5
Critical Heat Flux (CHF)	A condition where liquid cannot rewet the rod surface because of the rate of vapor production impends the liquid flow back to the hot surface	3	15.78	-
Natural circulation(NC)	Convection associated with fluid moving from a region of higher density (cooler) to a region of lower density (warmer)			2.5
• NC in SPF		1	5.263	
• NC in TPF		1	5.263	
Pressure drop (PD)	The change in pressure along the length of flow associated with frictional resistance			2
• PD in SPF		1	5.263	
• PD in TPF		1	5.263	
Flow regime (FR)	The characteristics of the flow in the channel: laminar, turbulent, bubbly, slug, etc.	3	15.78	2.5

4.2.2. Structural knowledge representation

The GSN based phenomenology pyramid corresponding to the PIRT presented in Table 4.3 is shown in Figure 4.2. As GSN is an argument modeling technique the objective (Goal: 1) in the GSN tree is stated as a claim, “Phenomenology pyramid for DNB is constructed”. As this phenomenology pyramid (PP) is constructed for single physics component (i.e., sub-channel thermal hydraulics) to assess the impact of macroscale effect on DNB prediction, this information is contained in the context blocks (Context: 1a and Context: 1b) in Figure 4.2. The phenomena (from the PIRT in Table 4.3) are classified based on boiling condition (Goal:1.1), flow redistribution mechanism (Goal 1.2) and glow condition (Goal: 1.2.1,1.2.2,1.2.3 and 1.3).

The reduced GSN network for the phenomenology pyramid (PP) is shown in Figure 4.3. It should be noted that all end nodes (or evidence nodes) in Figure 4.3 are phenomena from PIRT in Table 4.3. Based on the normalized importance factor in Table 4.3, the weight factor for different nodes in the phenomenology pyramid (PP) is calculated using the techniques described in section 3.5.4.3. This reduced GSN network for the phenomenology pyramid (PP) provides the basis for assessment of different validation attribute. It should be noted that the weight factor in Figure 4.3 are renormalized when reduced GSN network is transformed into the Bayesian network in the subsequent section of the framework.

As discussed in section 3.5.2, each phenomenon in the phenomenology pyramid (PP) is characterized by a set of QOI and corresponding system condition (parameter range). The characterization of phenomena is discussed in the next section along with classification and characterization of evidence. The experiments and specific model or correlation that are evaluated in different validation test are also enlisted in this section. The information related to each phenomenon and corresponding model and data is added to the solution node in the phenomenology pyramid (PP) (Figure 4.2) using the node dialog box in ASCE (tables and hyperlink to file can be added to these nodes in ASCE). In this way, all relevant information corresponding to model and data pyramid can be maintained inside a single pyramid.

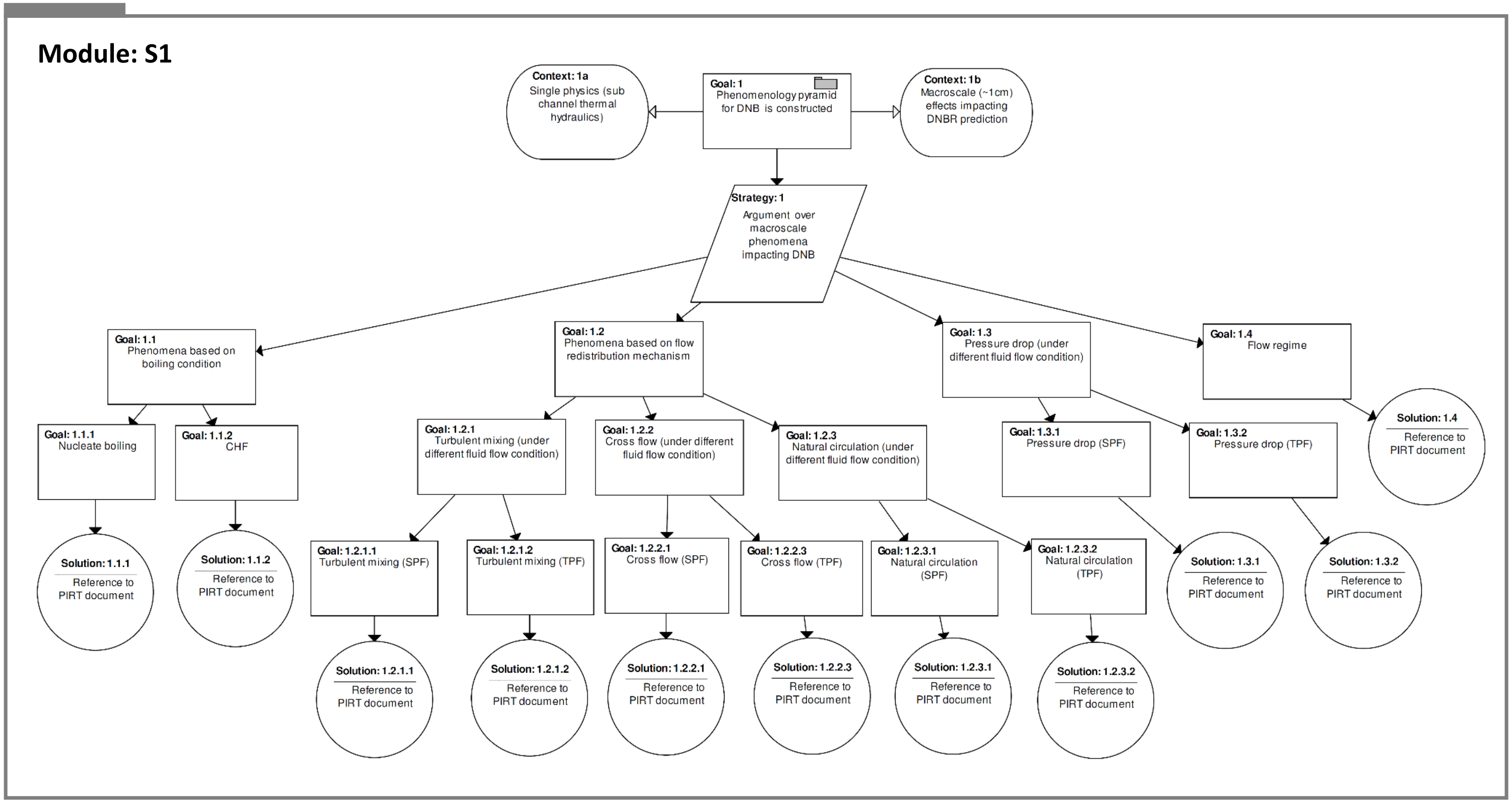


Figure 4.2: Phenomenology pyramid

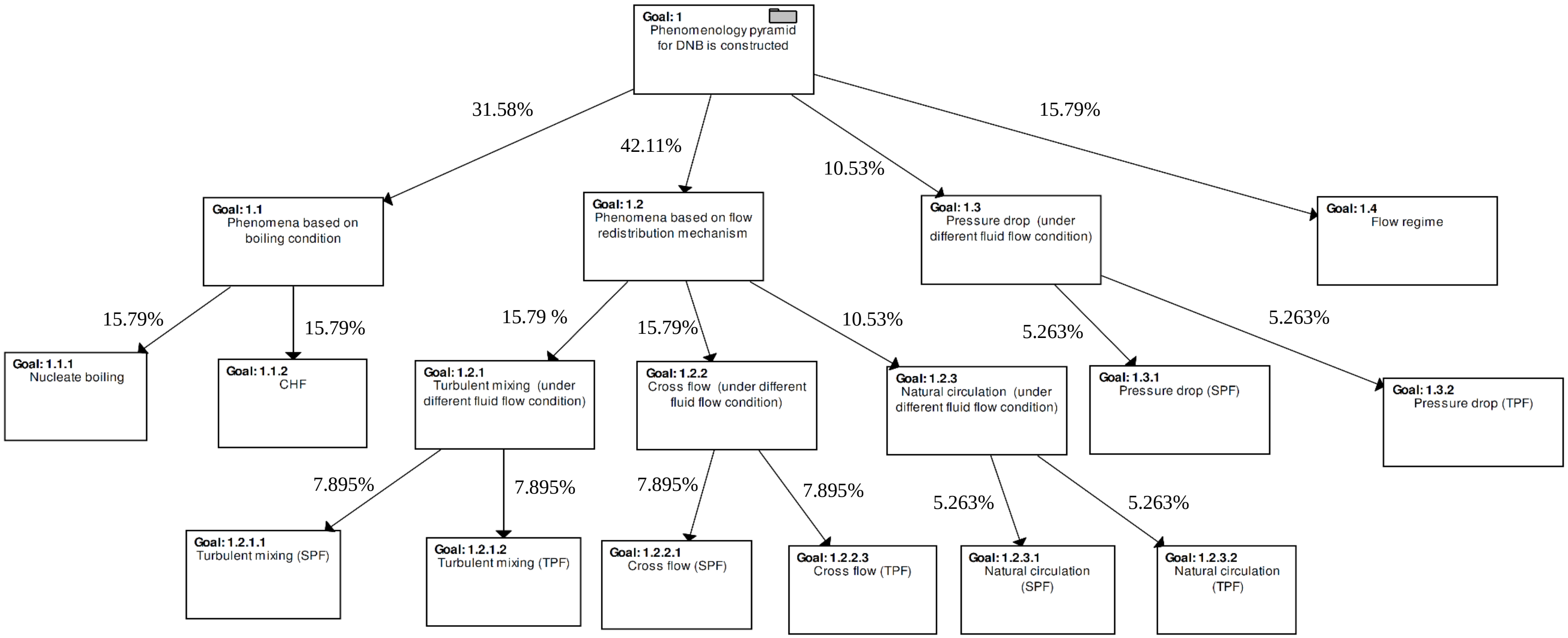


Figure 4.3: Reduced GSN network for the phenomenology pyramid (PP)

4.2.3. Classification and characterization of evidence

This section presents classification and characterization of evidence for validation assessment of CTF for DNB. The evidence presented in this section are based on the CTF V & V report [100]. It should be noted that the classification and characterization of phenomena and evidence is based on initial assessment and needs to be revised based on the input from subject matter expert (SME). Fields in the tables that are incomplete or have not been evaluated are labeled as TBA (i.e. To Be Assessed).

Classification and characterization of evidence for validation of each phenomenon is described by a set of three tables (e.g., Table 4.4, Table 4.5 and Table 4.6 for turbulent mixing in SPF):

- 1st Table: Characterization of a phenomenon
- 2nd Table: Classification and characterization of evidence used for the assessment of the phenomenon in 1st table
- 3rd Table: Description of evidence and their reference

The first table in the set of three tables present characterization of a phenomenon (see Table 4.4, Table 4.7, Table 4.10, Table 4.13, Table 4.16 and Table 4.19). It consists of three main columns. The first column consists of a set of QOI to characterize the phenomena. The second column consists of governing system condition for each QOI (in 1st column) based on dominant parameters. The third column defines the range of parameters (in the 2nd column) for phenomenon, model and data. The range of parameters have not been assessed; therefore, the empty fields in the table are labeled as TBA.

The second table in the set of three tables is focused on classification and characterization of evidence for assessment of code ability to simulate a phenomenon (see Table 4.5, Table 4.8,

Table 4.11, Table 4.14, Table 4.17 and Table 4.20). It consists of four main columns. The first column in this table consists of QOI defined in the table for characterization of phenomenon. The second column consists of coverage assessment result. Since, coverage assessment is not completed in this study, the respective columns are labeled as TBA. The third column in the table is based on the result of data relevance (R), physics scaling (PS) and data uncertainty assessment (U) of the experimental data used in the validation test for QOI. The fourth column consists of assessment of validation test result along with the description of specific model and correlation that has been evaluated. Grading in this table is based on the capability grades discussed in section 3.5.3.

The third table in the set of three tables consists of a description of all evidence used in "classification and characterization of evidence" for assessment of code ability to simulate a phenomenon and data applicability (see Table 4.6, Table 4.9, Table 4.12, Table 4.15, Table 4.18 and Table 4.21). These evidence are graded according to their level of detail (2^{nd} column) and credibility (4^{th} column) using the assessment grades described in section 3.5.3.2. As all evidence presented in this study are based on initial author assessment, they are assigned the grade "IA".

Validation of cross-flow, flow regime and CHF is not available in the CTF V & V report [100]; therefore, classification and characterization of evidence for these phenomena are incomplete and graded as "NA" in the evidence nodes for respective validation attribute. The specific description of gaps is noted in Table 4.22.

4.2.3.1. Turbulent mixing in single phase flow

This sub-section present tables for classification and characterization of evidence for assessment of CTF's ability to simulate turbulent mixing in single phase flow condition.

Table 4.4: Characterization of Turbulent mixing in single phase flow (TM in SPF)

$QOI = \left\{T_e, \frac{W'_{ij}}{\mu}, G_{axial\_profile}, G_{outlet}\right\}$		Governing system condition (set of parameters)		Parameter range for Phenomena, model and data			
Symbol	Description	Symbol	Description	Phenomena	Data		Model
					Range	Test	
T_e	Channel exit temperature $[F]$	q''	Average heat flux$\left[MBTU/_{Hr.ft^2}\right]$	TBA	0.1-0.6	CE 5x5	TBA
$\frac{W'_{ij}}{\mu}$	Turbulent transverse mixing rate (Non -dimensional) (i and j are channel number, μ *is dynamic viscosity*)	Re β	Reynolds number $\left(\frac{Transverse\ mass\ flux}{Axial\ mass\ flux}\right)$	TBA	5000-40000	Kumamoto 2x3	TBA
G_p	Mass flow rate distribution	h_{axial}	Axial location $[m]$	TBA	0-12 0-3	RPI 2x2 GE 3x3	TBA
G_{outlet}	Outlet mass flux $\left[kg/_{m^2s}\right]$	X_{sc}	Different subchannel (corner, side, center)	NA	NA	GE 3x3	NA

Table 4.5: Classification and characterization of evidence for turbulent mixing in single phase flow [with capability grade (CG→0/1/2/3/4)]

QOI	Coverage			Data applicability (DA) with capability grade (CG)				Validation test result (VTR) with capability grade (CG)			
	CMP	CME	CEP	CG			Evidence reference	Specific model /correlation evaluated	Metric	CG	Evidence reference
				R	PS	U					
T_e	TBA	TBA	TBA	2	1	3	CT 1.1. 1 CT 1.3. 1 CT 1.3. 2	**<u>CE 5x5</u>** Heat transfer models <u>Turbulent mixing model</u>	Bias=±10F	2	CT 1.3. 3
$\frac{W'_{ij}}{\mu}$	TBA	TBA	TBA	3	1	2	CT 1.3. 4 CT 1.3. 5	**<u>Kumamoto 2x3</u>** Turbulent mixing model with different mixing coefficient (β):			CT 1.3. 6
								Rogers and Rosehart correlation for β	Graphical	1	
								Blasius friction correlation with β=0.004	Graphical	1	
								Blasius friction correlation with β=0.007	Graphical	2	
								CTF friction correlation with β=0.007	Graphical	1	
G_p	TBA	TBA	TBA	1	0	2	CT 1.3. 7 CT 1.3. 8	**<u>RPI 2x2</u>** CTF friction correlation with no turbulent mixing	Graphical	2	CT 1.3. 9
G_p								Turbulent mixing model with→			
								CTF friction correlation and β=0.007	Graphical	2	
								CTF friction correlation and β=0.0035	Graphical	2	
G_{outlet}				2	0	0	CT 1.3. 7 CT 1.3. 8	**<u>GE 3x3</u>** Turbulent mixing model with Rogers and Rosehart correlation for β	Graphical Graphical Graphical	3 3 2	

Table 4.6: Evidence for turbulent mixing in single phase flow (based on CTF validation and verification report [100])

Index	Level of detail	Description	Credibility	Reference
CT 1.1. 1	LLC	CE 5x5 rod-bundle experiment facility Grids does not contain any mixing vanes (check)	IA	Section 3.4 in [100]
CT 1.3. 1	LLC	CE 5x5 tests were run at prototypic PWR pressure temperature and heat flux	IA	Section 3.4 in [100]
CT 1.3. 2	LLC	Temperature measurement made on the rod surface via Thermocouple attached inside the heater tube and at the outlet of the test section in the center of each 36 coolant channels Measurement error= 0.5 F	IA	Section 3.4 and section 6.1.1 in [100]
CT 1.3. 3	LLC	Average difference between CTF predicted channel exit temperature and experimental values for all test in CE 5x5 fall between ±10F Outlier observed at low heat flux which gave bias higher than 50F were neglected in validation test	IA	Section 6.1.1 in [100]
CT 1.3. 4	LLC	Kumamoto university 2x3 facility is an air water facility. Specifically designed for mixing and void drift study. Mixing channel is short, so inlet flow of individual channel is adjusted so that flow in mixing channel is in mechanical equilibrium	IA	Section 3.9 in [100]
CT 1.3. 5	LLC	Measurement made by gas chromatography for the gas phase and spectrometer for the liquid phase. Measurement error not reported in the report	IA	Section 3.9 in [100]
CT 1.3. 6	LLC	RMSE or bias was not reported in the results. Plots shows best results for Blasius friction correlation with β=0.007. β is a tuning parameter needs to be tuned based on the geometry of the facility	IA	Section 6.1.2 in [100]
CT 1.3. 7	LLC	**<u>RPI 2x2</u>** No spacer grids were used in the experiment **<u>GE 3x3</u>** BWR like simulation with general electric rods Pins holding the rods in place act as spacer	IA	Section 3.8 in [100] Section 3.5 in [100]
CT 1.3. 8	LLC	**<u>RPI 2x2</u>** Measurement uncertainty for channel mass flux is 5% **<u>GE 3x3</u>** Measurement uncertainty not reported	IA	[101], Section 6.1.3 in [100]
CT 1.3. 9	LLC	**<u>RPI 2x2</u>** CTF predict the correct flow distribution; however, not within the axial length of the test section, which is 1 m CTF model is extended to 7 m to show that correct split is eventually achieved **<u>GE 3x3</u>** CTF predict the correct flow distribution at a shorter length of 1.8m, which is the exit of the facility (shows much better result compared to RPI 2x2) G_{outlet} was evaluated at different locations-Side and inner region (channel) results are largely unaffected (rRMS between 0.8 -2.5 %); however, corner channel shows significant drop in accuracy (rRMS=9.8-13.2 %)	IA	Section 6.1.3 in [100]

4.2.3.2. Turbulent mixing in two phase flow

This sub-section present tables for classification and characterization of evidence for assessment of CTF's ability to simulate turbulent mixing in two phase flow condition.

Table 4.7: Characterization of turbulent mixing in two phase flow (TM in TPF)

$QOI = \{x, G_p, G_{outlet}\}$		Governing system condition (set of parameters)		Parameter range for Phenomena, model and data			
Symbol	Description	Symbol	Description	Phenomena	Data		Model
					Range	Test	
x	Thermodynamic quality	X_{sc}	Different subchannels (corner, center, side)	NA	NA	GE 3x3	NA
G_p	Mass flow rate distribution	h_{axial}	Axial location $[m]$ in subchannel	TBA	0-2.5	GE 3x3	TBA
G_{outlet}	Outlet mass flux $\left[{}^{kg}/{}_{m^2 s}\right]$	X_{sc}	Different subchannels (corner, center, side)	NA	NA	GE 3x3	NA

Table 4.8: Classification and characterization of evidence for of turbulent mixing in two phase flow [with capability grade (CG→0/1/2/3/4)]

QOI	Coverage			Data Applicability (DA) with capability grade (CG)				Validation test result (VTR) with capability grade			
	CMP	CME	CEP	CG R	CG PS	CG U	Evidence reference	Specific model /correlation evaluated	Metric	CG	Evidence reference
x	TBA	TBA	TBA	2	2	3	CT 1.3. 14 CT 1.2. 1 CT 1.3. 15	**<u>GE 3x3</u>** 1) Void drift model (Ka=1.4), turbulent mixing model($\beta_{sp} = 5$)	RMSE =0.036 % (corner) RMSE=0.014% (inner) RMSE=0.017% (side)	3	CT 1.3. 10
								2) No void drift and turbulent mixing model ($\beta_{sp} = 5$)	RMSE=0.132% (Corner) RMSE=0.032% (inner) RMSE=0.019% (side)	3	CT 1.3. 11
								3) Rogers and Rosehart correlation for β	RMSE=0.035% (Corner) RMSE=0.014% (inner) RMSE=0.017% (side)	3	CT 1.3. 12
G_{outlet}	TBA	TBA	TBA	2	2	3	CT 1.3. 14 CT 1.2. 1 CT 1.3. 16	**<u>GE 3x3</u>** 1) Turbulent mixing model with β=0.007	RMSE= 10.2% (corner) RMSE=5.1% (average)	1	CT 1.3. 13
								2) Turbulent mixing model with β=0.007	RMSE=23.1% (Corner) RMSE=9.6% (average)	1	
								3) Turbulent mixing model with Rogers and Rosehart correlation for β	RMSE=10.8% (corner) RMSE=4.9 (average)	1	
G_p				1	2	0	CT 1.3. 14 CT 1.2. 1 CT 1.3. 16	**<u>GE 3x3</u>** Turbulent mixing and void drift model, single phase mixing coefficient=0.007 and Beus two phase mixing multiplier =5	Graphical Corner channel	1	CT 1.3. 13
									Inner channel	2	
									Side channel	2	

Table 4.9: Evidence for of turbulent mixing in two phase flow (based on CTF validation and verification report [100])

Index	Level of detail	Description	Credibility	Reference
CT 1.3. 10	LLC	Validation result indicate predicted exit quality fall within experimental uncertainty except for the corner subchannel Corner channel quality prediction error is double the inner and side type channel prediction. Quality in corner is over predicted by CTF.	IA	Section 8.1.1 in [100]
CT 1.3. 11	LLC	Corner channel RMSE increases when void drift mode is turned off and only turbulent mixing model is used.	IA	Section 8.1.1 in [100]
CT 1.3. 12	LLC	When using Rogers and Rosehart correlation exit quality fall within experimental uncertainty except for the corner subchannel	IA	Section 8.1.1 in [100]
CT 1.3. 13	LLC	rRMS for exit mass flux (of individual subchannel) for two phase results are larger than single phase result in section 6.23. Corner channel is most poorly predicted of all the channel	IA	Section 8.1.1 in [100]
CT 1.3. 14	LLC	GE 3x3 is a classic test for assessing inter-subchannel mixing because mass flux and quality measurement are available for individual subchannel	IA	[102], Section 3.5 in [100]
CT 1.2. 1	LLC	In GE 3x3 pins holding the rods in place acts as spacer. Six pin type spacers are used	IA	Section 3.5 in [100]
CT 1.3. 15	LLC	2% uncertainty in quality measurement.	IA	Section 8.1.1 in [100]
CT 1.3. 16	LLC	Measurement uncertainty for flow measurement not reported	IA	NA

4.2.3.3. Natural circulation in single phase flow

This sub-section present tables for classification and characterization of evidence for assessment of CTF ability to simulate natural circulation.

Table 4.10: Characterization of natural circulation in single phase flow

$QOI = \{T_{sc}, v_l\}$		Governing system condition (set of parameters)		Parameter range for Phenomena, model and data			
Symbol	Description	Symbol	Description	Phenomena	Data		Model
					Range	Test	
T_{sc}	Subchannel-center temperature [F]	X_{sc}	Different subchannels (corner, center, side)	NA	NA	PNNL 2x6	NA
v_l	local velocity in sub-channel	X	Different axial positions in subchannel	NA	NA	Kumamoto 2x3	NA

Table 4.11: Classification and characterization of evidence natural circulation in single phase flow [with capability grade (CG→0/1/2/3/4)]

QOI	Coverage			Data Applicability (DA) with capability grade (CG))				Validation test result (VTR) with capability grade			
	CMP	CME	CEP	CG			Evidence reference	Specific model /correlation evaluated	Metric	CG	Evidence reference
				R	PS	U					
T_{sc}	TBA	TBA	TBA	2	2	0	CT 1.2. 2 CT 1.3. 19	Mixing model with Rogers and Rosehart correlation	Graphical	2	CT 1.3. 17
v_l	TBA	TBA	TBA	2	2	0	CT 1.2. 2 CT 1.3. 19	Mixing model with Rogers and Rosehart correlation	Graphical	2	CT 1.3. 18

Table 4.12: Evidence for natural circulation in single phase flow (based on CTF validation and verification report [100])

Index	Level of detail	Description	Credibility	Reference
CT 1.2. 2	LLC	PNNL2x6 proved provide benchmark data to study effect of buoyancy on flow patterns	IA	Section 3.3 in [100]
CT 1.3. 17	LLC	CTF capture the effect of the velocity distribution, which should be for all axial locations at rake locations Y=0.0 inch and Y=0.0581 inch. CTF overpredicts velocity for Y=-0.581	IA	Section 11.1 in [100]
CT 1.3. 18	LLC	CTF over predicts temperature in the lower axial region of the bundle (laminar flow) CTF matches well with the measured temperature at higher axial region (turbulent flow)	IA	Section 11.1 in [100]
CT 1.3. 19	LLC	Measurement uncertainty not reported	IA	NA

4.2.3.4. Pressure drop in single phase flow

This sub-section present tables for classification and characterization of evidence for assessment of CTF ability to simulate pressure drop in single phase flow condition.

Table 4.13: Characterization of pressure drop in single phase flow (PD in SPF)

$QOI = \{\Delta P_f\}$		Governing system condition (set of parameters)		Parameter range for Phenomena, model and data			
Symbol	Description	Symbol	Description	Phenomena	Data		Model
					Range	Test	
ΔP_f	Frictional pressure drop	Re	Reynolds number	TBA	80000-280000	BFBT	TBA

Table 4.14: Classification and characterization of evidence pressure drop in single phase flow [with capability grade (CG→0/1/2/3/4)]

QOI	Coverage			Data applicability (DA) with capability grade (CG)				Validation test result (VTR) with capability grade (CG)			
	CMP	CME	CEP	CG			Evidence reference	Specific model /correlation evaluated	Metric	CG	Evidence reference
				R	PS	U					
ΔP_f	TBA	TBA	TBA	3	2	3	CT 1.3. 21 CT 1.3. 22	CTF frictional pressure drop calculation	rRMS=6.4% (mean) rRMS=1.6% (min) rRMS=10%	2	CT 1.3. 20

Table 4.15: Evidence for pressure drop in single phase flow (based on CTF validation and verification report [100])

Index	Level of detail	Description	Credibility	Reference
CT 1.3. 20	LLC	Higher discrepancy between measured and predicted result is observed at lower Reynolds number, rRMS is between 1% to 10%	IA	Section 5.2.1 in [100]
CT 1.3. 21	LLC	BWR full size fine mesh bundle test (BFBT)-Steady state pressure drop benchmark test (8x8)	IA	Section 3.2 in [100]
CT 1.3. 22	LLC	Experimental uncertainty for pressure drop measurement is 1 % (check unit)	IA	Section 5.2.1 in [100]

4.2.3.5. Pressure drop in two phase flow

This sub-section present tables for classification and characterization of evidence for assessment of CTF ability to simulate turbulent mixing in two phase flow condition.

Table 4.16: Characterization of pressure drop in two phase flow (PD in TPF)

$QOI = \{\Delta P_f\}$		Governing system condition (set of parameters)		Parameter range for Phenomena, model and data			
Symbol	Description	Symbol	Description	Phenomena	Data		Model
					Range	Test	
ΔP_f	Frictional pressure drop	x_e	Average exit quality	TBA	6-26%	BFBT	TBA

Table 4.17: Classification and characterization of evidence for pressure drop in two phase flow [with capability grade (CG→0/1/2/3/4)]

QOI	Coverage			Data applicability (DA) with capability grade (CG)				Validation test result (VTR) with capability grade (CG)			
	CMP	CME	CEP	CG			Evidence reference	Specific model /correlation evaluated	Metric	CG	Evidence reference
				R	PS	U					
ΔP_f	TBA	TBA	TBA	3	2	3	CT 1.3. 24 CT 1.3. 25 CT 1.3. 23	<u>BFBT test</u> CTF frictional pressure drop calculation	rRMS=11% (mean)	2	CT 1.3. 23
ΔP_T	TBA	TBA	TBA	1	1	1	CT 1.3. 27 CT 1.3. 28		rRMS=6.3%	2	CT 1.3. 26

Table 4.18: Evidence for pressure drop in two phase flow (based on CTF validation and verification report [100])

Index	Level of detail	Description	Credibility	Reference
CT 1.3. 23	LLC	rRMSE lies between 2.9 to 19%, with an average of 11% The total bundle pressure drop match experimental results fairly close, it is the top span locations that produces large deviation from measurement	IA	Section 5.3.1.1 in [100]
CT 1.3. 24	LLC	BWR full size fine mesh bundle test (BFBT)-Steady state pressure drop benchmark test (8x8)	IA	Section 3.2 in [100]
CT 1.3. 25	LLC	<u>BFBT</u> Experimental uncertainty of pressure drop in BFBT is 1 % (check unit)	IA	Section 5.3.1.1 in [100]
CT 1.3. 26	LLC	<u>FRIGG test</u> CTF was able to match behavior of all three pressure components, i.e, acceleration pressure drop, frictional pressure drop, gravitational pressure drop rRMS=6.3%	IA	Section 5.3.1.2 in [100]
CT 1.3. 27	LLC	<u>FRIGG test</u> The fuel assembly lattice is much different than the typical U.S. PWR. It has circular shaped assembly bundle.	IA	Section 3.7 in [100]
CT 1.3. 28	LLC	<u>FRIGG test</u> The quantities used to plot experimental pressure drops were obtained from the original report using digitizer, so it may introduce additional error in measured values. The authors specification is not clear about how the components of the total pressure drop were obtained	IA	Section 5.3.1.2 in [100]

4.2.3.6. Nucleate boiling

This sub-section present tables for classification and characterization of evidence for assessment of CTF ability to simulate nucleate boiling.

Table 4.19: Characterization of nucleate boiling

$QOI = \{T_s\}$		Governing system condition (set of parameters)		Parameter range for Phenomena, model and data			
Symbol	Description	Symbol	Description	Phenomena	Data		Model
					Range	Test	
T_s	Rod surface temperature	q''	Rod power (Average heat flux) $\left[{MBTU}/{Hr.ft^2}\right]$	TBA	0.1-0.6	CE 5x5	TBA

Table 4.20: Classification and characterization of evidence for nucleate boiling [with capability grade (CG→0/1/2/3/4)]

QOI	Coverage			Data applicability (DA) with capability grade (CG)				Validation test result (VTR) with capability grade (CG)			
	CMP	CME	CEP	CG			Evidence reference	Specific model /correlation evaluated	Metric	CG	Evidence reference
				R	PS	U					
T_s	TBA	TBA	TBA	3	1	2	CT 1.3. 30 CT 1.3. 31	Heat transfer model Thom correlation	Bias=-5F (mean)	1	CT 1.3. 29

Table 4.21: Evidence for nucleate boiling (based on CTF validation and verification report [100])

Index	Level of detail	Description	Credibility	Reference
CT 1.3. 29	LLC	CE 5x5 involved 5x5 electrically heated rod bundle with varying operating condition. Heat transfer mechanism in the bundle ranges from single phase convection to saturated boiling	IA	Section 4.1 in [100]
CT 1.3. 30	LLC	Comparison of prediction with experimental results highlight outliers with discrepancy greater than 50 F at higher heat flux These outliers are removed which leads to a mean discrepancy -5F between experiment and measurement	IA	Section 4.1 in [100]
CT 1.3. 31	LLC	2σ scatter in measurement is reported	IA	Section 4.1 in [100]

Table 4.22: Description of gap for validation of CTF for departure from nucleate boiling

Phenomena	Gap Description	Reference
Cross-flow	Lack of data	NA
Flow regime	Lack of data to support transient and transition flow regime	NA
Critical heat flux (CHF)	Model lack predictive capability for different surface and fuel bundle geometry Data for validation is available but test not completed yet	CTF theory manual [103] data reference for CHF[104]
Nucleate boiling (NB)	Model does not capture surface effect	CTF theory manual [103]
Natural circulation in two phase flow condition (NC in TPF)	Data for validation is available but test not completed	NA

4.2.4. Formulation of decision model

This section of the framework illustrates the formulation of the decision model for validation assessment of CTF. The main module of the decision model is shown in Figure 4.5. Goal 1 represents the top claim of the decision model, i.e., "Validation assessment of CTF for DNB is completed." This claim is broken down into two sub-claims (Goal 1.1 and Goal 1.2) based on the nature of validation evidence. These sub-claims accounts for assessment based on direct validation attribute (Goal 1.1) and process quality assurance factor (Goal 1.2). The decision model presented in Figure 4.5 is based on the strong assumption that validation of CTF for DNB is based on the assessment of the capability of CTF to simulate the phenomena identified by the PIRT. This assumption is specified in assumption (Assumption:1b) block in the GSN based decision model.

The structure of the decision model is defined by different validation attributes discussed in section 3.5.3. Direct validation attributes are evaluated based on the claim regarding the assessment of data applicability (Goal 1.1.1) and validation results (Goal 1.1.2). Process quality assurance (PQA) is evaluated based on the claim regarding assessment of process quality assurance (PQA) factors for the phenomenology pyramid (PQA for PP, Away goal 1.2.2) and process quality assurance factors for the validation evidence assessment (PQA for VEA, Away goal 1.2.1). Data applicability (Goal 1.1.2) is evaluated based on the claim regarding data uncertainty assessment (DUA, Away goal 1.1.1), physics scaling assessment (PSA, Away goal 1.1.1.2.1) and data relevance assessment (DRA, Away goal 1.1.1.2.2). Validation results are assessed based on the claim regarding data coverage assessment (DCA, Away goal 1.1.2.1) and assessment of validation test result (VTR, Away goal 1.1.2.2).

All away goal in the main module for validation assessment of CTF (Module D1, Figure 4.5) are resolved in individual GSN module. The GSN modules for data uncertainty assessment (DUA), physics scaling assessment (PSA), data relevance assessment and validation test result are shown in Figure 4.6 to Figure 4.10. The assessment of all direct validation attribute is based on the assessment of individual attribute for all phenomenon in the phenomenology pyramid (PP) in Figure 4.3 (i.e., the phenomena identified by the PIRT). Following the steps of transformation discussed in section 3.5.4, we obtain the computable network (Bayesian network) for quantitative maturity assessment. The Bayesian network corresponding to the GSN module in Figure 4.5 to Figure 4.12 is shown in Figure 4.13 to Figure 4.23.

The target level for all higher-level nodes (or attributes) in the decision model is based on the target level for evidence nodes. The required target level for all direct validation attribute for each phenomenon assessment is assumed to be “High”. The target level for process quality assurance of phenomenology pyramid (PQA for PP) and process quality assurance for validation evidence assessment (PQA for VEA) process is shown in Figure 4.19 and Figure 4.21, respectively. The target level for all attribute in the main decision module is shown in Figure 4.23

It is evident from the tables for classification and characterization of evidence in section 4.2.3 that each phenomenon is supported by multiple sets of evidence acquired from different validation test. Multiple evidence are incorporated into the evidence node by following the scheme for multiple evidence assessment illustrated in section 3.5.4.2. The estimation of the probability distribution for validation test results (VTR) for turbulent mixing in single phase flow (TM in SPF) based on the capability grade in Table 4.5 (column 4 for validation test result, VTR) is shown in Table 4.23 and Figure 4.4 for further illustration. Following similar technique, the probability distribution for all evidence nodes with multiple evidence is estimated.

Table 4.23: Estimation of probability distribution of validation test result (VTR) for turbulent mixing in single phase flow, TM (SPF), based on multiple evidence from Table 4.5

Probability, $P(VR_{M_{i1}}) = \frac{f_{CG} \times w_e}{\sum(f_{CG} \times w_e)}$	0%	46%	40%	14%
Weight assignment$(f_{CG} \times w_e)$	0	4×70	2×30	2×30
Frequency of grade (f_{CG})	0	4	2	3
Capability Grade (CG)	NA '0'	Low '1'	Medium '2'	High '3'
Evidence weight (w_e)	70 % (counter evidence)		30 % (supporting evidence)	

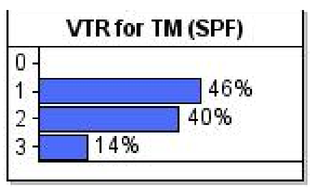


Figure 4.4: Probability distribution of VTR for TM (SPF)

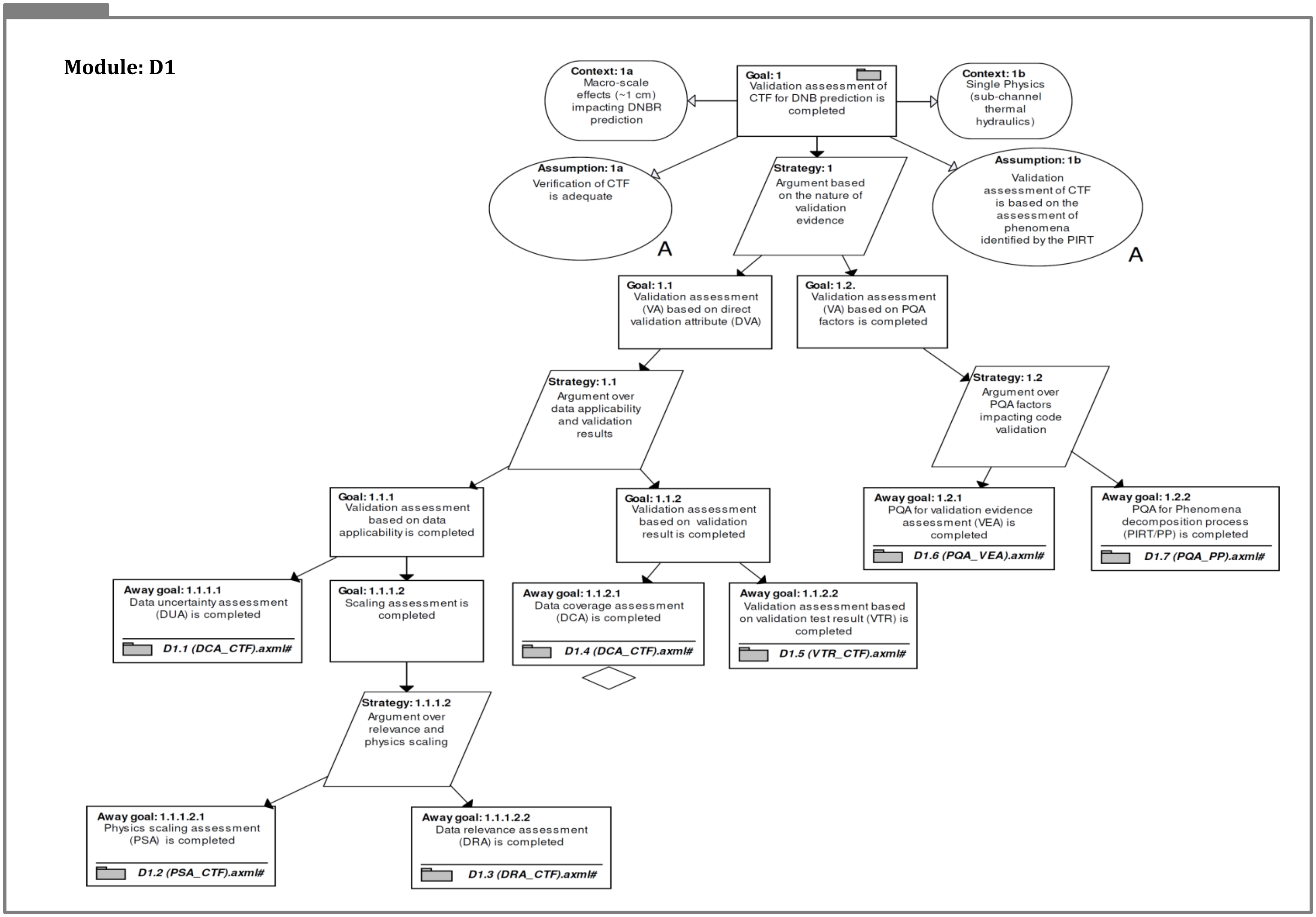


Figure 4.5: Decision model for validation assessment of CTF for DNB (Main module D1)

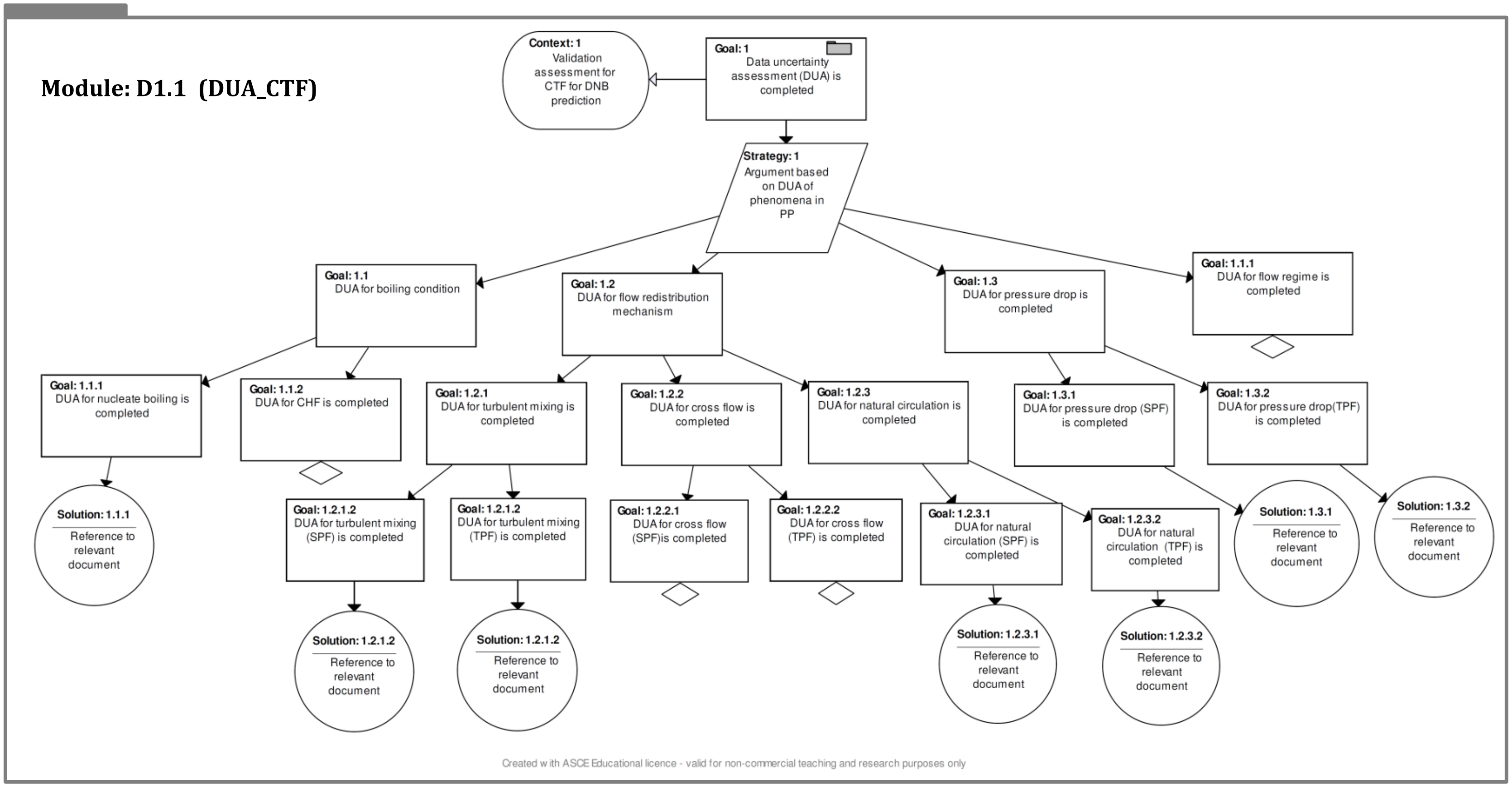


Figure 4.6: Module for data uncertainty assessment (DUA) of CTF, Module: D1.1 (DUA_CTF), corresponding to Away goal: 1.1.1.1 in decision module for validation of CTF (Figure 4.5)

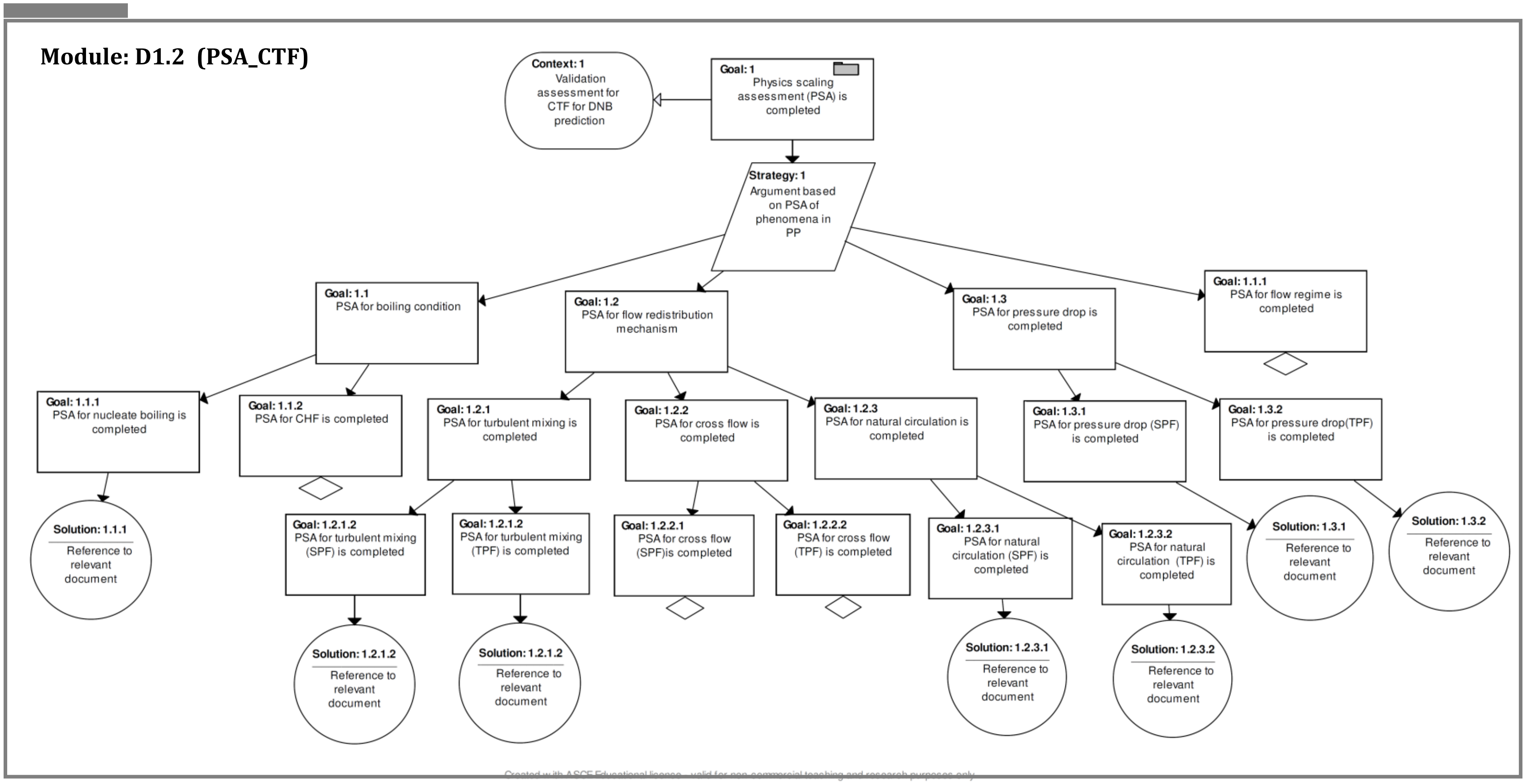


Figure 4.7: Module for physics scaling assessment (PSA) of CTF, Module: D1.2 (PSA_CTF), corresponding to *Away goal*: 1.1.1.2.1 in decision module for validation of CTF (Figure 4.5)

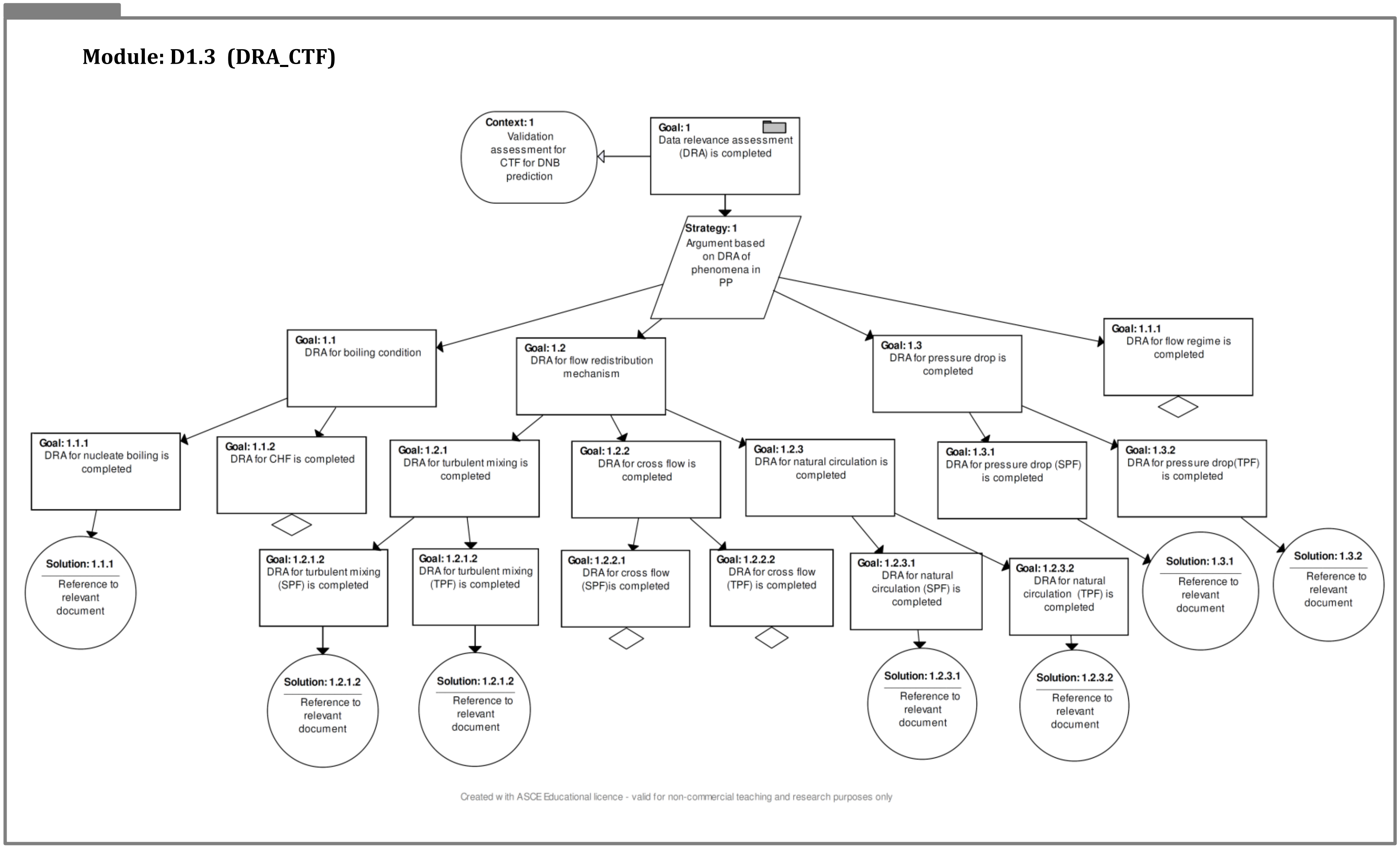


Figure 4.8: Module for data relevance assessment (DRA) of CTF, Module: D1.3 (DRA_CTF), corresponding to Away goal: 1.1.1.2.2 in decision module for validation of CTF (Figure 4.5)

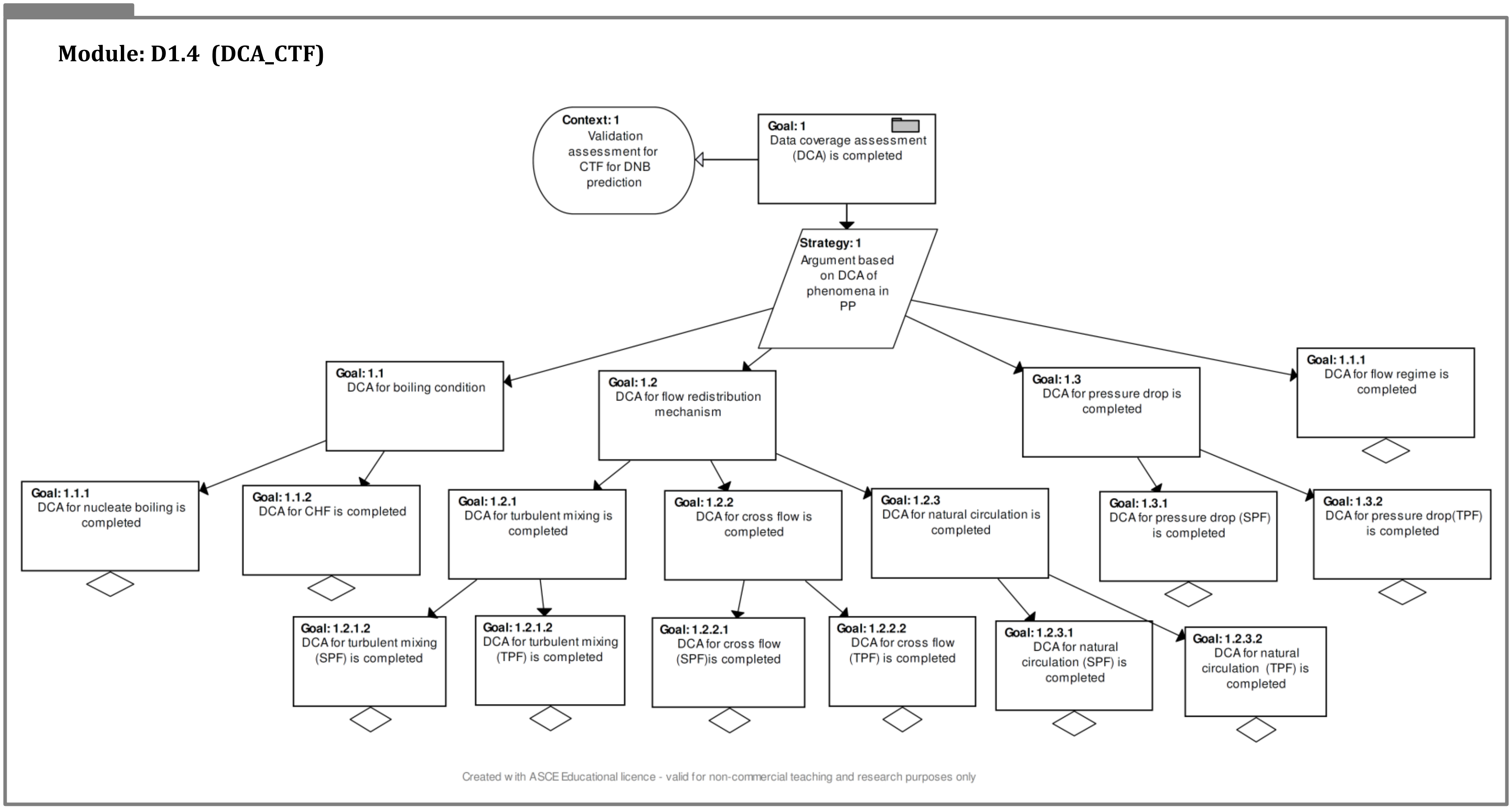


Figure 4.9: Module for data coverage assessment (DCA) of CTF, Module: D1.4 (DCA_CTF), corresponding to Away goal: 1.1.2.1 in decision module for validation of CTF (Figure 4.5)

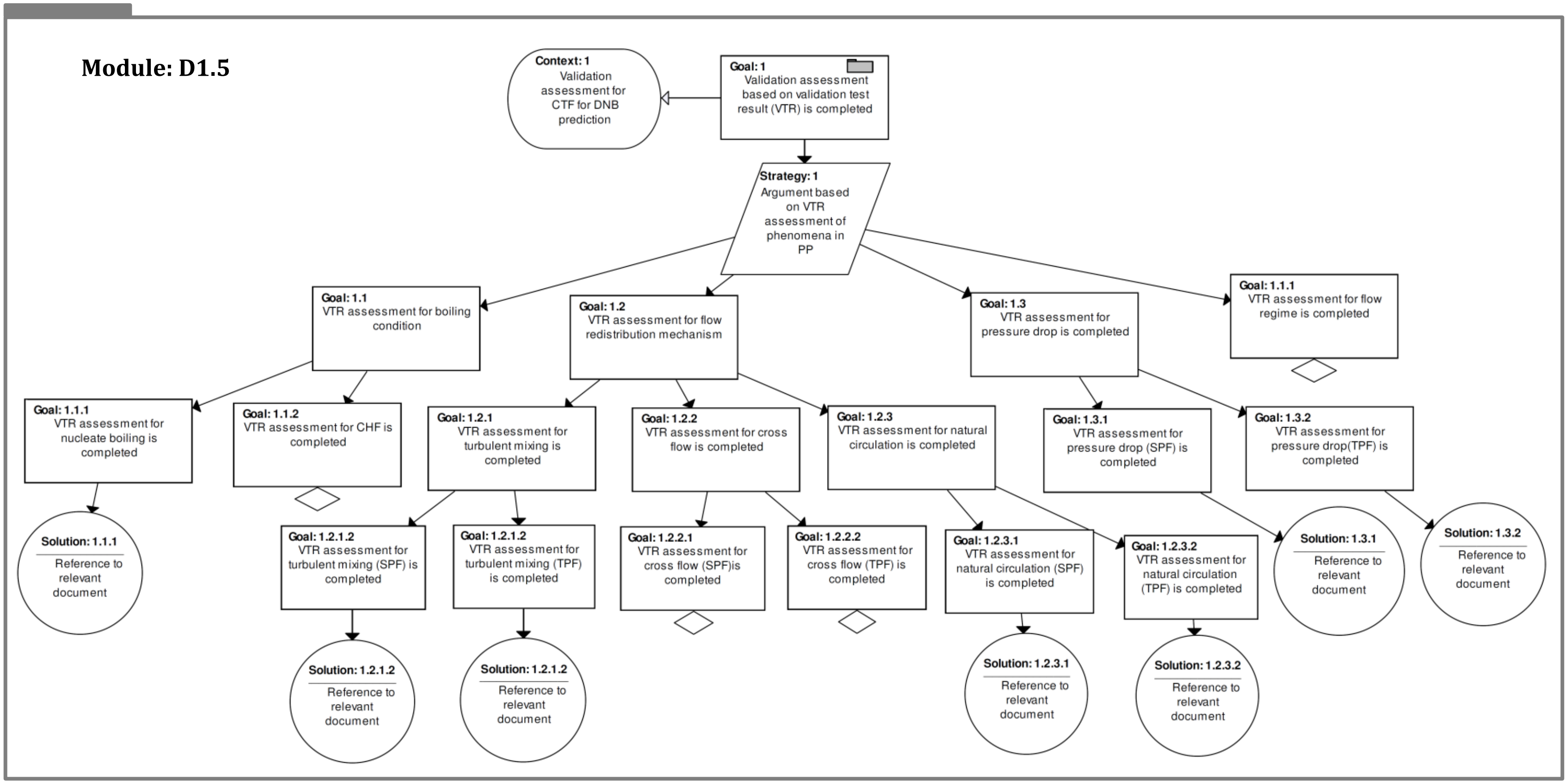


Figure 4.10: Module for assessment of validation test result (VTR) of CTF, Module: D1.5 (VTR_CTF), corresponding to Away goal: 1.1.2.2 in decision module for validation of CTF (Figure 4.5)

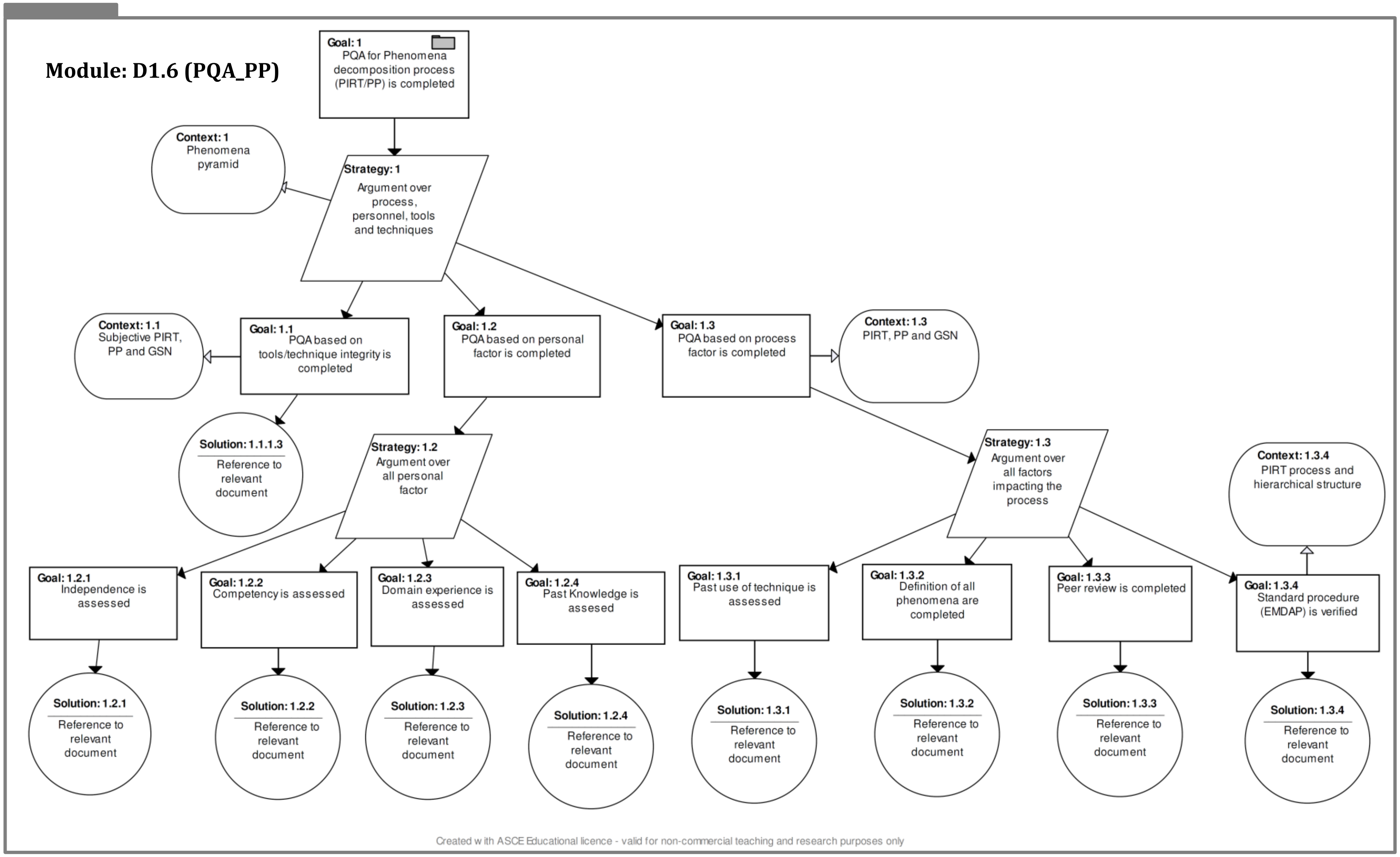


Figure 4.11: Module for process quality assurance for Phenomenology pyramid or PIRT (PQA_PP), Module: D1.6 (PSA_CTF), corresponding to Away goal: 1.2.2 in decision module for validation of CTF (Figure 4.5)

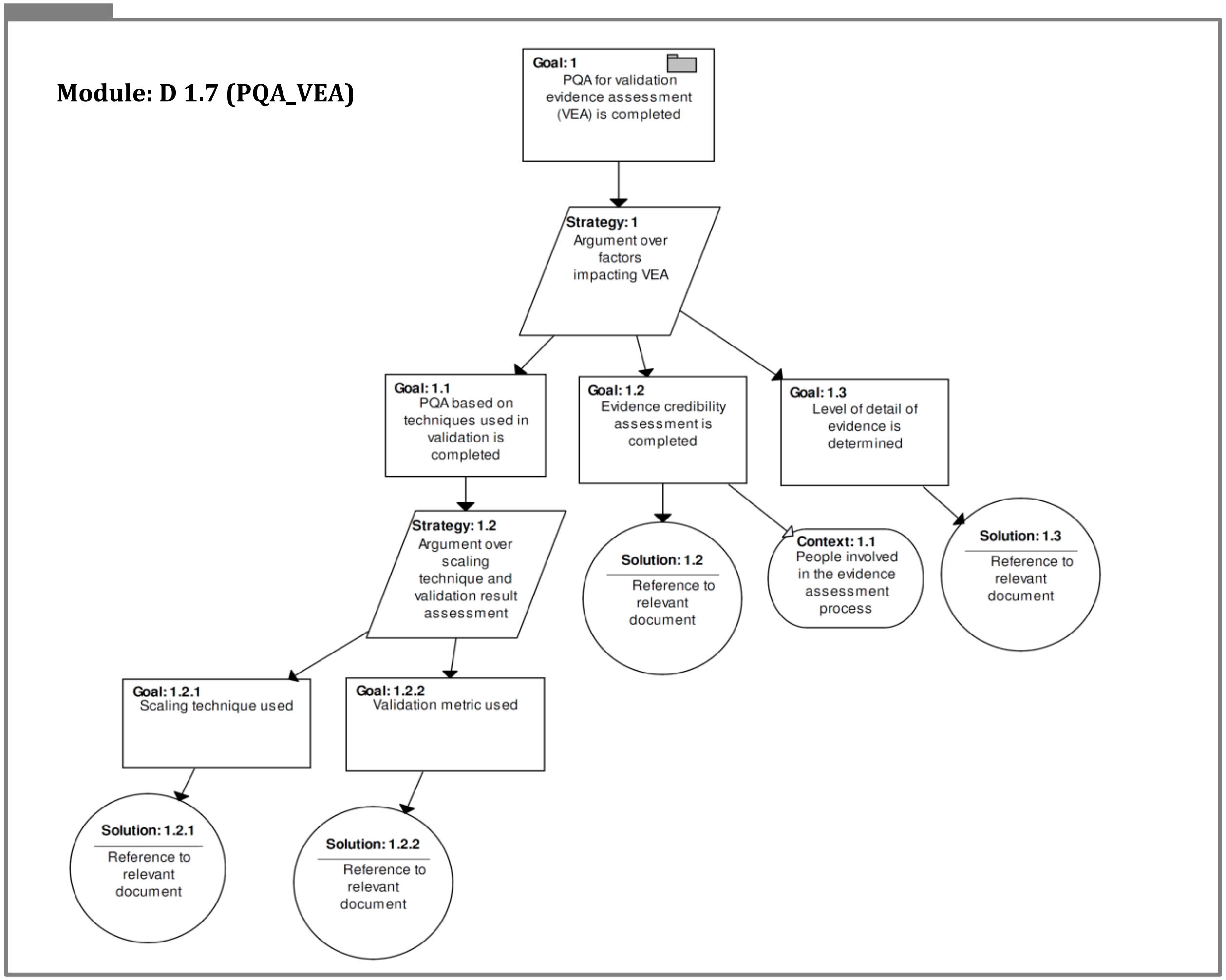


Figure 4.12: Module for PQA for validation evidence assessment, Module: D1.7 (PSA_VEA), corresponding to Away goal: 1.2.1 in decision module for validation of CTF (Figure 4.5)

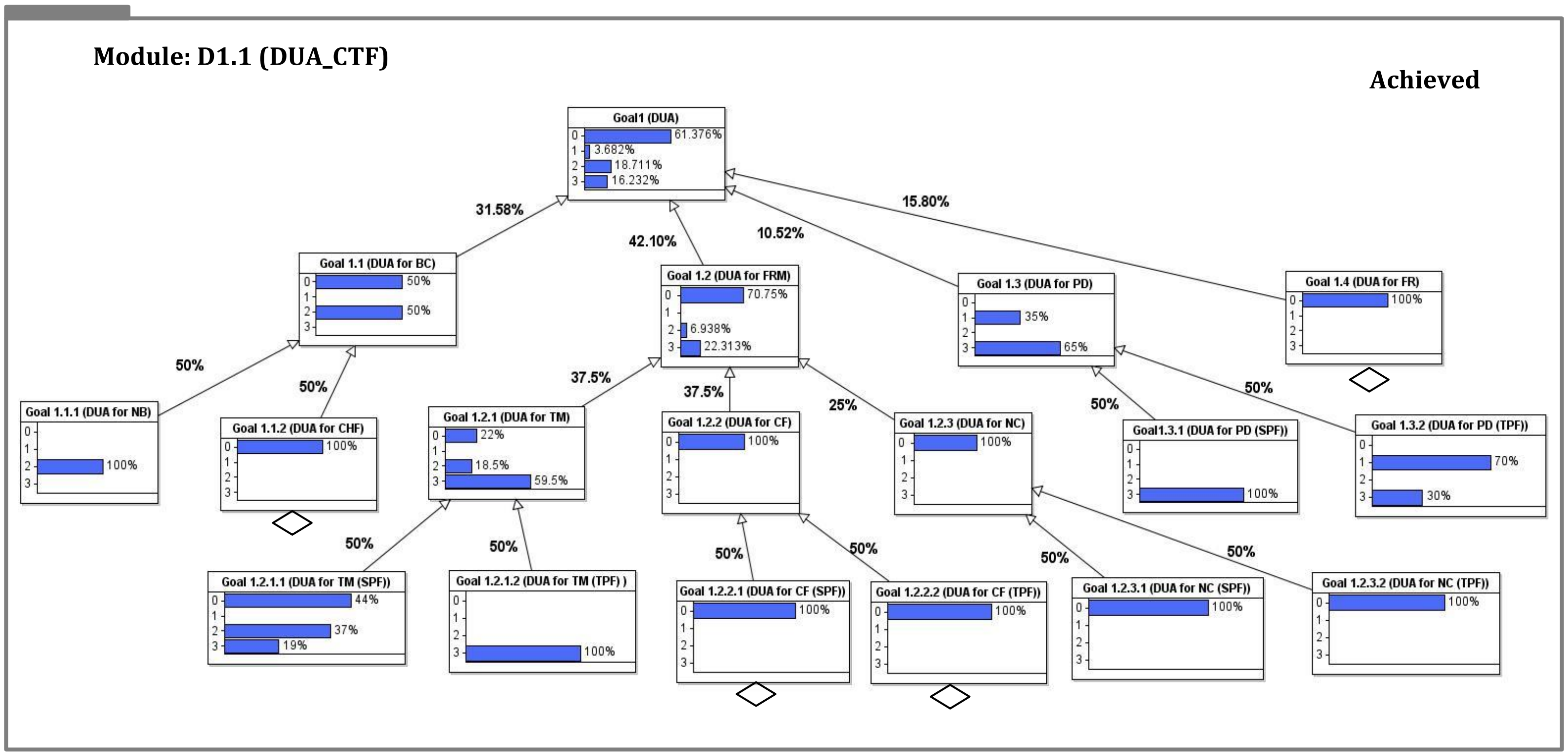


Figure 4.13: Data uncertainty assessment (DUA) for CTF using the Bayesian network (Achieved) based on GSN module in Figure 4.6

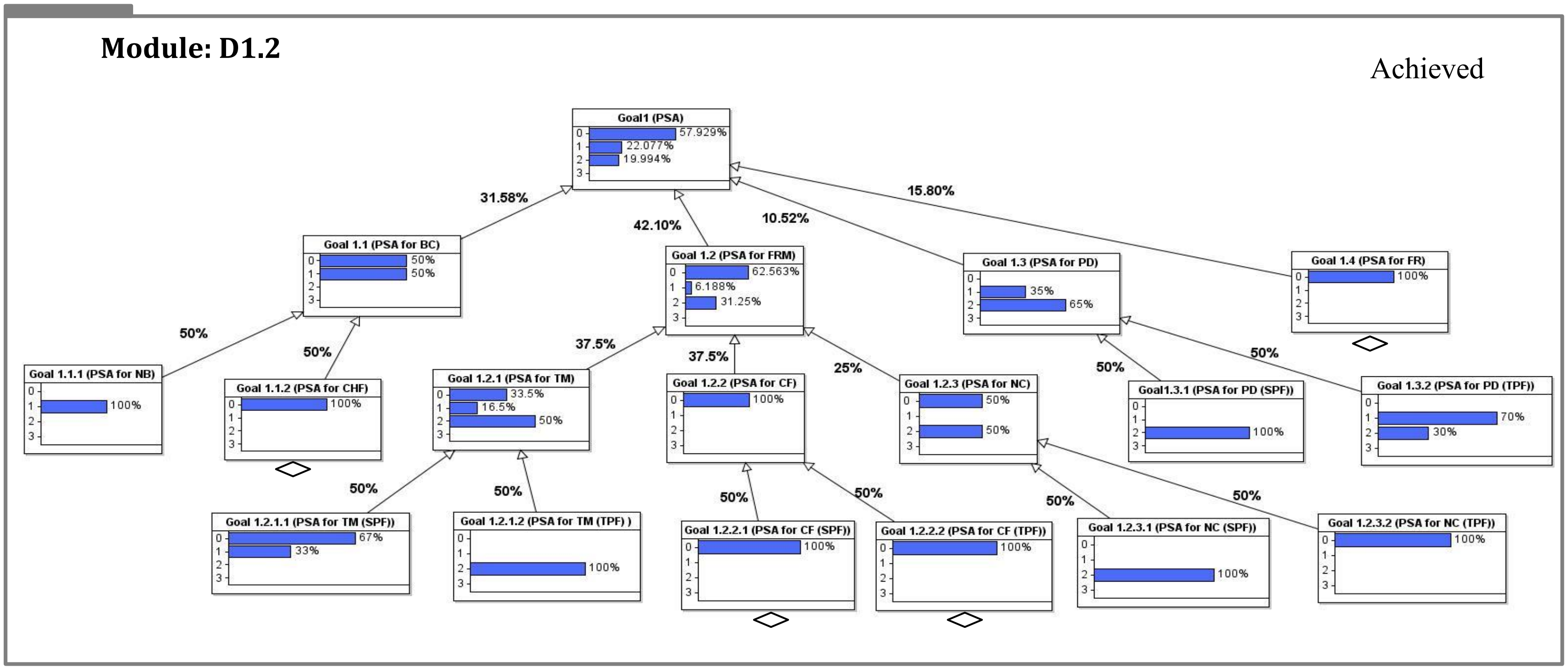


Figure 4.14: Physics scaling assessment (PSA) of CTF using the Bayesian network (Achieved) based on GSN module in Figure 4.7

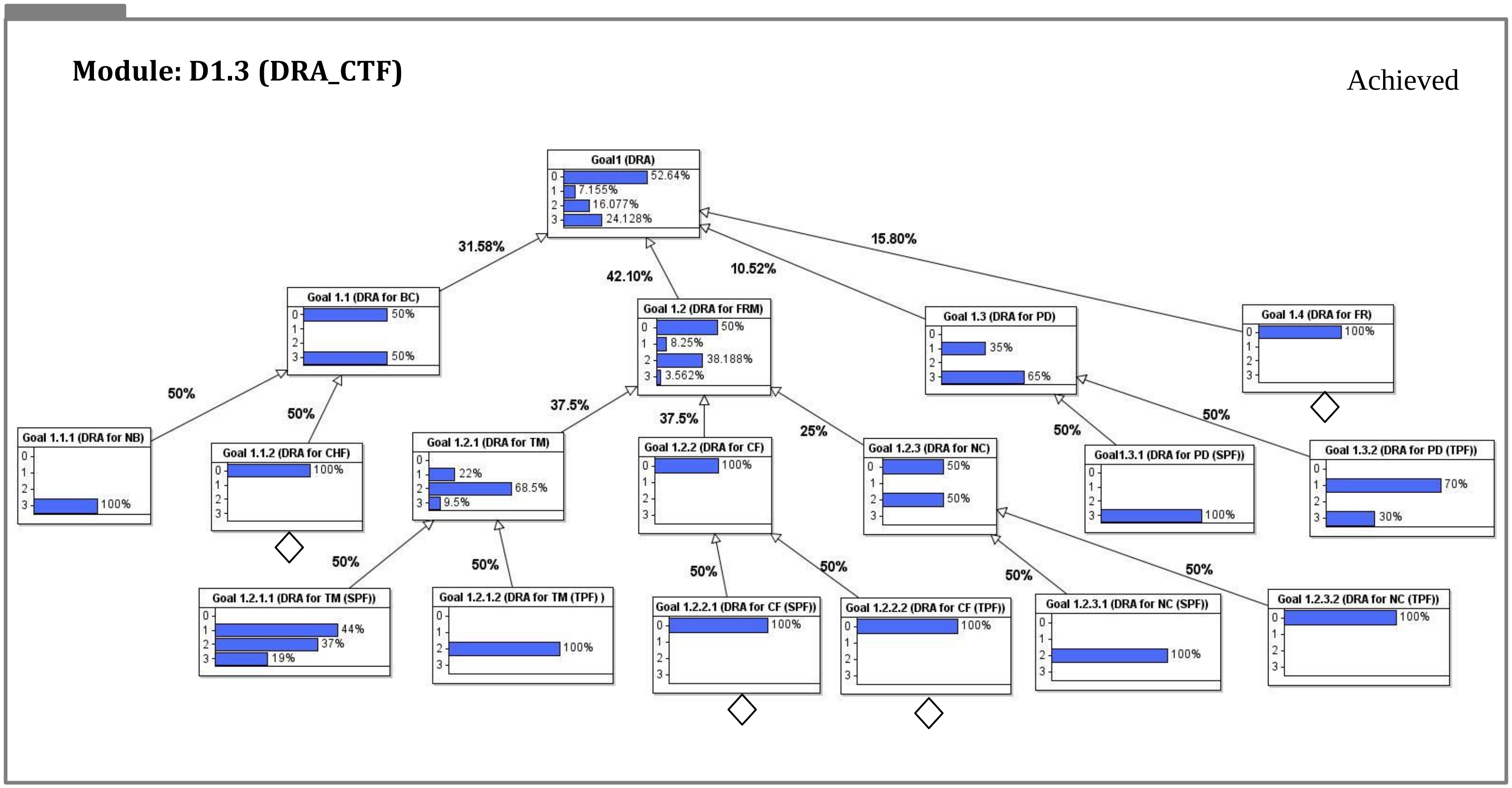


Figure 4.15: Data relevance assessment (DRA) of CTF using the Bayesian network (Achieved) based on GSN module in Figure 4.8

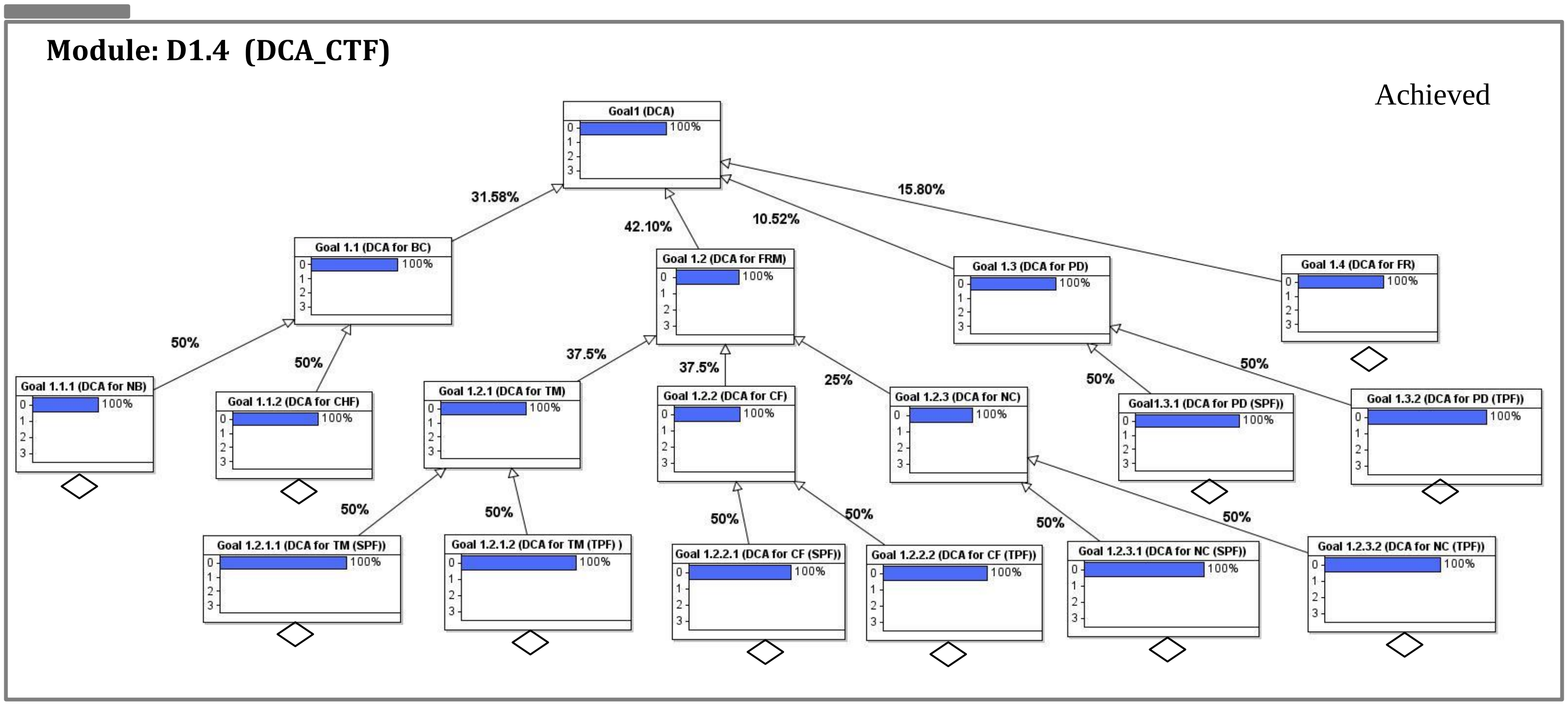


Figure 4.16: Data coverage assessment (DCA) of CTF using the Bayesian network (Achieved) based on GSN module in Figure 4.9

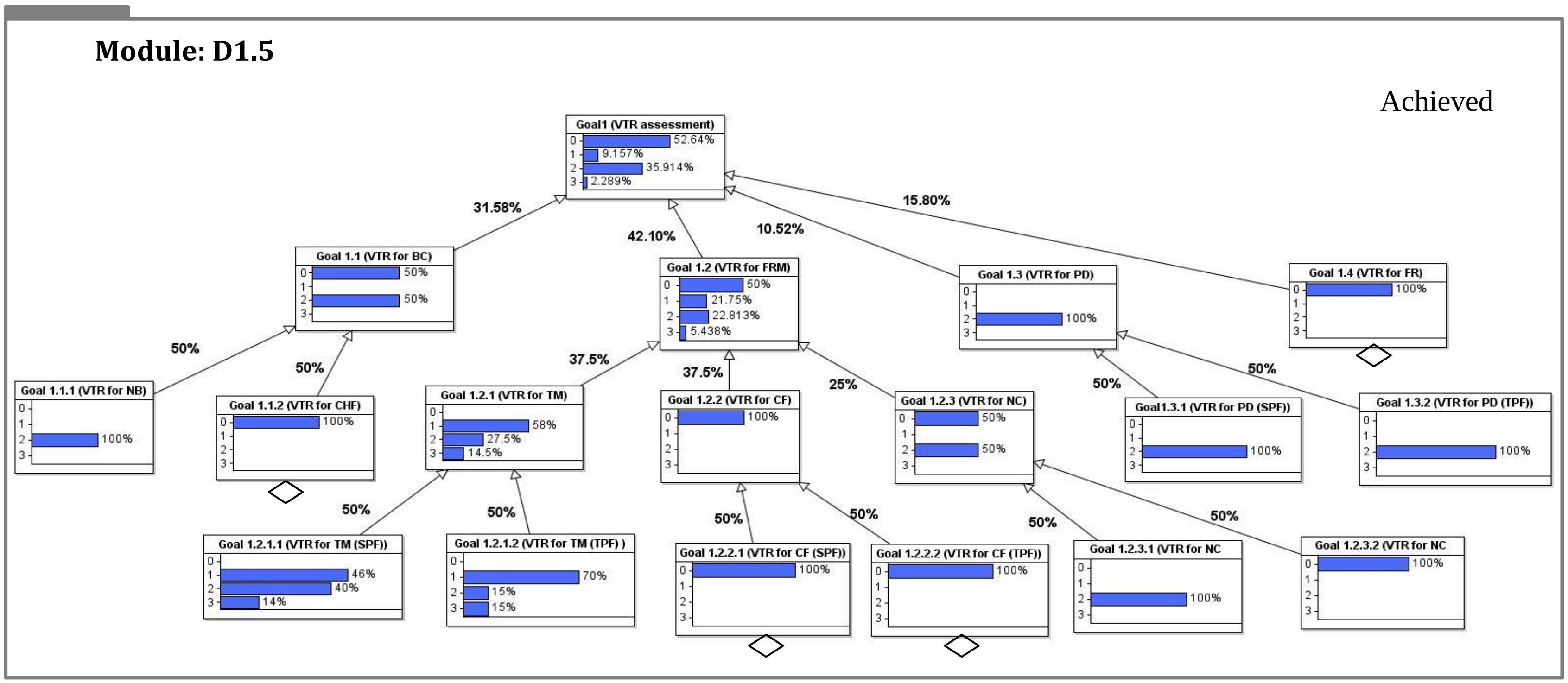


Figure 4.17: Assessment of validation test result (VTR) using the Bayesian network (Achieved) based on GSN module in Figure 4.10

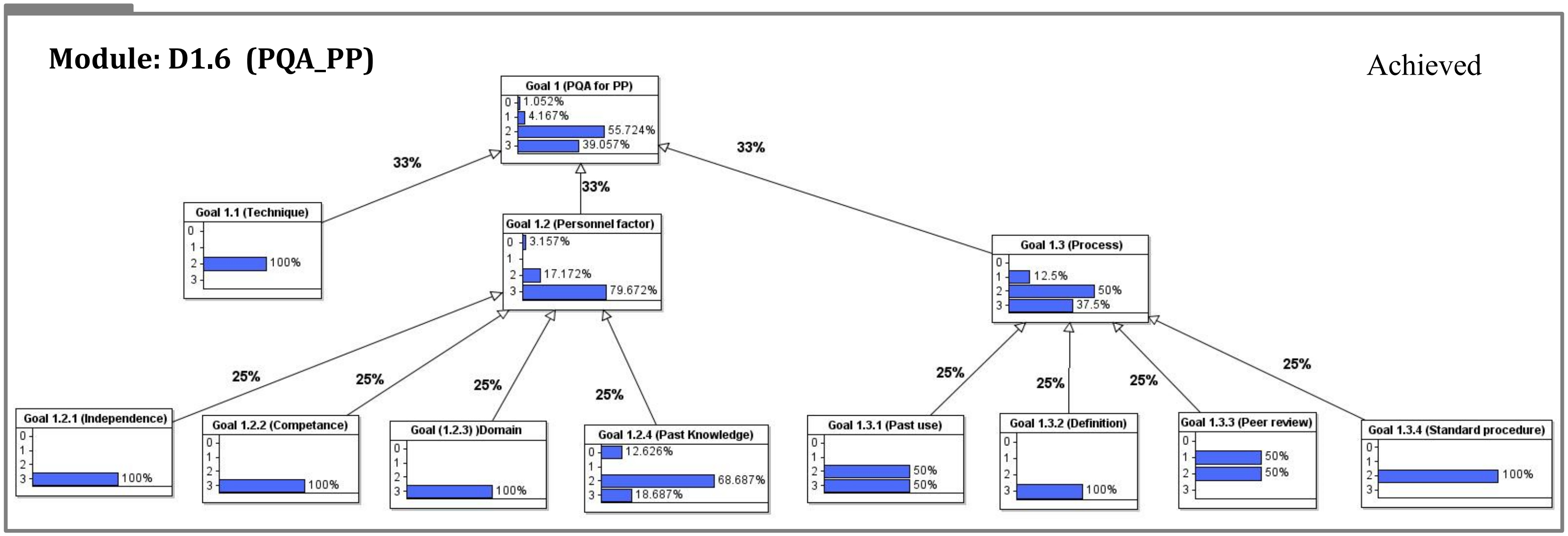


Figure 4.18: Process quality assurance (PQA) for phenomenology pyramid using the Bayesian network (achieved) based on GSN module in Figure 4.11

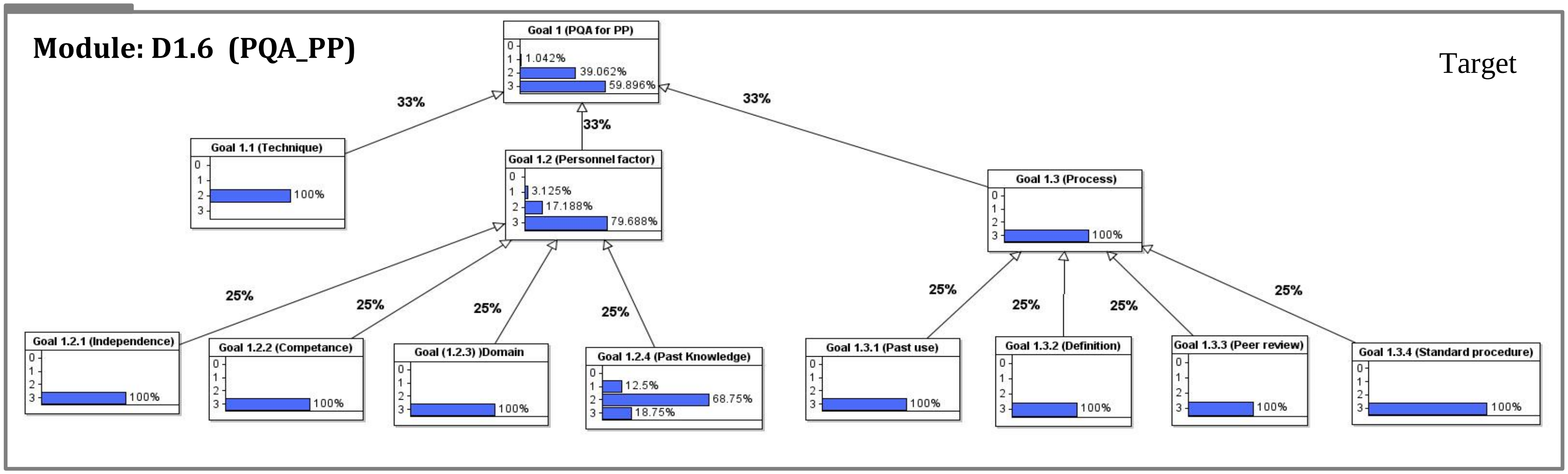


Figure 4.19: Process quality assurance (PQA) for phenomenology pyramid using the Bayesian network (Target) based on GSN module in Figure 4.11

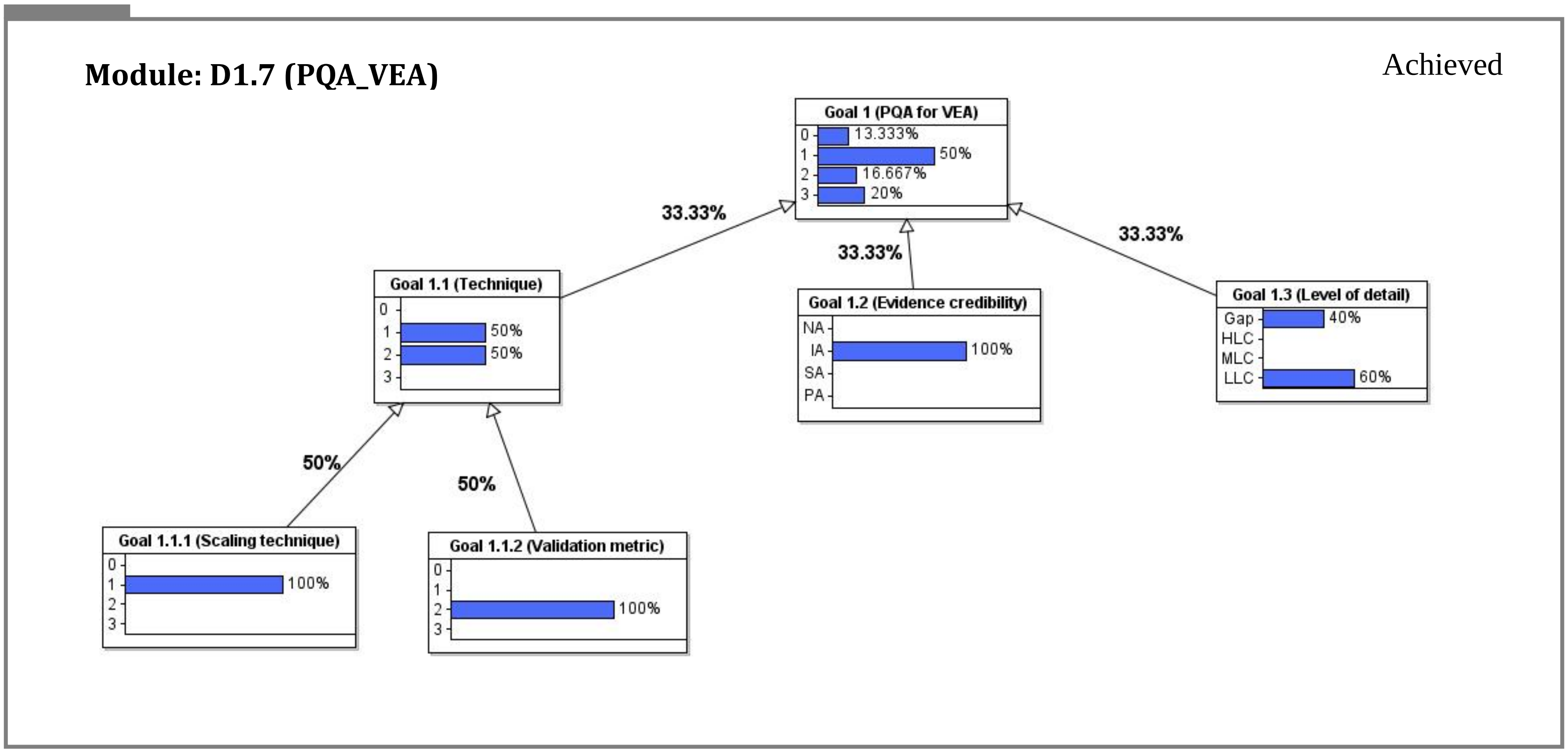


Figure 4.20: Process quality assurance (PQA) for validation evidence assessment (VEA) process (Achieved) based on GSN module in Figure 4.12

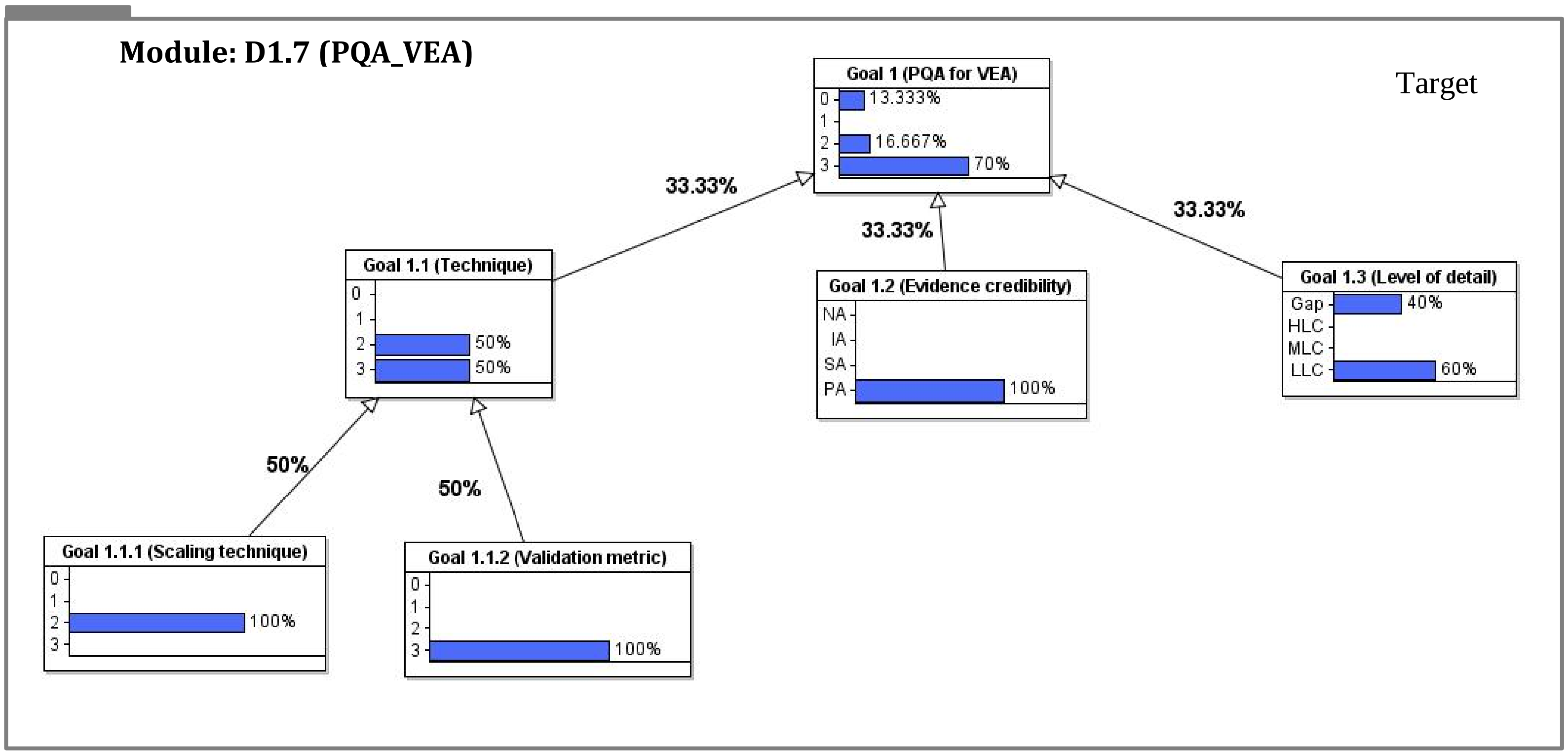


Figure 4.21: Process quality assurance (PQA) for validation evidence assessment (VEA) process (Target) based on GSN module in Figure 4.12

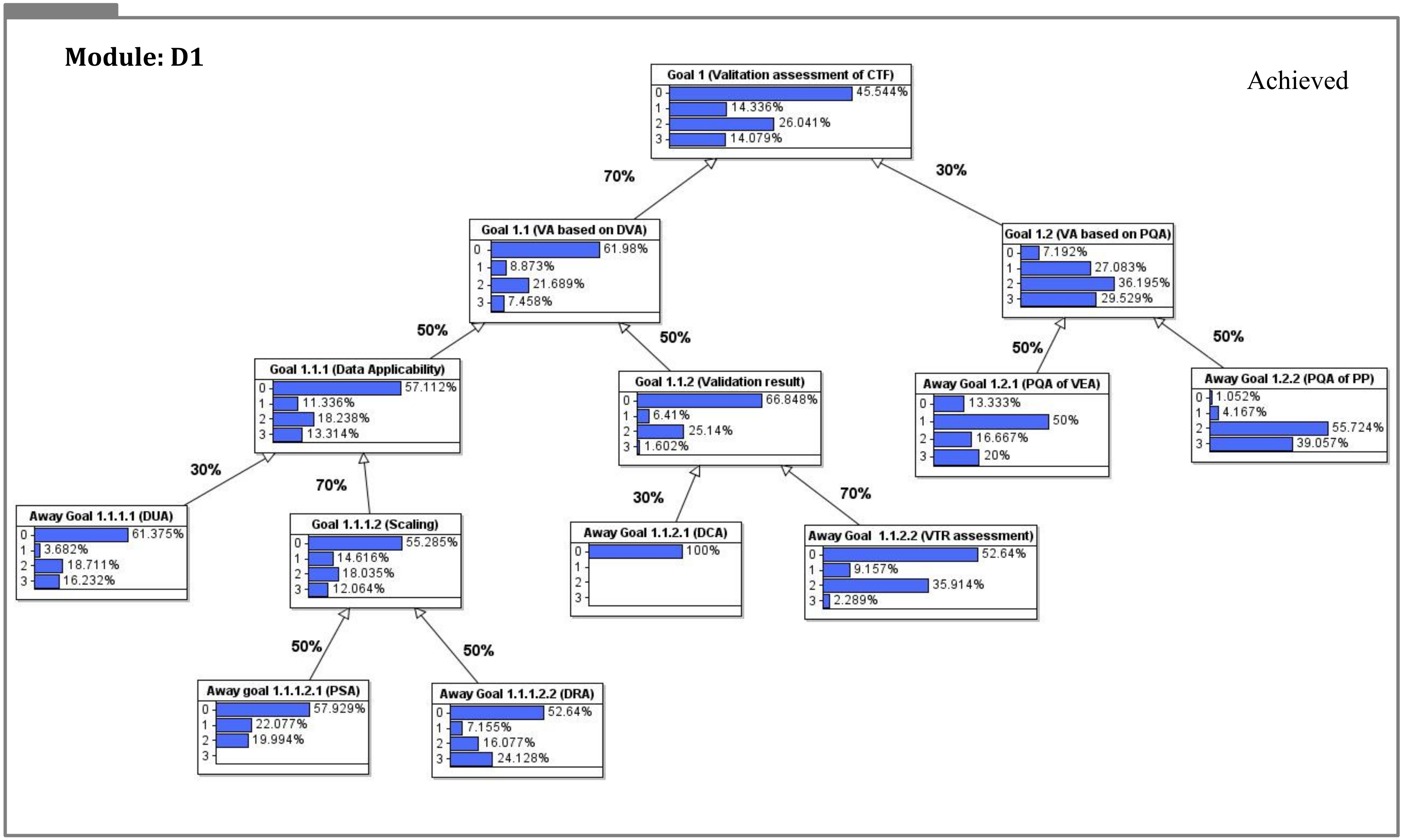


Figure 4.22: Validation assessment for CTF using the Bayesian network (Achieved) based on GSN module in Figure 4.5

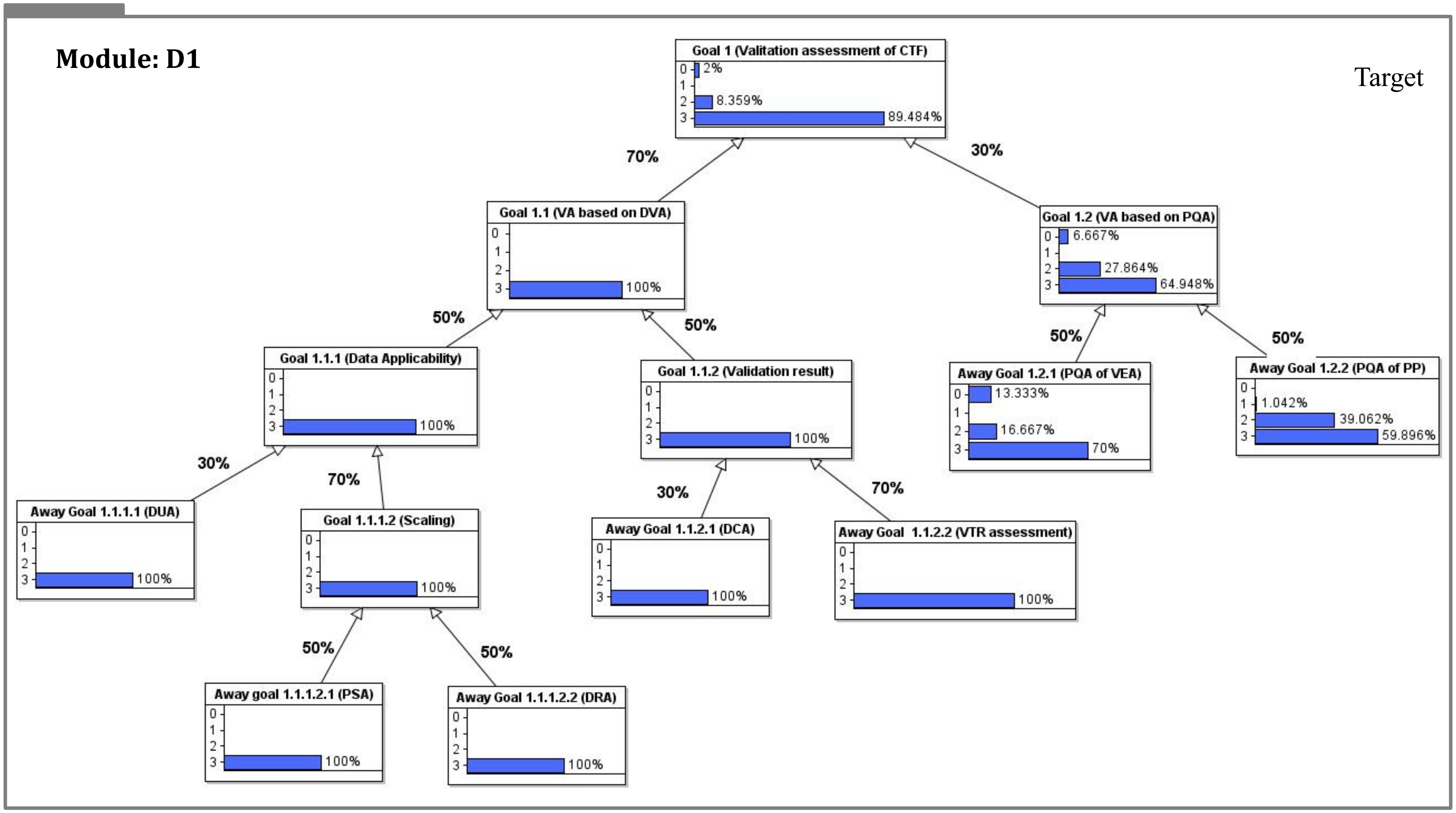


Figure 4.23: Validation assessment of CTF using the Bayesian network (Target) based on GSN module in Figure 4.5

4.2.5. Evaluation and interpretation of result

This section present evaluation and interpretation of result obtained from the formalized decision model developed in the previous section. The evaluation is based on the expected distance metric [Eq. (3.8)] introduced in section 3.5.5. E_N close to zero implies, achieved level is close to the target level. E_N close to 1 implies, achieved level is far from the target level. Table 4.24 presents results of assessment for all the primary validation attribute. It is evident from these result that validation assessment of CTF is incomplete and there is lack evidence to support validation of some phenomena. A more detailed interpretation of the result is provided by assessment of individual validation attribute for each phenomenon (see Table 4.25 to Table 4.28). The interpretation of result presented in these tables is supported by the evidence presented in section 4.2.3. It should be noted that different validation attributes for a phenomenon are evaluated based on the same set of data. One of the experiment (FRIGG test) used in the validation assessment of PD in TPF has “low” grade for data relevance [R], scaling [PS] and uncertainty [U]. However, the grade for validation test result is same as that for the other test (see Table 4.17 for further reference). Therefore, FRIGG test should not be used in validation assessment of CTF as it does not add any value to the current assessment. Process quality assurance for validation evidence assessment process indicates that scaling assessment is based on observation only and proper scaling is needed. It also indicates that the evidence presented in this study are based on initial author assessment and needs to be revised by subject matter expert (see Table 4.29).

Table 4.24: Estimation and interpretation of result for all primary validation attribute

Estimated distance metric		Interpretation/comment
$E_N(DCA)$	1	Data coverage assessment is incomplete
$E_N(PSA)$	0.79	Lack of data to validate some phenomena, PSA for some phenomena is not completed
$E_N(VTR)$	0.70	Lack of data to validate some phenomena, additional test required to increase confidence in simulation of tested phenomena and some model needs improved
$E_N(DUA)$	0.70	Lack of data to validate some phenomena, measurement uncertainty for some test is not reported.
$E_N(DRA)$	0.63	Lack of data to validate some phenomena, DRA for some phenomena is not completed
$E_N(PQA)$	0.25	PQA for VEA needs improvement

Table 4.25: Estimation and interpretation of result for data relevance assessment (DRA)

Estimated distance metric		Interpretation/comment	Evidence reference
$E_N(FR)$	1	No data set available, lack of data to simulate transient and transition flow pattern	Table 4.22
$E_N(CF)$	1	No data set available for cross-flow	Table 4.22
$E_N(CHF)$	1	Data for CHF testing is available, but DRA is incomplete	Table 4.22, data reference [104]
$E_N(NC\_TPF)$	1	DRA for NC in two phase flow condition is incomplete	Table 4.22
$E_N(PD\_TPF)$	0.47	Some data (FRIGG test) is less relevant (Fuel assembly in the test is less relevant to the U.S. PWR)	Table 4.18
$E_N(TM\_SPF)$	0.42	Some experiment used in the test does not have spacer grid, more relevant data required	Table 4.6 (CT 1.3. 7)
$E_N(NC\_SPF)$	0.33	Data is medium level relevant, addition data required for higher confidence	Table 4.11
$E_N(TM\_TPF)$	0.33	Pins holding the rod act as spacer, more relevant data required	Table 4.9 (CT 1.2. 1)
$E_N(PD\_SPF)$	0	High level DRA	Table 4.15
$E_N(NB)$	0	High level DRA	Table 4.21

Table 4.26: Estimation and interpretation of result for physics scaling assessment (PSA)

Estimated distance metric		Interpretation/comment	Evidence reference
$E_N(FR)$	1	No data set available for PSA	Table 4.22
$E_N(CF)$	1	No data set available for PSA	Table 4.22
$E_N(CHF)$	1	Data set available, but PSA is incomplete	Table 4.22, data reference [104]
$E_N(NC\_TPF)$	1	PSA for NC in TPF is incomplete	Table 4.22
$E_N(TM\_SPF)$	0.89	PSA for some data set is incomplete	Table 4.6
$E_N(NB)$	0.66	Low confidence in PSA	Table 4.21
$E_N(PD\_TPF)$	0.57	FRIGG test is not appropriately scaled for the application	Table 4.18
$E_N(NC\_SPF)$	0.33	PSA is adequate, but additional data is required	Table 4.11
$E_N(TM\_TPF)$	0.33	Confidence in PSA is medium level	Table 4.9
$E_N(PD\_SPF)$	0.33	Confidence in PSA is medium level	Table 4.15

Table 4.27: Estimation and interpretation of result for data uncertainty assessment (DUA)

Estimated distance metric		Interpretation/comment	Evidence reference
$E_N(FR)$	1	No data set available for DUA	Table 4.22
$E_N(CF)$	1	No data set available for DUA	Table 4.22
$E_N(CHF)$	1	Data set available but DUA is incomplete	Table 4.22, data reference [104]
$E_N(NC\_SPF)$	1	Measurement uncertainty not reported	Table 4.11
$E_N(NC\_TPF)$	1	Measurement uncertainty not reported	Table 4.22
$E_N(TM\_SPF)$	0.56	Measurement error for some data set is not reported	Table 4.6
$E_N(PD\_TPF)$	0.47	Some data set are of low quality (FRIGG test)	Table 4.18
$E_N(NB)$	0.33	Confidence in DUA is medium level	Table 4.21
$E_N(TM\_TPF)$	0	High quality data	Table 4.9
$E_N(PD\_SPF)$	0	High quality data	Table 4.15

Table 4.28: Estimation and interpretation of result for assessment of validation test results (VTR)

Estimated distance metric		Interpretation/comment	Evidence reference
$E_N(FR)$	1	Data for validation not available	Table 4.22
$E_N(CF)$	1	Data for validation not available	Table 4.22
$E_N(CHF)$	1	Validation for CHF not yet completed	Table 4.22, data reference [104]
$E_N(NC\_TPF)$	1	No validation test reported for NC in TPF	Table 4.22
$E_N(TM\_TPF)$	0.52	Validation result show higher error in corner sub-channel for all the tests, model needs improvement	Table 4.9 , Table 4.8
$E_N(TM\_SPF)$	0.44	Validation result show higher error in corner sub-channel, model needs improvement	Table 4.6 (CT 1.3. 9)
$E_N(NC\_SPF)$	0.33	Confidence in VR is medium, additional test needed	Table 4.11
$E_N(PD\_SPF)$	0.33	Higher discrepancy in validation result at low Reynolds number	Table 4.15
$E_N(PD\_TPF)$	0.33	Medium level confidence, additional test required	Table 4.18
$E_N(NB)$	0.33	Medium level confidence, additional test required	Table 4.21

Table 4.29: Estimation and interpretation of result for process quality assurance (PQA) factors for validation evidence assessment (VEA) process

Estimated distance metric		Interpretation/comment	Evidence reference
$E_N(Evidence\_credibility)$	0.66	Evidence evaluation are based on initial author assessment, needs to be reviewed by subject matter experts	Evidence assessment by Paridhi Athe (Ph.D. Student)
$E_N(Scaling\_Technique)$	0.50	Scaling assessment is based on observation, proper scaling analysis is required	CTF V&V report[100]
$E_N(Validation\_metric)$	0.33	Assessment based on graphical or deterministic validation metric	CTF V&V report[100]
$E_N(level\_of\_detail)$	0	Adequate	NA

Table 4.30: Estimation and interpretation of result for process quality assurance (PQA) factors for phenomenology pyramid (PP)

Estimated distance metric		Interpretation/comment	Evidence reference
$E_N(process)$	0.25	Peer review of phenomenology pyramid is incomplete	NA
$E_N(Personnel\_factor)$	0.01	Acceptable	NA
$E_N(Technique)$	0	Adequate	NA

4.2.6. Refinement

Based on the estimation and interpretation of results, we can formulate the following list of action items:

(1) Refinement of decision model: It is evident from the estimation and interpretation of results for PQA factors in Table 4.29 and Table 4.30 that we have higher confidence in PQA of phenomenology pyramid (PP) compared to PQA of validation evidence assessment (VEA). Therefore, weight factor for PQA of phenomenology pyramid should be higher compared to PQA of validation evidence assessment in the decision model. Hence, its recommended to change the weight factor ratio for PQA of PP (Away goal: 1.2.2) and PQA of VEA (Away goal: 1.2.1) in the decision model (in Figure 4.22 and Figure 4.23) from 50:50 to a ratio that is skewed towards PQA of VEA.

Characterization of phenomena should be completed based on the input of SME and this information should be used to complete the coverage assessment.

(2) Refinement of model: Higher error is observed across the corner sub-channel for all the tests used in the assessment of turbulent mixing model (for both single phase flow and two-phase flow). Table 4.28 for validation test results shows $E_N(TM\_SPF) = 0.44$ and $E_N(TM\_TPF) > 0.5$; therefore, turbulent mixing models needs improvement.

(3) Refinement of data: we have following action items for refinement of data:

(a) Validation assessment of CTF is greatly affected by the lack of data. Data for validation assessment of cross-flow and flow regime is not available. It is evident from the PIRT table that both cross-flow and flow regime are important phenomena for DNB; therefore, if the current budget permits, new data acquisition for cross-flow and flow regime should be conducted.

(b) Experimental data for CHF is available. These data set should be used in the validation assessment of CHF.

(c) FRIGG test is not suitable for validation assessment of pressure drop (PD) in two-phase flow (TPF) as $E_N(PD\_TPF) \sim 0.5$ for all attribute (data relevance, scaling, uncertainty) related to data applicability assessment. Therefore, this test should be discarded.

(d) Measurement uncertainty for some data set is not reported. These information needs to be updated.

(4) Refinement of PQA factors: Based on the assessment of PQA factor for validation evidence assessment, we have the following action items:

(a) Current scaling assessment is based on observation only, proper scaling analysis based on dimensionless scaling group should be conducted.

(b) Evidence assessment needs to be reviewed by subject matter expert (SME).

The priority set for the action item is based on the expected distance metric. Items with E_N value close to 1 have higher priority while items with E_N value close to 0 have lower priority.

4.3. Summary remarks

The proposed framework supports the validation assessment of CTF in following way:

- Provides classification and characterization of evidence for validation assessment of different phenomena identified by the PIRT.
- Bring clarity and traceability in the assessment process. Indicator in GSN helps in explicitly specifying undeveloped entities. GSN also helps in explicitly specifying different assumption and contextual information in the framework. Modular GSN helps in manage large networks of decision making elements in the framework.
- Helps in determining the level of maturity for different validation attribute and sub-attribute based on the quality of evidence (i.e. capability grade).
- Facilitates abstraction of maturity information from lower level attribute to higher level attribute for validation using the Bayesian network.
- Provides a measure of distance between target maturity and achieved level of maturity using expected distance metric. Expected distance measure also helps in deciding priority set for refinement.

CHAPTER 5: ASSESSMENT FRAMEWORK DEMONSTRATION-CASE STUDY II

The case study presented in this chapter is based on a multiphysics CASL challenge problem called CRUD-Induced Power Shift (CIPS).

Chalk River Unidentified Deposits (CRUD) refers to the deposition of porous corrosion products on the surface of the nuclear fuel rods. These chemical products are iron and nickel-based compounds that are produced by corrosion of the metallic surface of the steam generator in PWR. Some of the corrosion products get released into the coolant in particulate form and eventually finds their way to the reactor fuel rods. Deposition of CRUD leads to poor heat transfer, changes in flow pattern and accelerated corrosion. Furthermore, boron compounds get accumulated inside the porous CRUD. CRUD formation is accelerated under sub-cooled boiling condition. As boron is a neutron poison, a shift in the power spectrum is observed. This shift is termed as CRUD-Induced Power Shift (CIPS) [42].

5.1. Objective of the case study

Demonstrate formulation of different elements of the framework for maturity assessment of multiphysics codes and test if the proposed framework can provide a significant improvement in the assessment of the selected codes.

5.2. Demonstration of the framework

All the elements of the framework for this case study are discussed in the following sub-sections.

5.2.1. Preprocessing for the framework development

The preprocessing steps (step 1 and step 2) for the development of the framework for the assessment of CASL code for Multiphysics application are shown in Table 5.1 and Table 5.2. The information presented in this table are based on the CASL report [20]. The PIRT table (step 3) for this challenge problem is prepared by a team of experts in CASL. It is provided in the CASL report [20] with definition of phenomena, knowledge and importance ranking.

Table 5.1: Specify the issue, simulation tool and decision objective (Step 1)

Issue	Chalk River Unidentified Deposits induced power shift (CIPS)
Simulation tool	Multiphysics CASL code – Individual physic codes (Neutronics code, Sub-channel TH code, Fuel modeling code and Coolant Chemistry code) and coupled codes
Decision objective	Assess adequacy of different CASL codes for simulation of CRUD induced power shift

Table 5.2: Specify scenario, system condition, FOM (Step 2) [20]

Specify scenario	Transient and normal operation scenario
System condition	PWR system condition during transient and normal operation (with changing fuel burn up and CRUD deposition)
Quantity of interest	• Boron mass (scalar) • Boron mass distribution (vector) • Axial offset (scalar)

5.2.2. Structural knowledge representation

Figure 5.1 to Figure 5.8 show representation of the phenomenology pyramid for CIPS using GSN. This phenomenology pyramid is based on the CIPS PIRT in the CASL report [20] (prepared by a team of experts in the CASL). Complexity resolution for CIPS challenge problem is based on physics decoupling, so we have added this information in the illustration of the top goal (G1 in Figure 5.2) as an assumption (block A1a in Figure 5.2). The second assumption (block A1b in

Figure 5.2) related to the top goal is that *all relevant phenomena within individual physics were identified by the PIRT process*. This assumption (block A1b) can be graded based on the expert input using the confidence and importance indicator in Figure 3.9. Less confidence in this assumption would undermine the entire process of phenomena resolution (PIRT). It will also impact the formulation of data and model pyramid as they are completely based on the structure of the phenomenology pyramid (PP). Strategy for decomposition is based on the identification of phenomena in all governing physics (FP, CC, TH, Neutronics). As the decomposition is performed in the context of the FOM, this contextual information is stored in block C1. Figure 5.3 shows decomposition of Goal G 1.1 for governing phenomena in FP. Further decomposition is based on phenomena in pellet, gap, and cladding. Solution node at the end of the GSN tree contains evidence in the form of excerpt or link from research papers, CASL reports or other documentations. In this way, supporting evidence for a complete branch can be incorporated in the GSN tree. Documents, excerpts, and links can also be included in other nodes of GSN network in the ASCE.

Decomposition of goal G 1.1.1, *governing phenomena in fuel pellet* is shown in Figure 5.4. As all the phenomena in this sub-tree are less important with respect to the figure of merit (axial offset and boron mass distribution), the importance flag for all the nodes indicates low importance (gray flag in the top right corner of each block).

CIPS PIRT document contains some assumptions related to the phenomena. We designate the assumptions appropriately in the GSN blocks for assumptions (e.g. A 1.4 in Figure 5.8). It can be observed from the GSN tree for the CIPS - phenomenology pyramid, how different GSN blocks offer structure and improve clarity in the representation of phenomena.

We can also better express the connection between different phenomena using GSN. GSN tree helps in expressing connection between different phenomena. It is important to understand that the decomposition is based on expert knowledge. Therefore, it may have large subjective uncertainty. This uncertainty can be addressed by incorporating strong evidence (excerpts from relevant literature, CASL report, or research paper) at the solution nodes. Therefore, it is important to provide sufficient evidence at the solution nodes to support the decomposition of the GSN tree. The structure of the pyramid can be modified with repeated iterations, and additional information can be added as and when required.

The characterization of phenomena (identification of QOI, system condition for individual phenomena) for comprehensive validation assessment is not completed for this challenge problem.

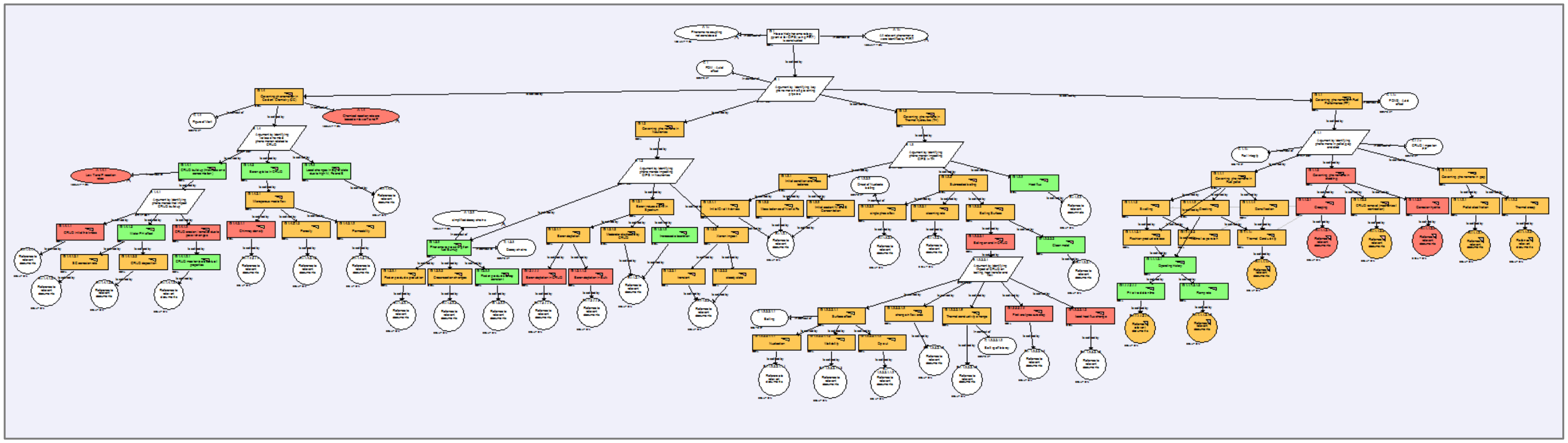


Figure 5.1: GSN tree for CIPS-Phenomenology Pyramid (CIPS-PP)

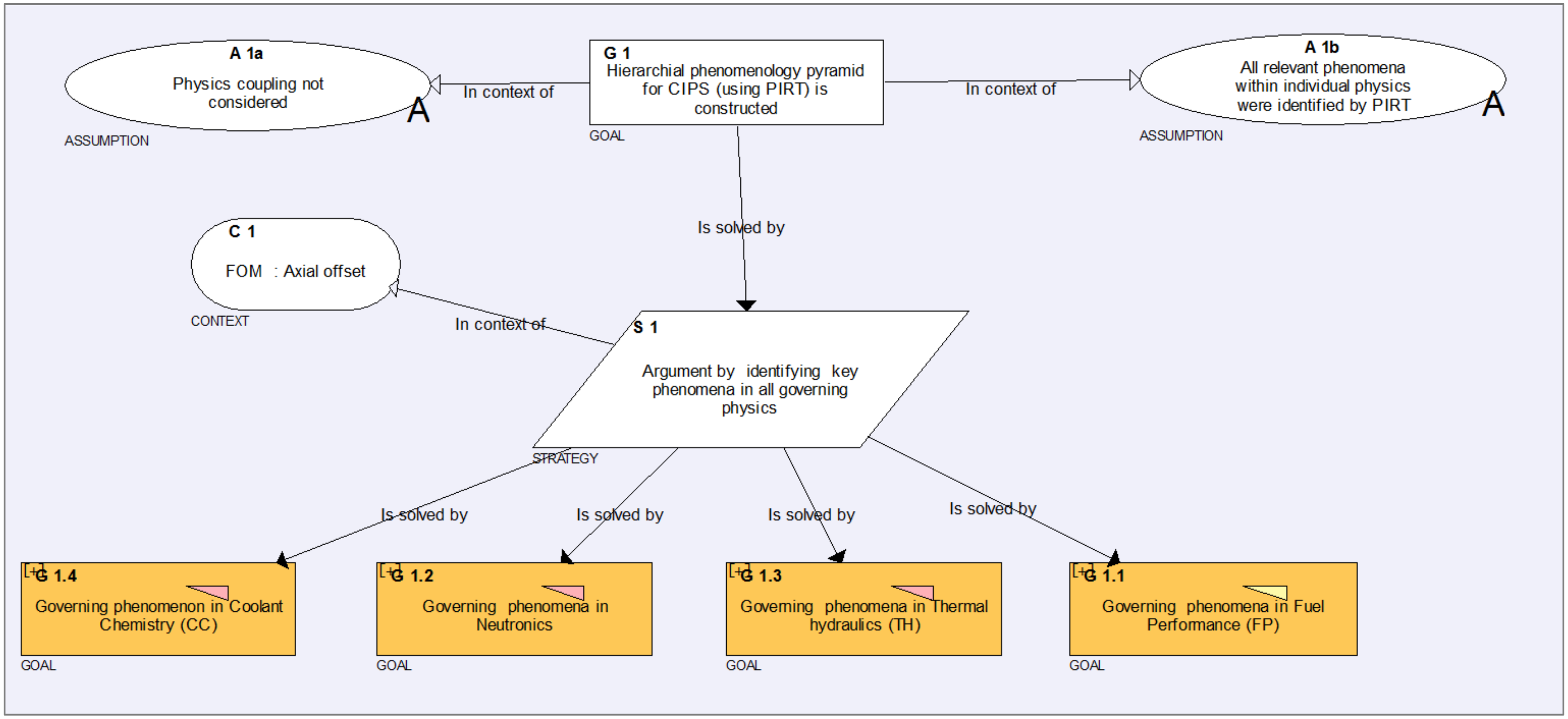


Figure 5.2: GSN tree for CIPS- phenomenology pyramid (showing only Top goal and sub-goals)1

1 All nodes corresponding to each subtree(i.e G 1.1, G 1.2, G 1.3 and G 1.4) has been collapsed for clarity and visibility . “[+]” sign above the goal index indicate that the node has further expansion. ASCE 4.2 facilitate expansion and collapse of any part of the network.

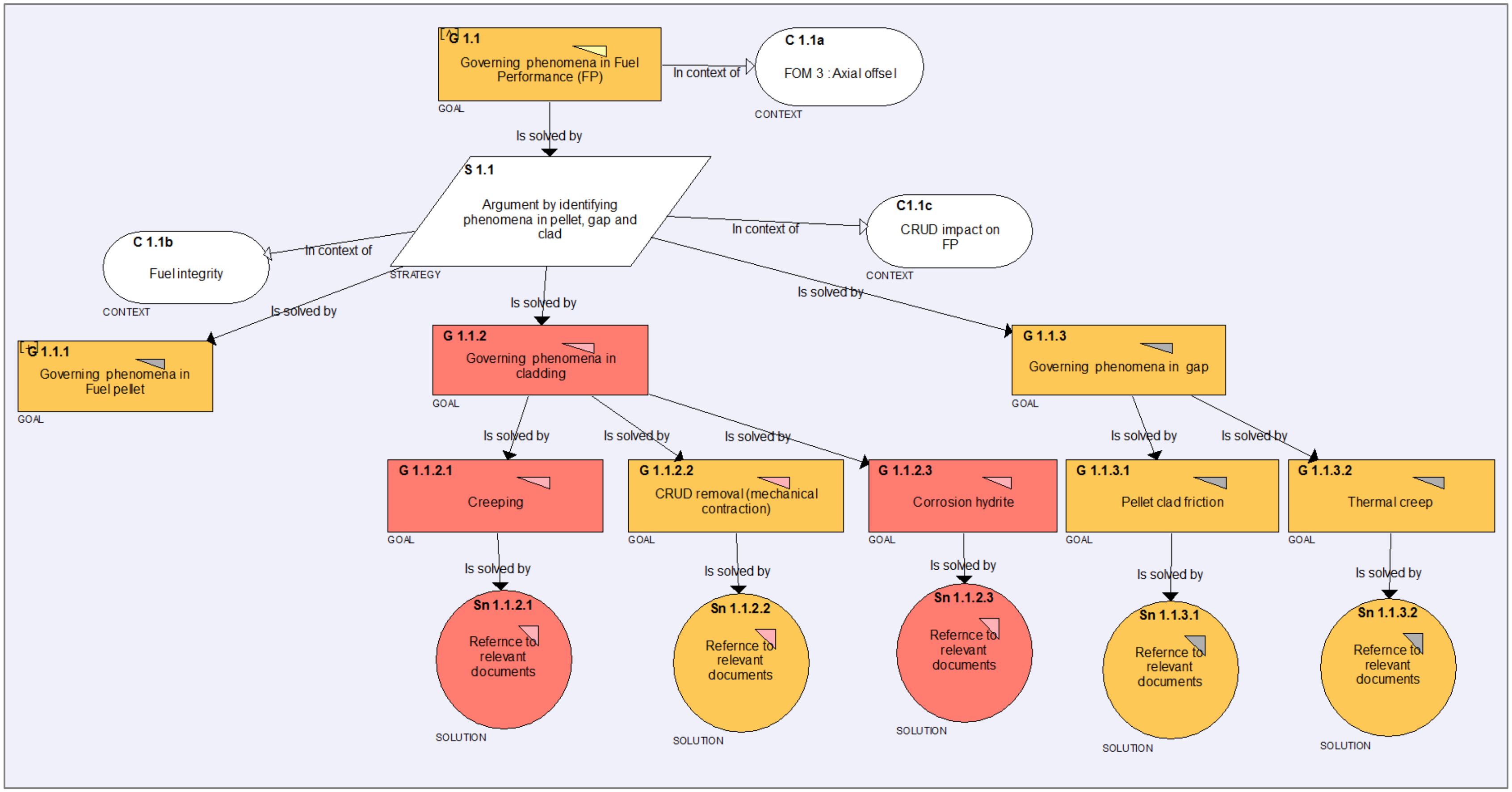


Figure 5.3: Decomposition of sub-goal - G 1.1 (in CIPS- phenomenology pyramid)

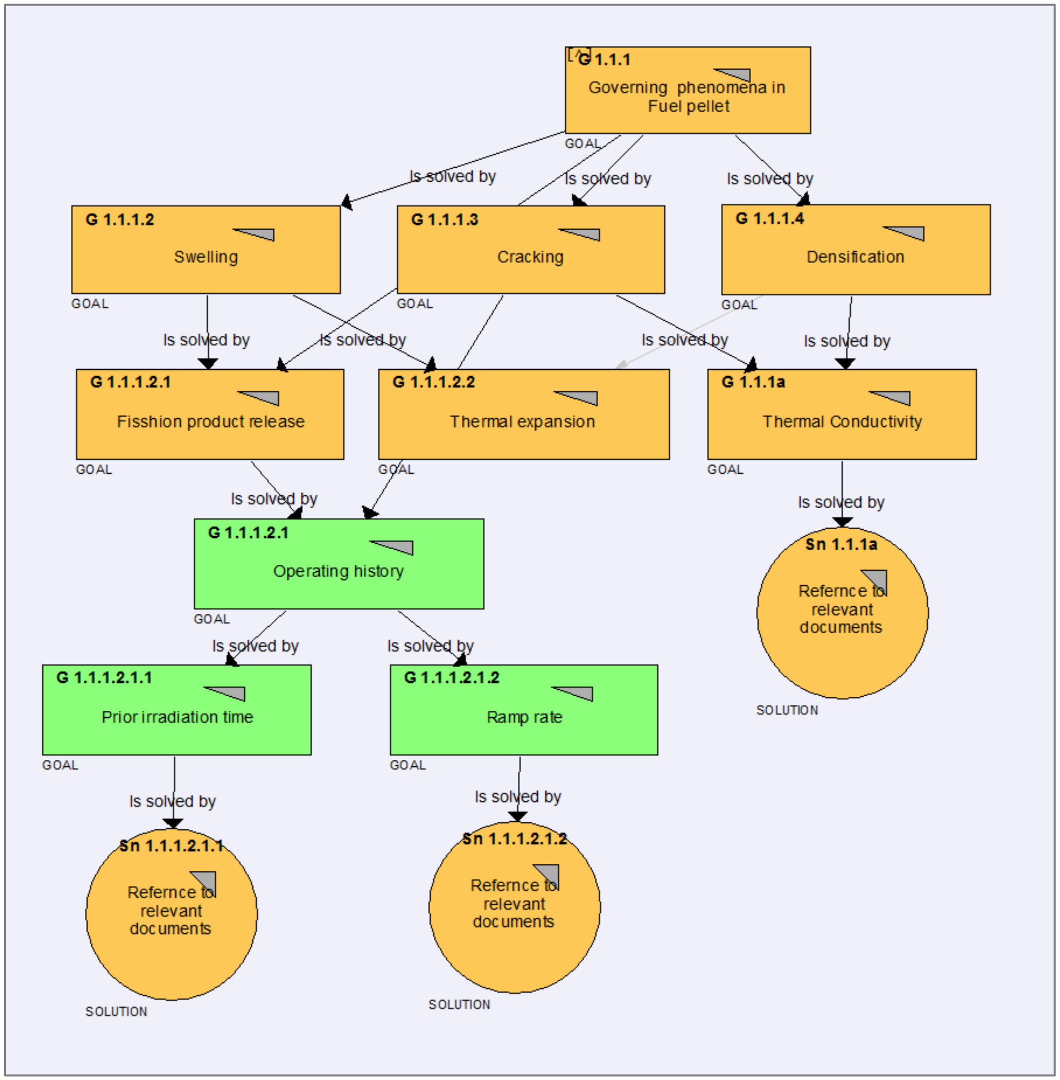


Figure 5.4: Decomposition of sub-goal G 1.1.1 (in CIPS- phenomenology pyramid)

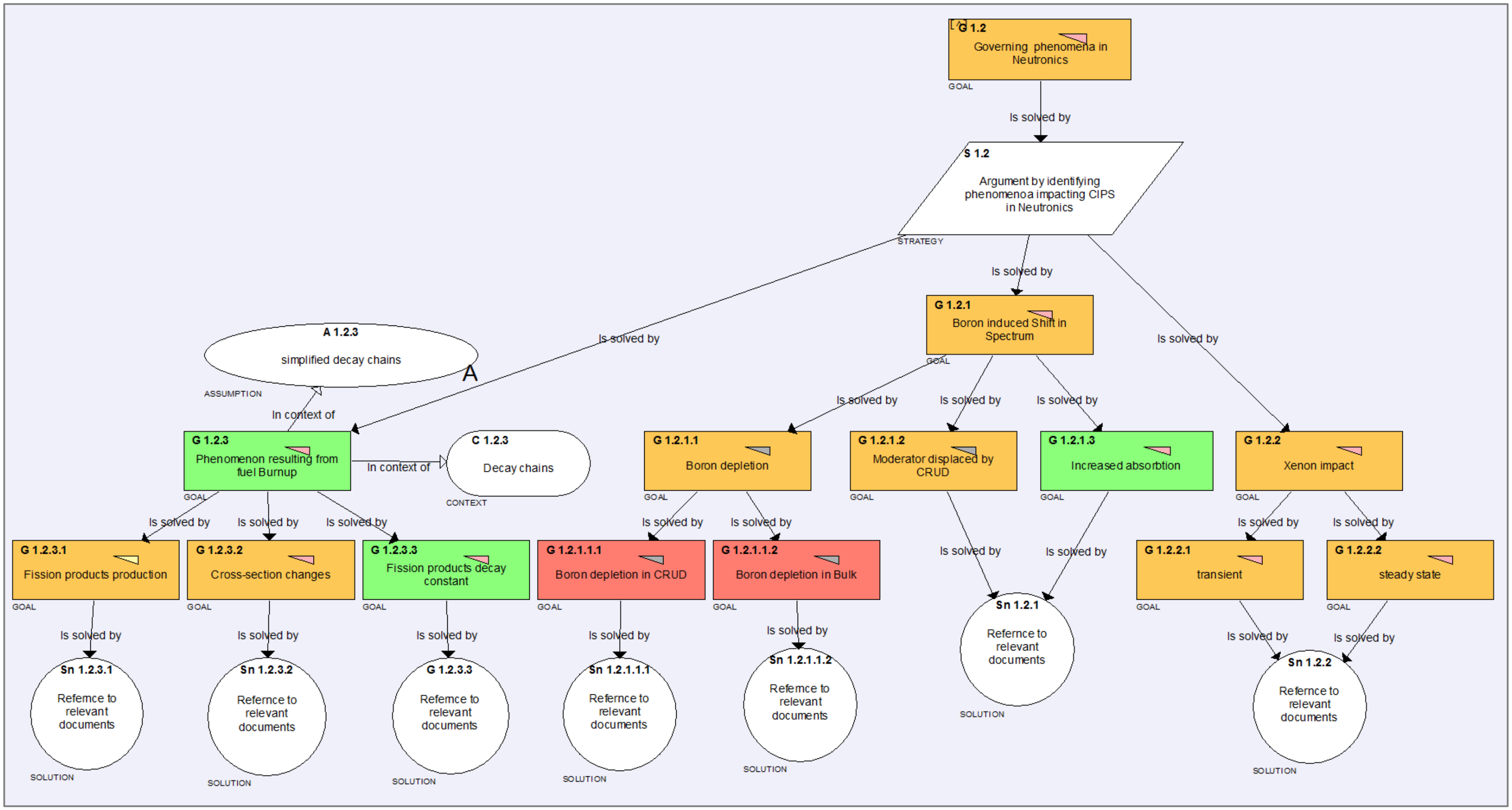


Figure 5.5: Decomposition of sub-goal G 1.2 (in CIPS- phenomenology pyramid)

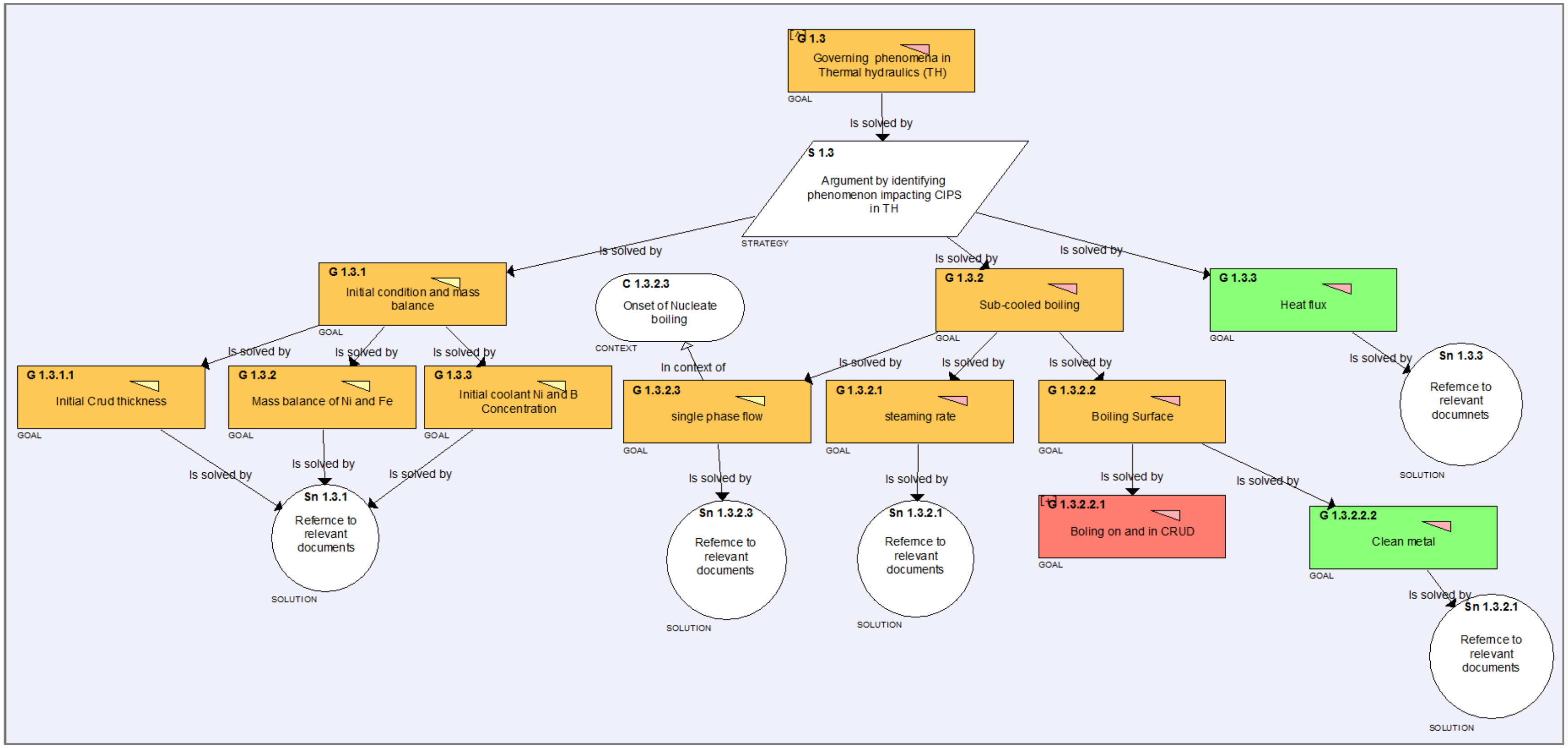


Figure 5.6: Decomposition of sub-goal G 1.3 (in CIPS- phenomenology pyramid)

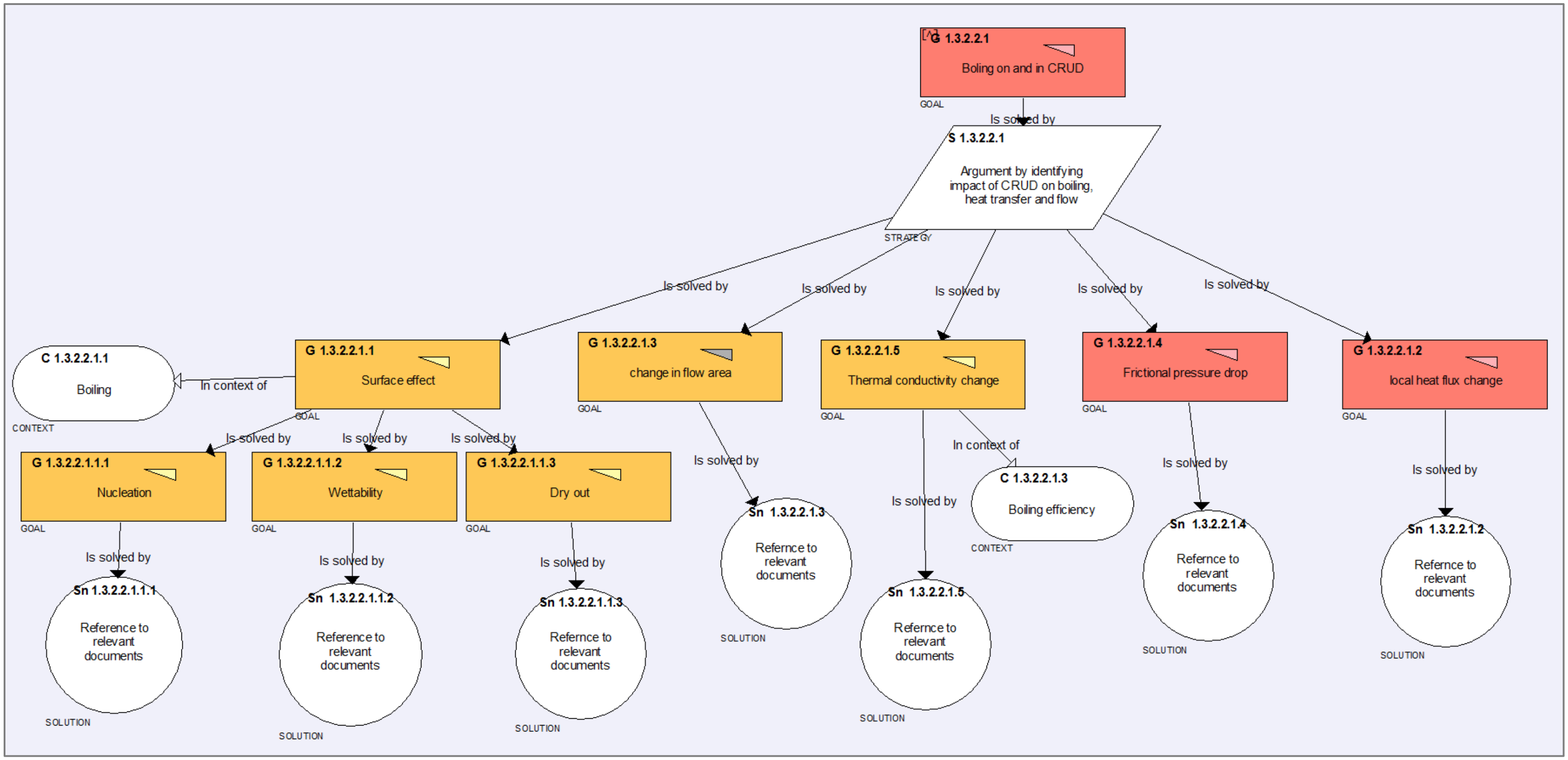


Figure 5.7: Decomposition of sub-goal G 1.3.2.2.1 (in CIPS- phenomenology pyramid)

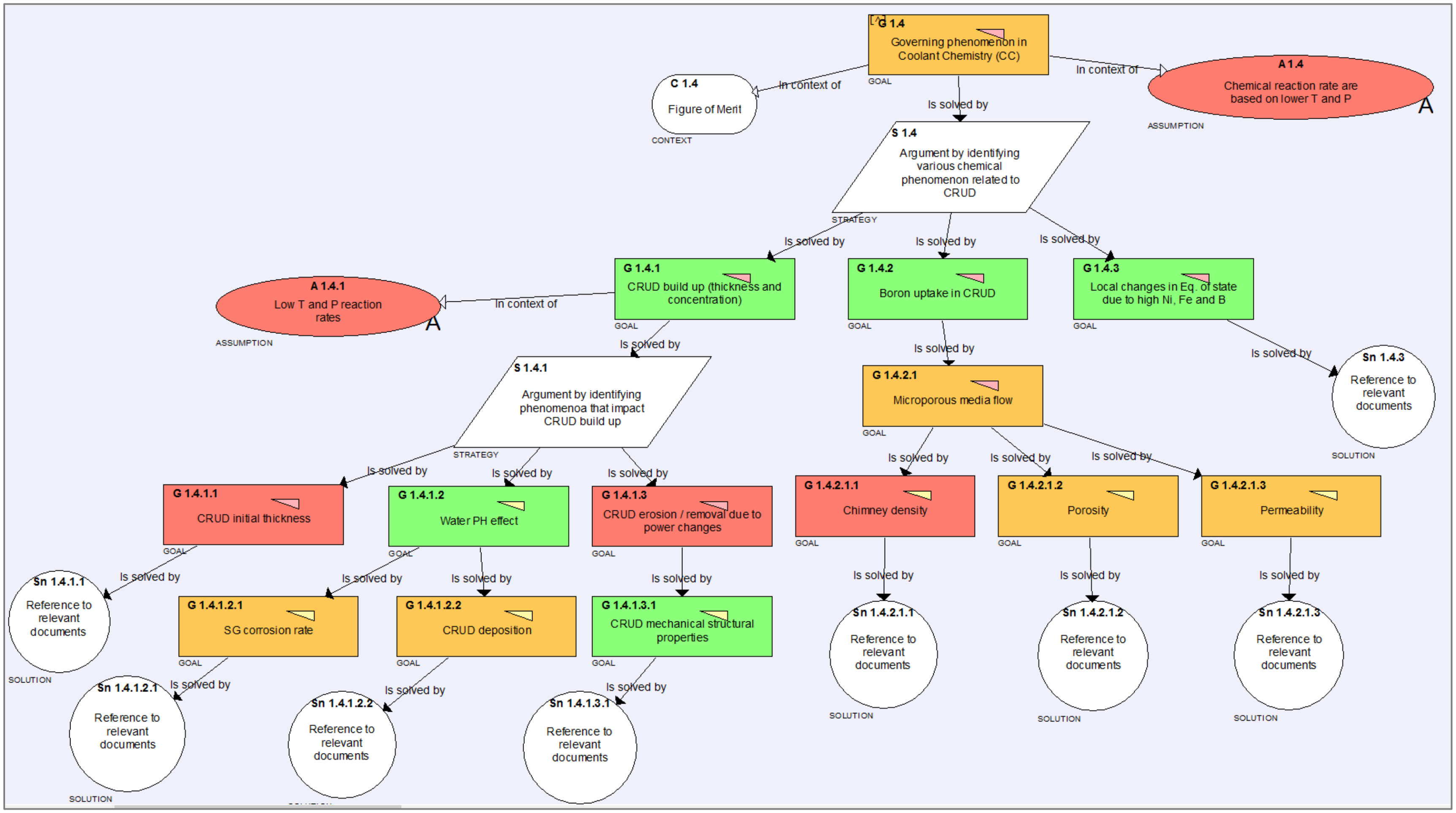


Figure 5.8: Decomposition of sub-goal G 1.4 (in CIPS- phenomenology pyramid)

5.2.3. Classification and characterization of evidence

The classification and characterization of evidence for this challenge problem based on the PCMM attributes is completed in the CASL report [20]. A table of evidence from the CASL report is shown in Table 5.3. The evidence are referred by different abbreviations (MP for neutronics code, VE for coupled code, CT for sub-channel TH code, MA for coolant chemistry code) corresponding to each simulation code. The description of all the evidence is provided in the appendix section of the CASL report [20]. Classification is based on the level of detail of evidence and relevance to PCMM attribute.

Table 5.3: Evidence table in CASL report [20]

PCMM attribute	**Significance/Relevance**			**Gap/ Overall Evaluation**
	H	**M**	**H**	**M**
PMMF: Physics and Material Model Fidelity	MP.2.3.3 MP.2.3.4 VE.1.3.1 VE.1.3.2 VE.1.3.3	MP.3.3.1 MP.3.3.2 MP.3.3.3 MP.3.3.4 MP.3.3.5 MP.3.3.6 MP.3.3.8 MP.3.3.9	MP.2.3.3 MP.2.3.4 VE.1.3.1 VE.1.3.2 VE.1.3.3	MP.3.3.1 MP.3.3.2 MP.3.3.3 MP.3.3.4 MP.3.3.5 MP.3.3.6 MP.3.3.8 MP.3.3.9
SQA: Software Quality Assurance (including documentation)	MP.1.1.3 CT.1.1.1 MA.2.3.1	MP.1.1.2 MP.1.1.4 CT.1.2.1 CT.1.2.2 CT.1.3.1 CT.1.3.2 CT.1.3.5 CT.1.3.6 CT.1.3.7	MP.1.1.3 CT.1.1.1 MA.2.3.1	MP.1.1.2 MP.1.1.4 CT.1.2.1 CT.1.2.2 CT.1.3.1 CT.1.3.2 CT.1.3.5 CT.1.3.6 CT.1.3.7
CVER: Code Verification	MP.1.2.2 MP.2.3.4 CT.1.2.3 MA.2.3.2	MP.1.3.1 MP.1.3.2 CT.1.3.3	MP.1.2.2 MP.2.3.4 CT.1.2.3 MA.2.3.2	MP.1.3.1 MP.1.3.2 CT.1.3.3
SVER: Solution Verification	MP.2.1.1 MP.2.1.4 CT.1.1.4 CT.1.2.4 MA.2.3.3	MP.2.1.2 MP.2.1.3 MP.2.3.3 MP.2.3.4 CT.1.3.4	MP.2.1.1 MP.2.1.4 CT.1.1.4 CT.1.2.4 MA.2.3.3	MP.2.1.2 MP.2.1.3 MP.2.3.3 MP.2.3.4 CT.1.3.4
SVAL: Separate Effects Validation	MP.3.1.1	MP.2.3.1 MP.3.1.3 CT.2.2.1	MP.3.1.1	MP.2.3.1 MP.3.1.3 CT.2.2.1
IVAL: Integral Effects Validation	MP.3.1.1 MA.1.2.2 MA.1.2.3 MA.1.2.4 VE.1.1.2 VE.1.2.1 VE.1.2.2	MP.3.1.2 MP.3.1.3 CT.2.1.2	MP.3.1.1 MA.1.2.2 MA.1.2.3 MA.1.2.4 VE.1.1.2 VE.1.2.1 VE.1.2.2	MP.3.1.2 MP.3.1.3 CT.2.1.2
UQSA: Uncertainty Quantification & Sensitivity Analysis			VE.1.3.5 VE.1.3.6 VE.1.3.7	None [0]

5.2.4. Formulation of decision model

This section of the framework illustrates the formulation of decision model for validation assessment of CTF. The main module of the decision model is shown in Figure 5.9. Goal 1 represents the top claim of the decision model, i.e., "Maturity assessment of CASL codes for CIPS challenge problem is completed." This claim is broken down into two sub-claims (Goal 1.1 and Goal 1.2) based on the assessment of direct maturity evaluation attributes (PCMM attributes) and PQA factors. Assessment of PQA factor is based on PQA of phenomenology pyramid/PIRT and PQA of evidence assessment process (EAP). Corresponding to the maturity assessment of individual codes and coupled simulation code we have five claims (goal blocks in GSN). Away goal 1.1.2 in Figure 5.9 corresponds to maturity assessment of coupled simulation code for CIPS. Away goal 1.1.1.1 to Away goal 1.1.1.4 in Figure 5.9 corresponds to maturity assessment of individual physics codes. All away goals are resolved in individual GSN modules. Figure 5.10 shows the GSN module for maturity assessment of neutronics code. GSN module for PQA of phenomenology pyramid (M 1.6) is shown in Figure 5.11. PQA factors are assessed based on personnel factors related to people qualification, process factor related to definition of phenomena, execution of standard procedure and past use of PIRT/phenomenology pyramid for complexity resolution of Multiphysics problems. These factors are described in detail in section 3.5.3. GSN module for PQA of evidence assessment process (EAP) is shown in Figure 5.12. PQA of evidence assessment process (EAP) is based on the significance level of detail of evidence, credibility of evidence and tools and technique used in the validation and verification of codes. These factors are described in detail in section 3.5.3.

Quantitative maturity assessment is performed by transforming the GSN representation of the decision model to the Bayesian network following the techniques described in section 3.5.4.

Bayesian network corresponding to the GSN representation of decision model in Figure 5.9 to Figure 5.12 are shown in Figure 5.13 to Figure 5.20, respectively. Both target and achieved level are shown for comparison of results.

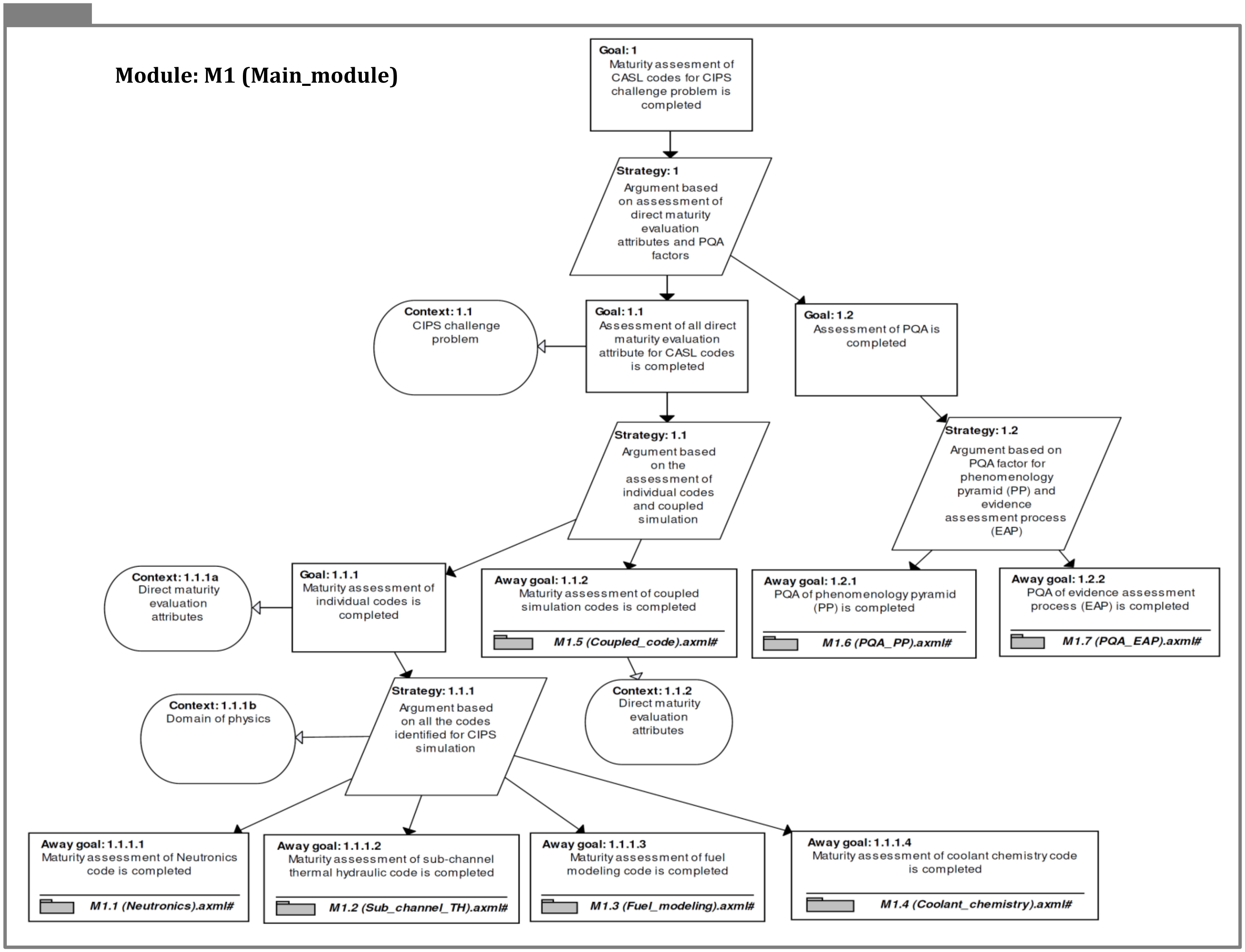


Figure 5.9: Main decision module for maturity assessment of CASL codes for CIPS challenge problem

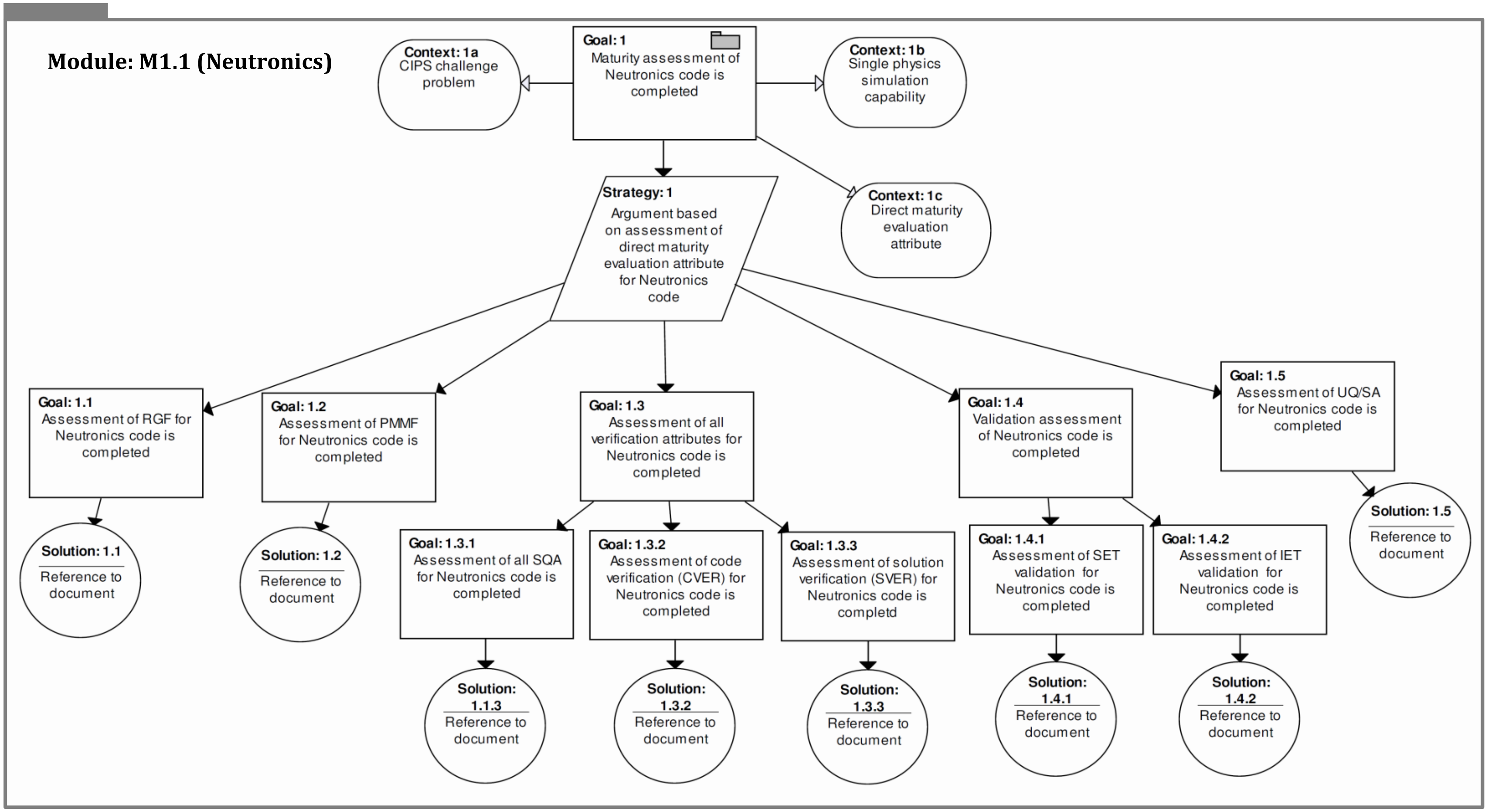


Figure 5.10: GSN module for maturity assessment of Neutronics code, corresponding to Away goal 1.1.1.1 in the main decision module shown in Figure 5.9

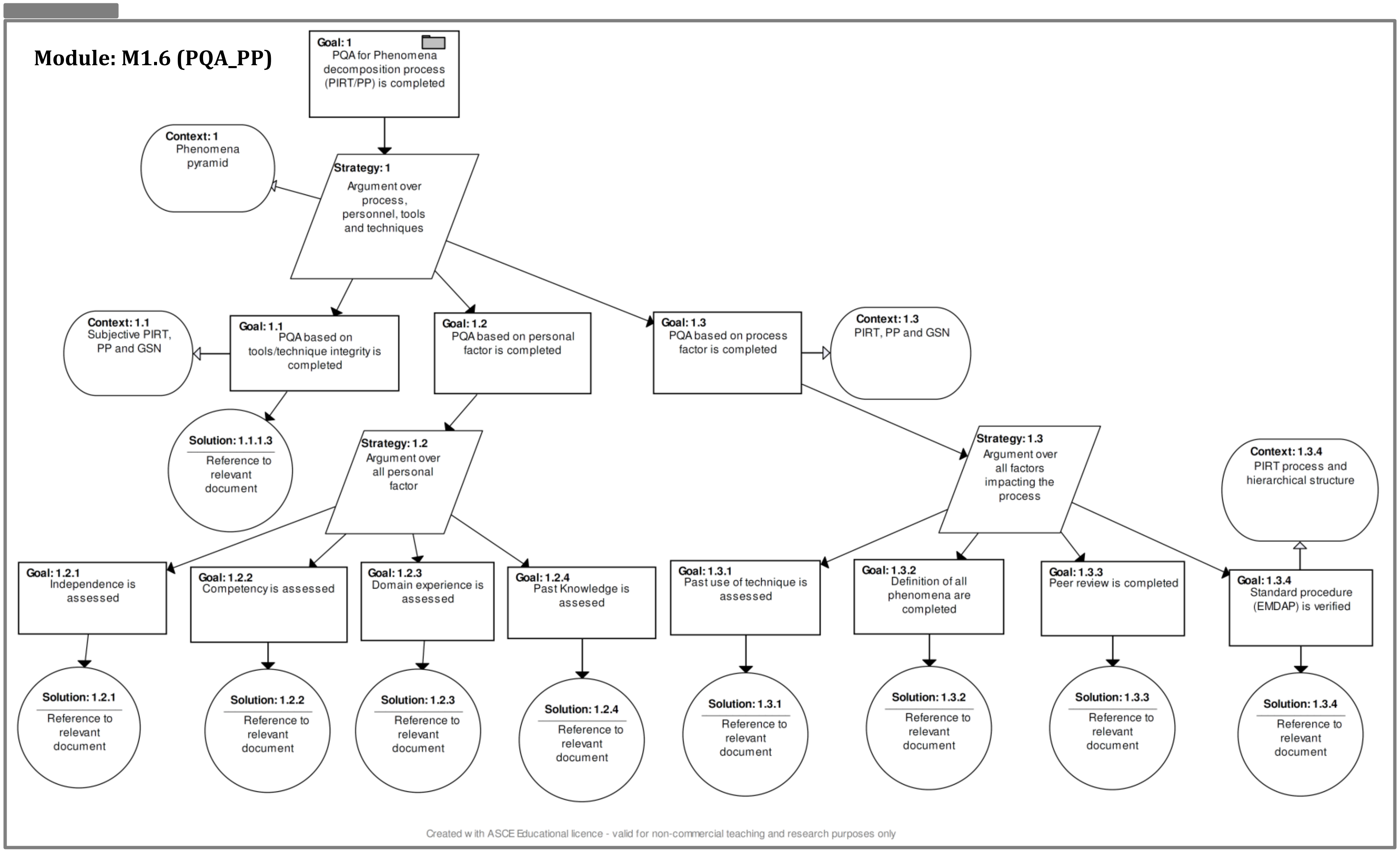


Figure 5.11: Module for PQA of phenomenology pyramid, corresponding to Away goal 1.2.1 in the main decision module shown in Figure 5.9

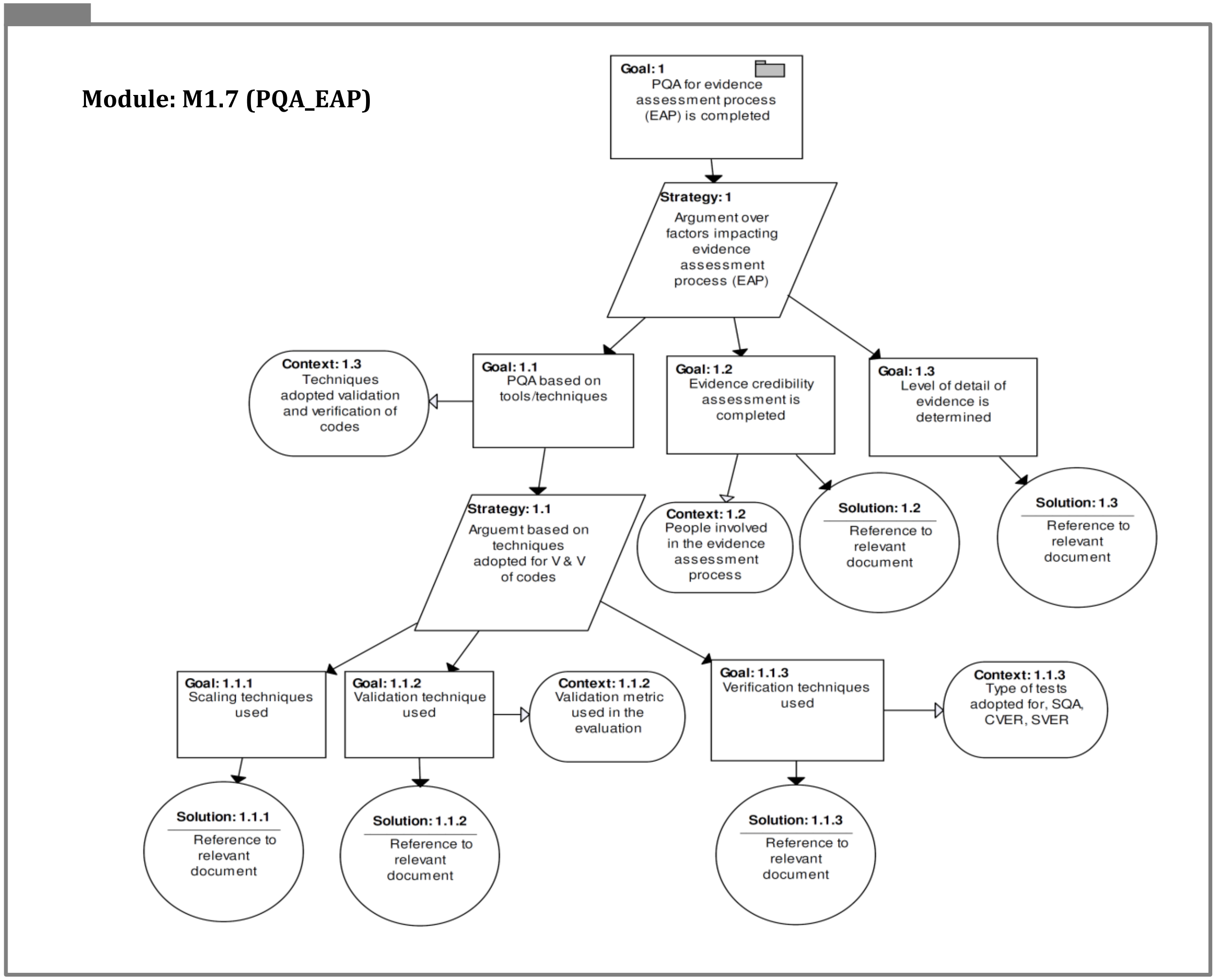


Figure 5.12: Module for PQA of evidence assessment process, corresponding to Away goal 1.2.2 in the main decision module shown in Figure 5.9

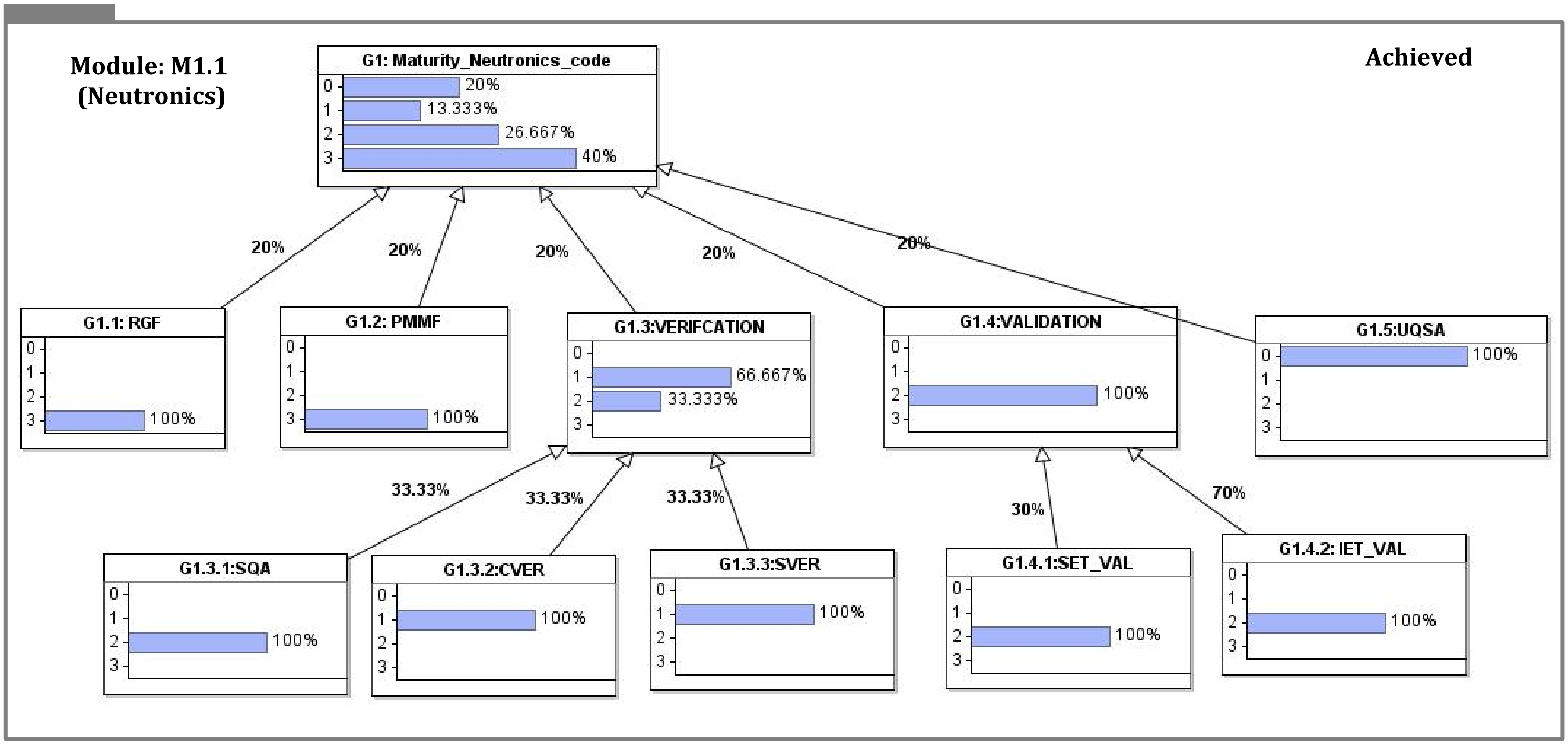


Figure 5.13: Bayesian network corresponding to the GSN module for Neutronic code in Figure 5.10 (Achieved)

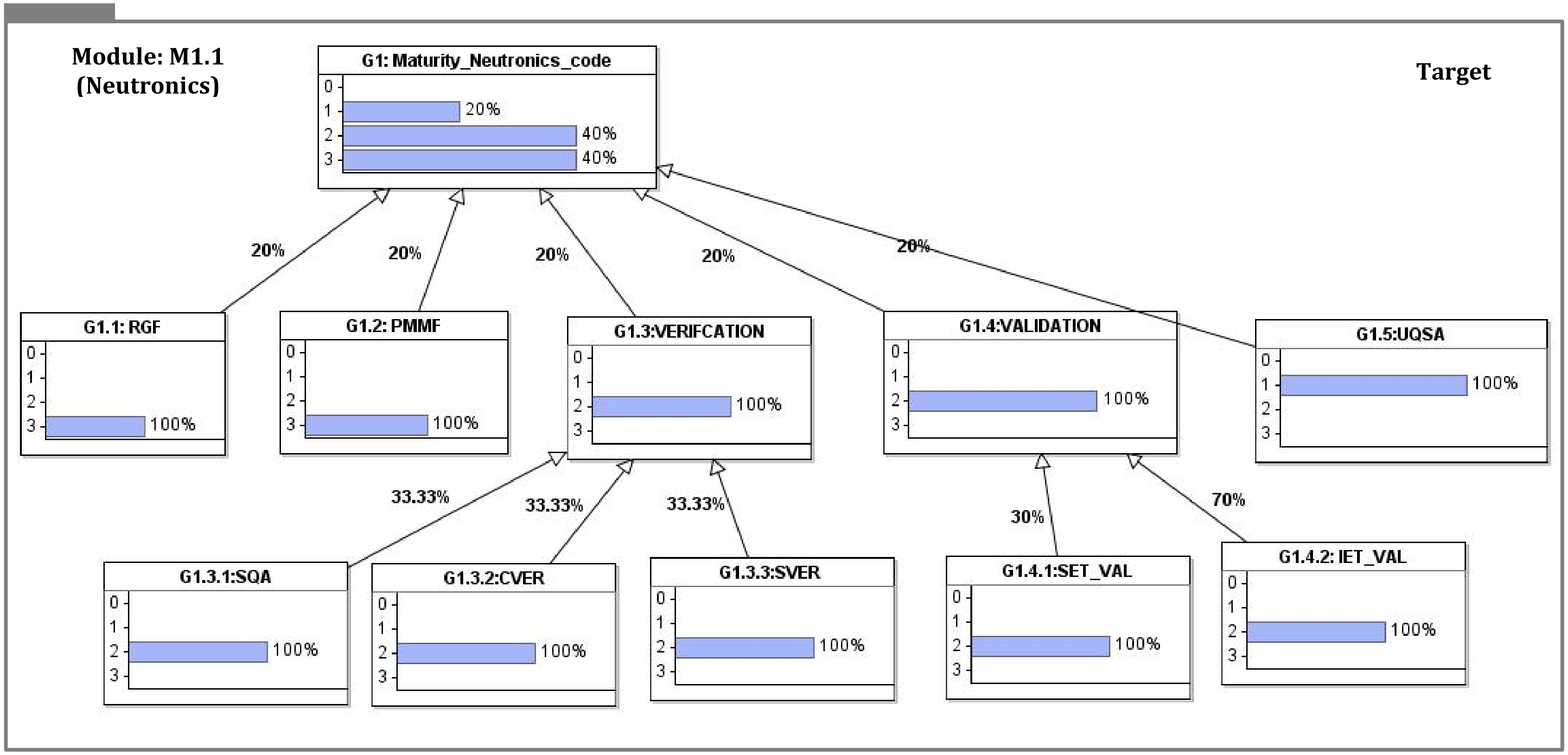


Figure 5.14: Bayesian network corresponding to the GSN module for Neutronic code in Figure 5.10 (Target)

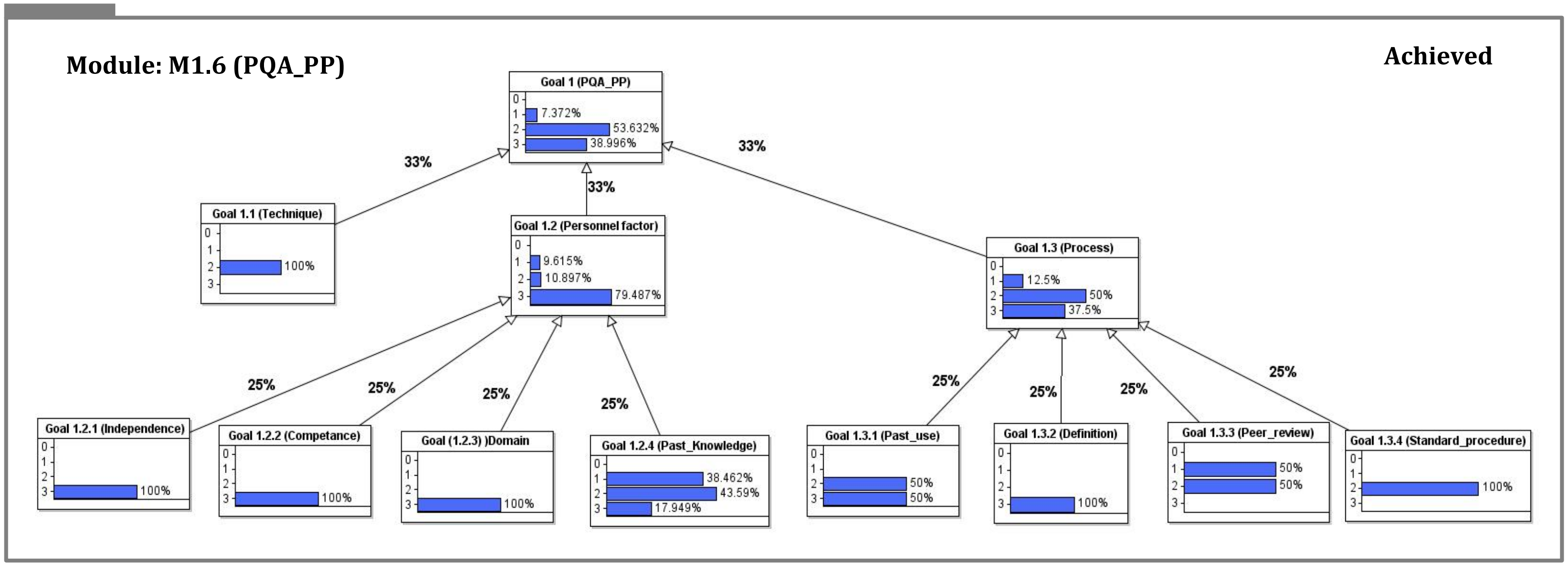


Figure 5.15: Bayesian network corresponding to the GSN module for the PQA of phenomenology pyramid (PP) in Figure 5.11 (Achieved)

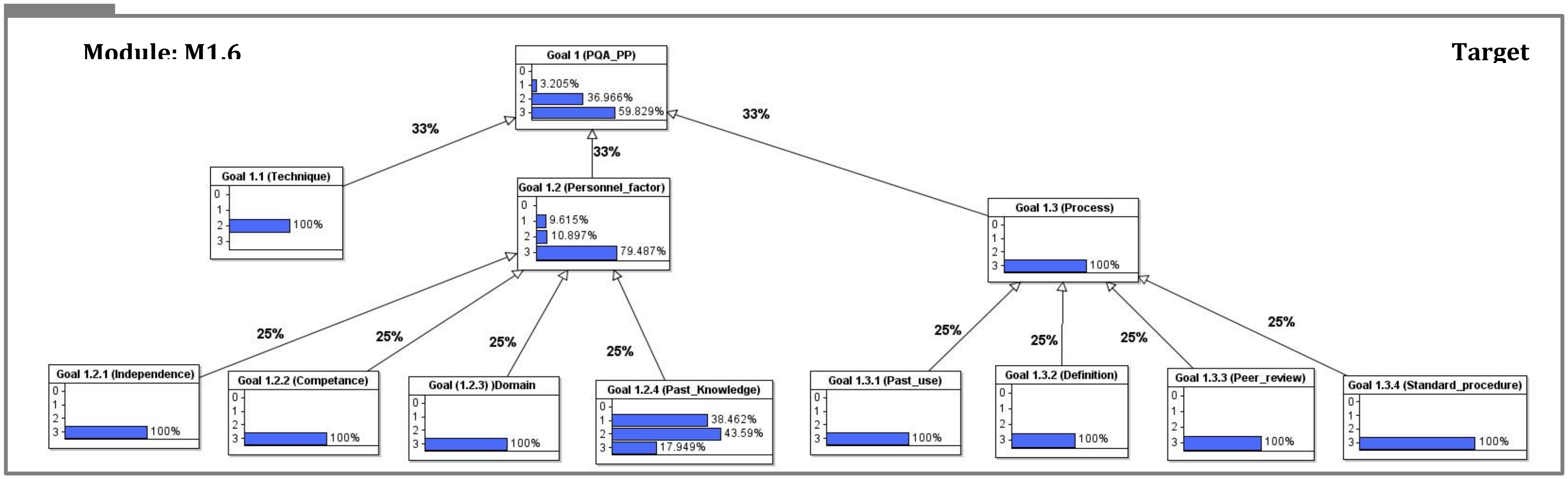


Figure 5.16: Bayesian network corresponding to the GSN module for the PQA of phenomenology pyramid (PP) in Figure 5.11 (Target)

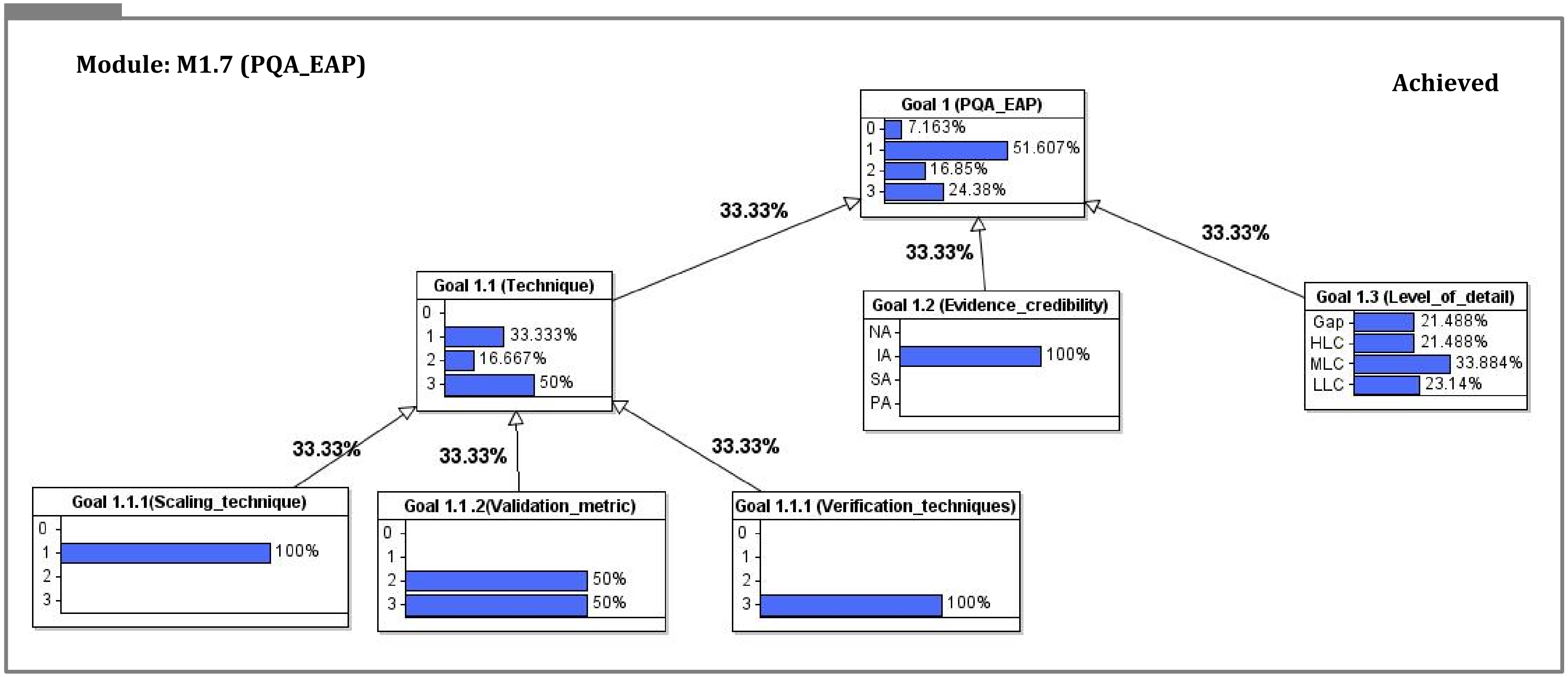


Figure 5.17: Bayesian network corresponding to the GSN module for the PQA of evidence assessment process (EAP) in Figure 5.12 (Achieved)

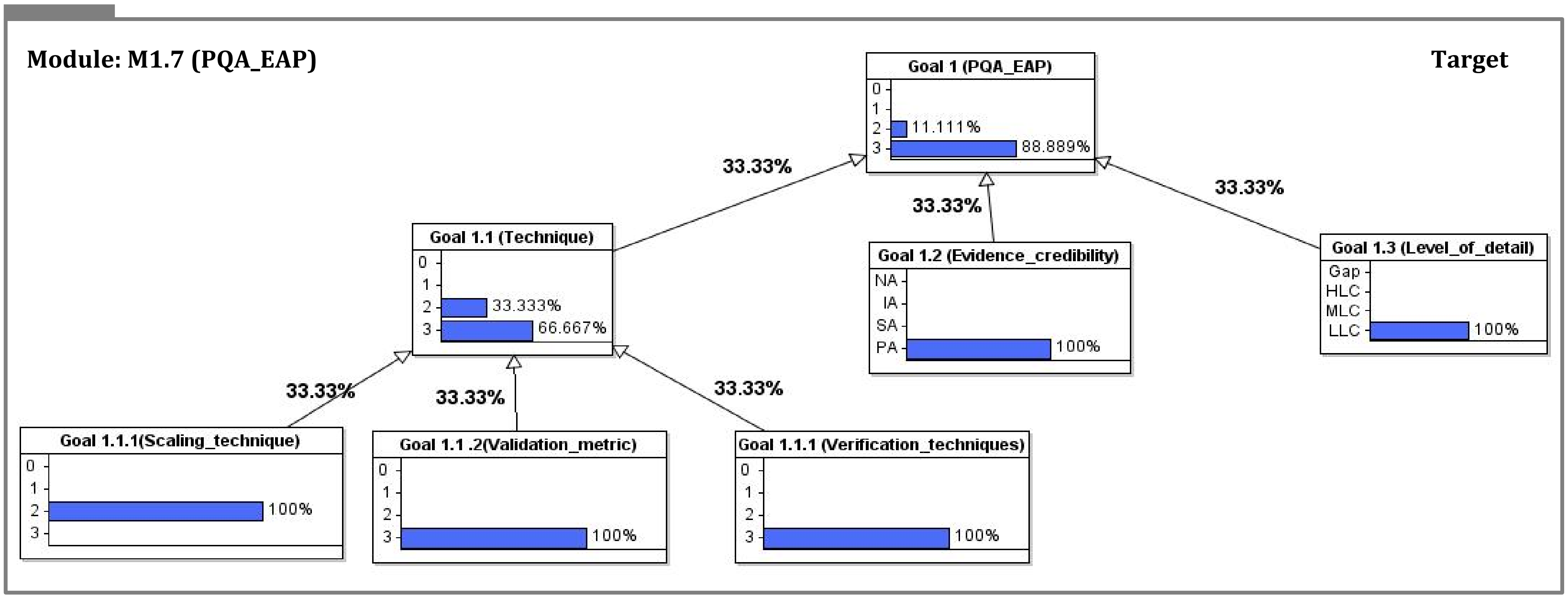


Figure 5.18: Bayesian network corresponding to the GSN module for the PQA of evidence assessment process (EAP) in Figure 5.12 (Target)

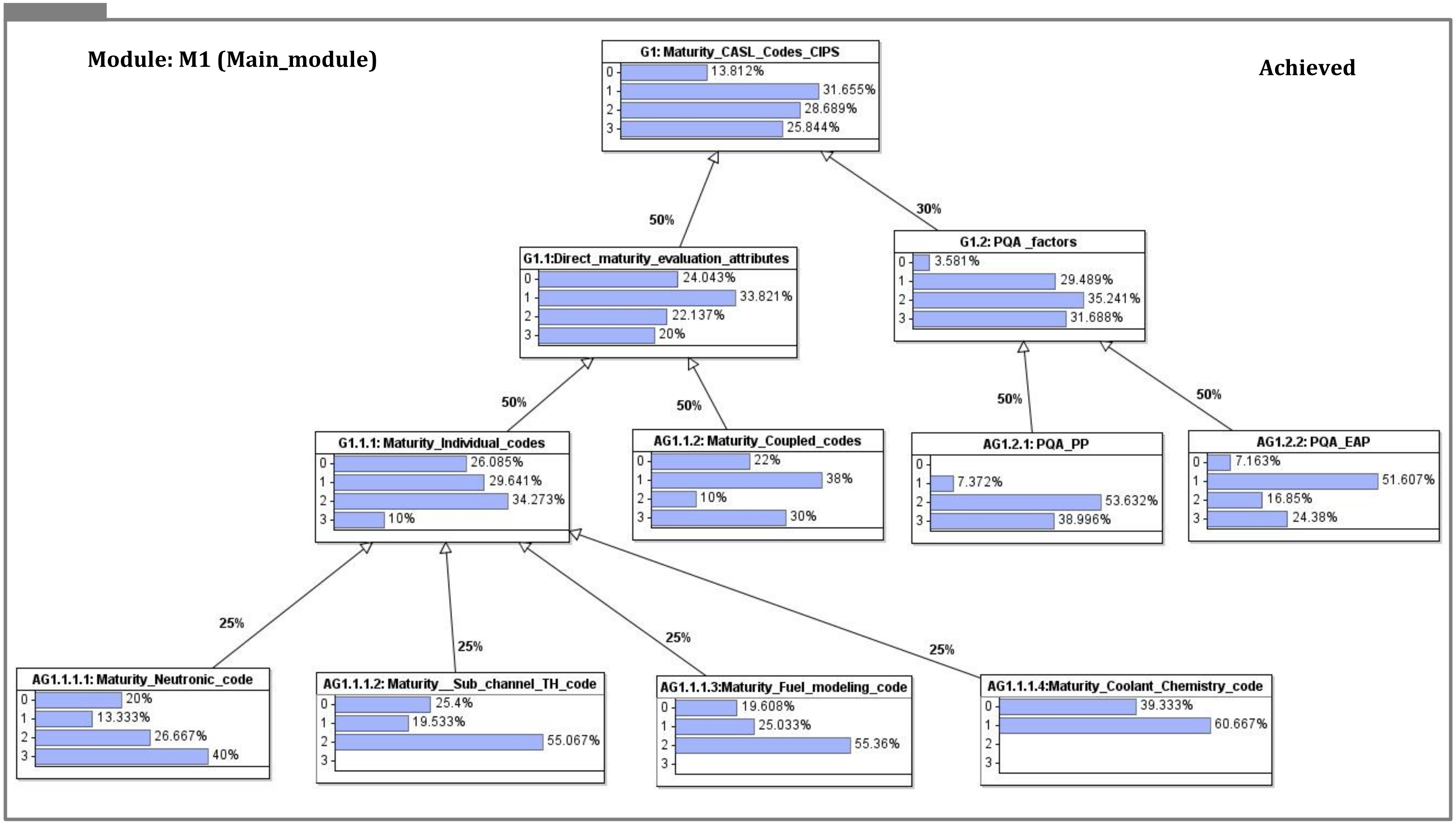


Figure 5.19: Bayesian network corresponding to the main module in Figure 5.9 (Achieved)

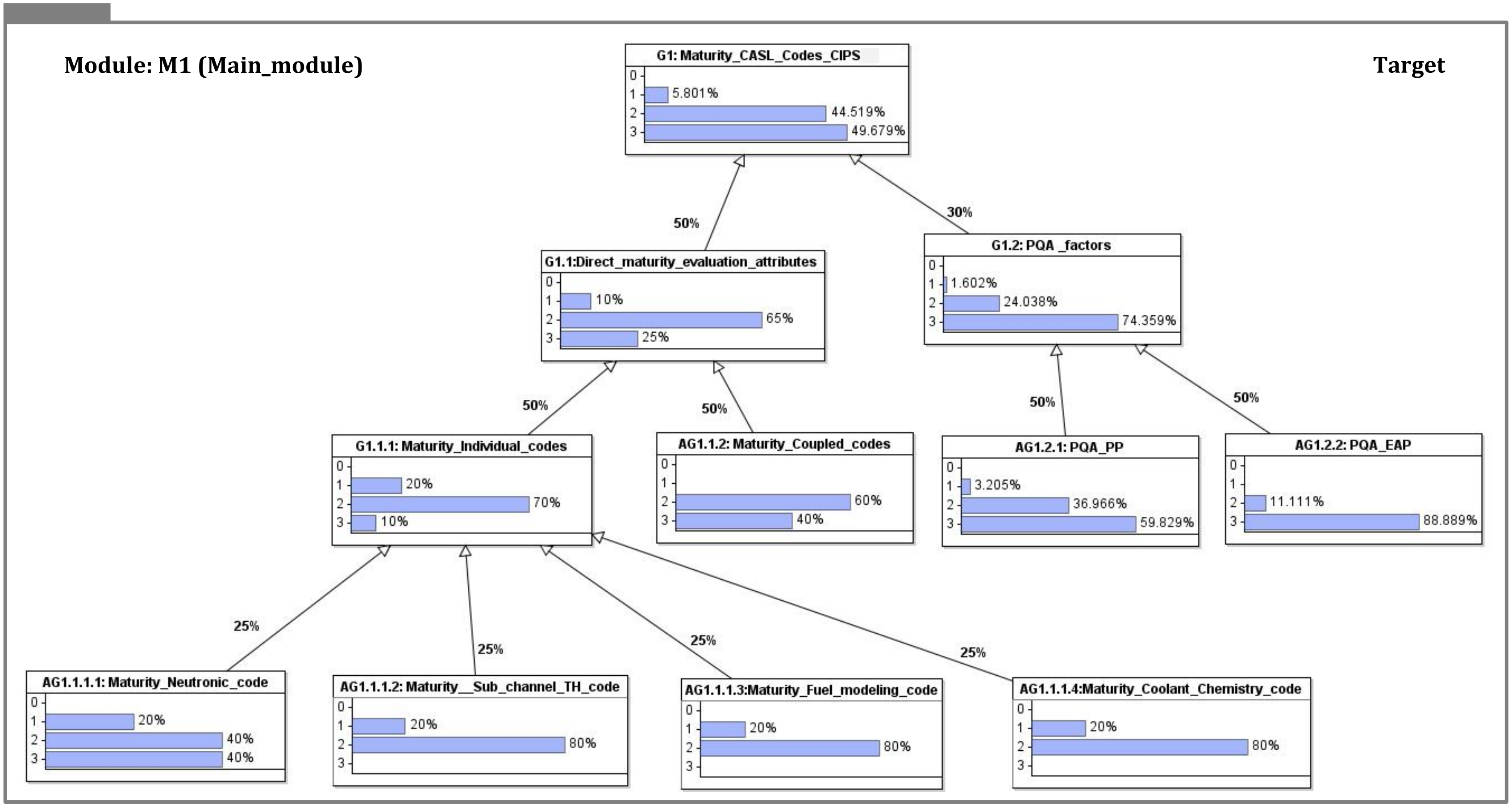


Figure 5.20: Bayesian network corresponding to the main module in Figure 5.9 (Target)

5.2.5. Evaluation and interpretation of result

This section present evaluation and interpretation of result obtained from the formalized decision model developed in the previous section. The evaluation is based on the expected distance metric [Eq. (3.8)] introduced in section 3.5.5. E_N close to zero implies, achieved level is close to the target level. E_N close to 1 implies, achieved level is far from the target level.

Table 5.4: Estimation and interpretation of result for all primary validation attribute

Estimated distance metric		Interpretation/comment
$E_N(Coolant\_chemistry)$	0.66	Need improvement in all PCMM attribute
$E_N(PQA\_EAP)$	0.45	More detailed evidence from the V & V manual of codes are required
$E_N(Coupled\_code)$	0.38	Lack of data for validation, additional test required, verification is incomplete
$E_N(Sub\_channel\_TH)$	0.28	UQ/SA, verification, SET validation is incomplete
$E_N(Fuel\_modeling)$	0.24	Verification, UQ/SA, SET validation is incomplete,
$E_N(Neutronics)$	0.15	Additional verification test required, UQ/SA is incomplete
$E_N(PQA\_PP)$	0.10	Peer review of phenomenology pyramid is required

Table 5.5: Estimation and interpretation of result for Neutronics code

Estimated distance metric		Interpretation/comment	Evidence reference
$E_N(UQSA)$	1	UQ/SA is incomplete	2017 CASL V & V assessment report [20]
$E_N(CVER)$	0.5	Additional code verification tests required to reach the target level	
$E_N(SVER)$	0.5	Additional solution verification tests required to reach the target level	
$E_N(RGF)$	0	Adequate (reached the target maturity)	
$E_N(PMMF)$	0	Adequate (reached the target maturity)	
$E_N(SQA)$	0	Adequate (reached the target maturity)	
$E_N(SET\_VAL)$	0	Adequate (reached the target maturity)	
$E_N(IET\_VAL)$	0	Adequate (reached the target maturity)	

Table 5.6: Estimation and interpretation of result for Sub-channel TH code

Estimated distance metric		Interpretation/comment	Evidence reference
$E_N(UQSA)$	1	UQ/SA is incomplete	2017 CASL V & V assessment report [20]
$E_N(SET\_VAL)$	0.95	SET validation is incomplete due to lack of data	
$E_N(IET\_VAL)$	0.6	Validation of some phenomena are incomplete	
$E_N(CVER)$	0.5	Additional code verification tests required to reach the target level	
$E_N(SVER)$	0.5	Additional solution verification tests required to reach the target level	
$E_N(RGF)$	0	Adequate (reached the target maturity)	
$E_N(PMMF)$	0	Adequate (reached the target maturity)	
$E_N(SQA)$	0	Adequate (reached the target maturity)	

Table 5.7: Estimation and interpretation of result for Fuel modeling code

Estimated distance metric		Interpretation/comment	Evidence reference
$E_N(UQSA)$	1	UQ/SA is incomplete	2017 CASL V & V assessment report [20]
$E_N(CVER)$	0.5	Additional code verification tests required to reach the target level	
$E_N(SVER)$	0.5	Additional solution verification tests required to reach the target level	
$E_N(SET\_VAL)$	0.5	SET validation is incomplete	
$E_N(IET\_VAL)$	0.15	Marginal improvement required	
$E_N(RGF)$	0	Adequate (reached the target maturity)	
$E_N(PMMF)$	0	Adequate (reached the target maturity)	
$E_N(SQA)$	0	Adequate (reached the target maturity)	

Table 5.8: Estimation and interpretation of result for Coolant chemistry code

Estimated distance metric		Interpretation/comment	Evidence reference
$E_N(SQA)$	1	SQA is incomplete	2017 CASL V & V assessment report [20]
$E_N(CVER)$	1	CVER is incomplete	
$E_N(SET\_VAL)$	1	SET validation is incomplete	
$E_N(UQSA)$	1	UQ/SA is incomplete	
$E_N(RGF)$	0.5	Some improvement required	
$E_N(PMMF)$	0.5	Some improvement required	
$E_N(SVER)$	0.5	Additional solution verification tests required to reach the target level	
$E_N(IET\_VAL)$	0.5	Addition IET validation required	

Table 5.9: Estimation and interpretation of result for coupled simulation code

Estimated distance metric		Interpretation/comment	Evidence reference
$E_N(SQA)$	1	SQA is incomplete	2017 CASL V & V assessment report [20]
$E_N(SET\_VAL)$	1	SET validation is incomplete	
$E_N(CVER)$	0.5	Additional code verification tests required to reach the target level	
$E_N(SVER)$	0.5	Additional solution verification tests required to reach the target level	
$E_N(IET\_VAL)$	0.5	Additional test required, lack of data to validate the coupling	
$E_N(UQSA)$	0.5	UQ/SA is incomplete	
$E_N(PMMF)$	0.1	Some improvement required	
$E_N(RGF)$	0.07	Adequate (close to target maturity)	

Table 5.10: Estimation and interpretation of result for process quality assurance (PQA) factors for evidence assessment process (EAP)

Estimated distance metric		Interpretation/comment	Evidence reference
$E_N(Evidence\_credibility)$	0.66	Evidence are based on initial author assessment, needs to be reviewed by subject matter experts	2017 CASL V & V assessment report [20]
$E_N(Scaling\_Technique)$	0.5	Scaling assessment is based on observation, proper scaling analysis is required	NA
$E_N(level\_of\_detail)$	0.47	Need more detailed evidence from the V & V manual of codes to support the assessment results	2017 CASL V & V assessment report [20]
$E_N(Validation\_metric)$	0.17	Need better validation metrics for some tests	NA
$E_N(Verification\_technique)$	0	Adequate (reached the target maturity)	NA

Table 5.11: Estimation and interpretation of result for process quality assurance (PQA) factors for phenomenology pyramid

Estimated distance metric		Interpretation/comment	Evidence reference
$E_N(process)$	0.25	Peer review of pyramid structure and characterization of phenomenology pyramid is not completed	NA
$E_N(Personnel\_factor)$	0	Adequate	NA
$E_N(Technique)$	0	Adequate	NA

5.2.6. Refinement

Based on the estimation and interpretation of results, we can formulate the following list of action items:

(1) Refinement in decision model/framework:

   a. Complete characterization of all the phenomena needs to be completed for rigorous validation assessment of individual codes and coupled simulation.

   b. Structure of phenomenology pyramid should be reviewed by SME.

   c. Weight factors should be reviewed and adjusted.

(2) Refinement of data: It is evident from estimation and interpretation of results in the previous section that the assessment of CASL codes is greatly affected by the lack of data. If the current budget permits, new data acquisition should be conducted.

(3) Refinement items related to model:

   a. SQA and verification tests for codes should be completed.

   b. UQ/SQ for all codes should be completed.

(4) Refinement related to PQA factors:

   a. Need higher level of detail for some evidence. These evidence should be filtered out from the V & V manuals of codes.

   b. Peer review of phenomenology pyramid should be conducted.

The priority set for the action item is provided by the expected distance metric. Items with E_N value close to 1 have higher priority while items with E_N value close to 0 have lower priority.

5.3. Summary remarks

This chapter presents a case study of maturity assessment of CASL codes for CIPS to demonstrate the framework. Based on this case study we can draw the following concluding remarks regarding the proposed framework:

(1) Provides abstraction of information from lower level attributes to higher level attributes in the decision model using the Bayesian network.

(2) Provides a measure of distance between target maturity and achieved level of maturity using expected distance metric. Expected distance measure also helps in deciding priority set (action items) for refinement in the decision model.

(3) PQA helps in monitoring the process quality, efficiency of tools and techniques, and people qualification.

(4) The quality of the maturity framework is governed by the level of detail of the decision schema. In the current case study, the assessment is based on the primary attribute set in PCMM. These attributes are not further divided into sub-attribute. Finer assessment information can be extracted from the maturity assessment framework if the decision schema is expanded to include further lower level attributes. The lower level attributes for validation assessment are described in section 3.5.3. However, as the characterization of phenomena in this case study is not completed, in-depth validation assessment is not obtained. Therefore, characterization of phenomena is included as an action item in the refinement section of the framework.

CHAPTER 6: ANALYSIS OF THE FRAMEWORK

This chapter presents the analysis of the proposed framework based on the sensitivity analysis and different sources of uncertainty in the decision model.

6.1. Sensitivity analysis

Sensitivity analysis helps in determining the effect of the uncertainty in the grades (state or maturity level) of the lower level attribute on the grade of the higher-level attributes. As we use the Bayesian network for quantitative maturity assessment in the framework, we describe the sensitivity analysis using different configurations of the Bayesian network.

Figure 6.1 shows a Bayesian network corresponding to a decision model with four attributes, N1, N2, N3, and N4. Each attribute is graded based on the available evidence. Attributes N1 and N2 have high uncertainty in grades, N3 has low uncertainty and N4 has no uncertainty. Figure 6.2 shows the impact of uncertainty in N1, N2, N3, and N4 on each grade of D. It is evident from this figure that the attribute with uncertain grades leads to higher uncertainty in the decision.

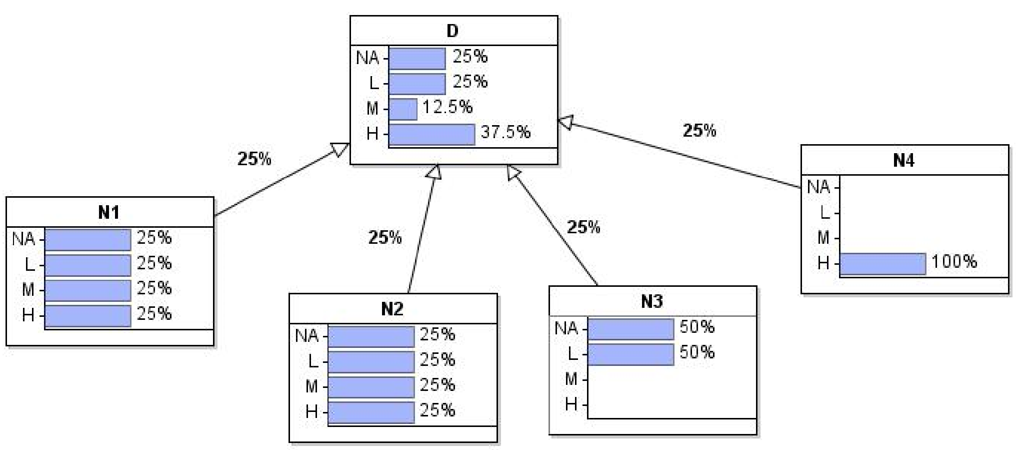


Figure 6.1: Impact of uncertainty in attributes

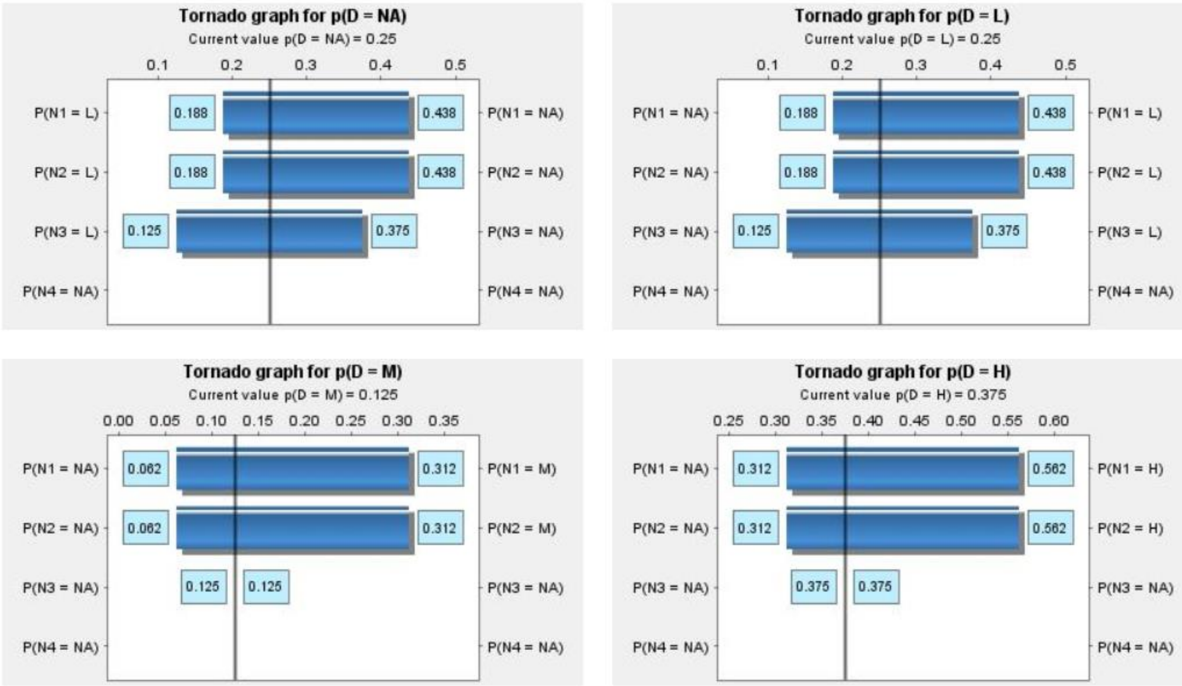


Figure 6.2: Sensitivity analysis for the Bayesian network in Figure 6.1

Figure 6.3 shows another Bayesian network corresponding to a hierarchical decision model to show how the weight factors used in the computation of the conditional probability table impact the sensitivity of the decision model. In this example, attribute D1 is weighted more compared to attribute D2 at the second level in the hierarchy (3:1). Sub-attributes N1 and N2 are weighted equally while N3 and N4 are weighted with a ratio of 1:3. Figure 6.4 shows the tornado plot for E. It is evident from this figure that the sensitivity in the maturity grades for E is affected by the weight factors of the attributes (or nodes). Furthermore, the weight factors of the higher-level nodes dominate in the sensitivity analysis of E.

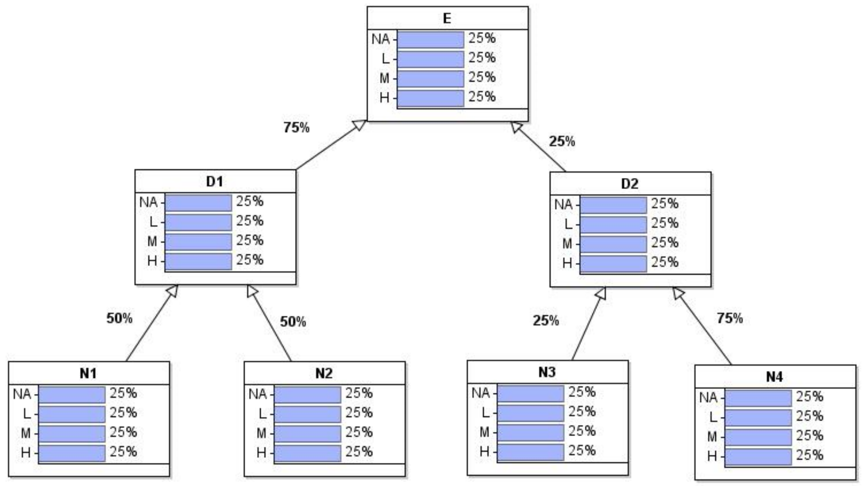


Figure 6.3: Bayesian network with different weight factors

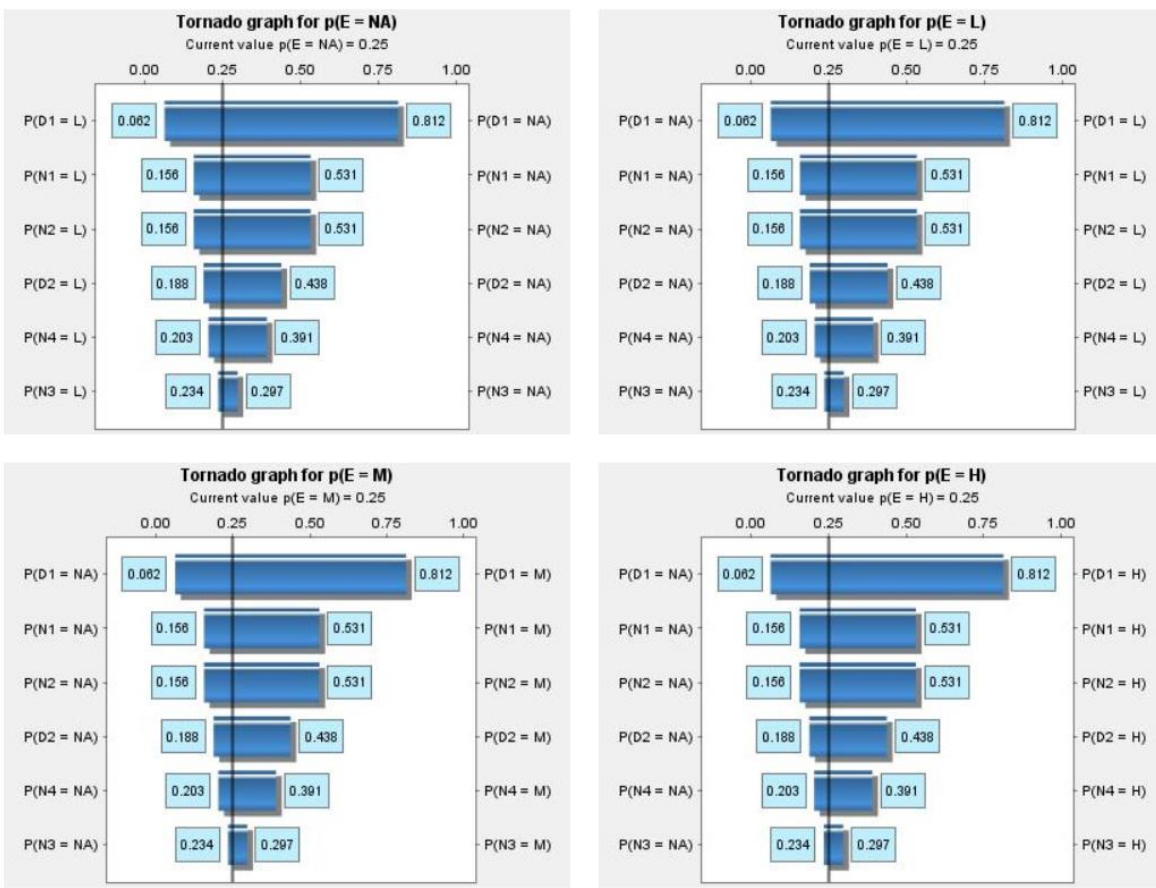


Figure 6.4: Sensitivity analysis for the Bayesian network in Figure 6.3

Figure 6.5 shows the BN with no uncertainty in grades of N1 and N2. Tornado plot corresponding to Figure 6.5 is shown in Figure 6.6. In this case, E is sensitive to D2, N3, and N4, only.

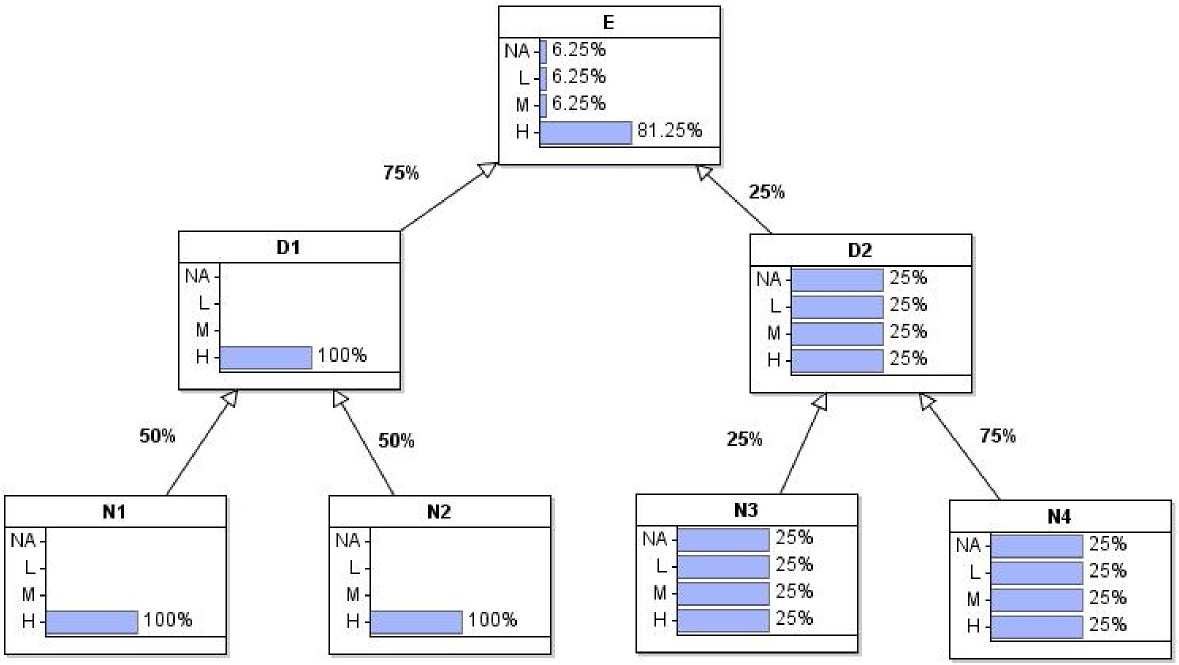


Figure 6.5: Bayesian network with different weight factor and no uncertainty in attribute N1 and N2

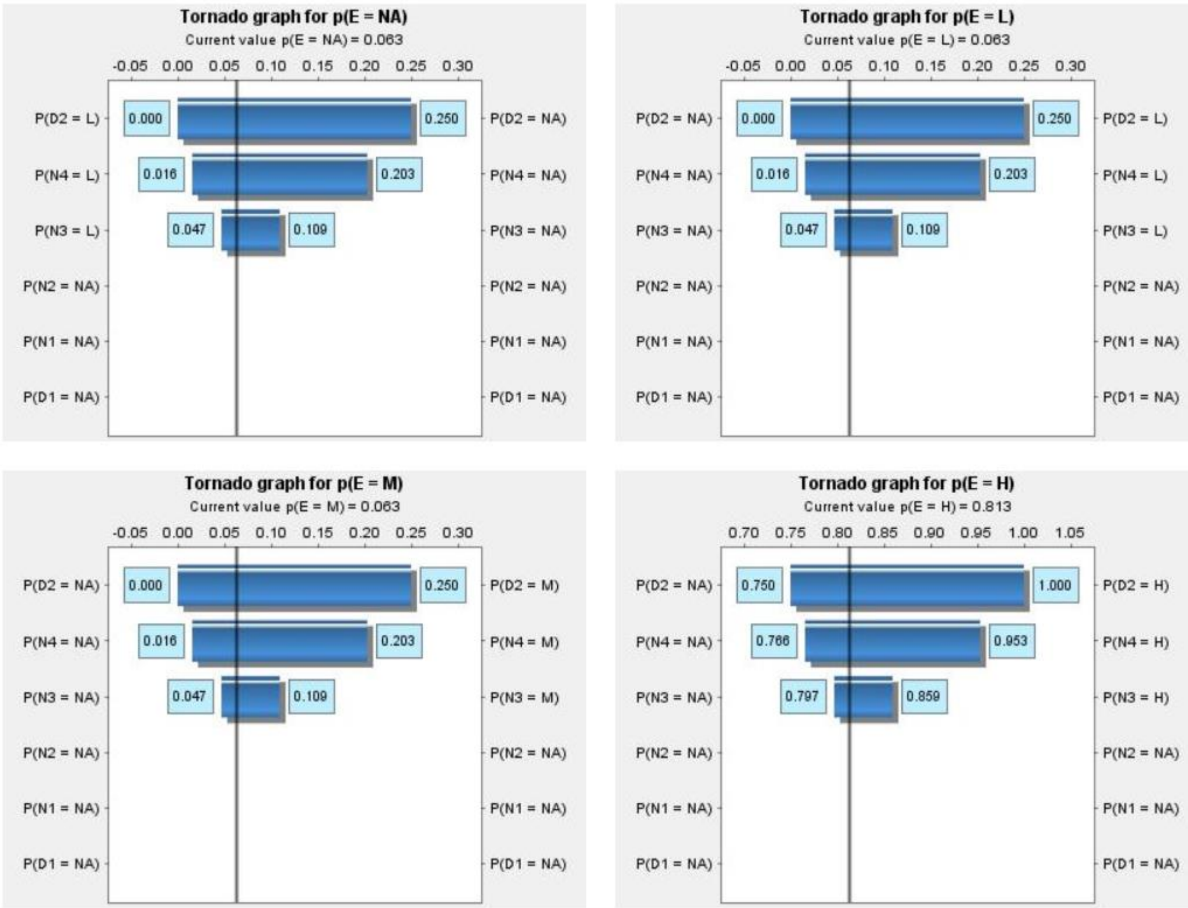


Figure 6.6: Sensitivity analysis for the Bayesian network in Figure 6.5

Figure 6.7 shows the result of sensitivity analysis for validation assessment of CTF in Figure 4.23. It is evident for this figure that the decision regarding the validation assessment of CTF is most sensitive to the validation test results (VTR in Figure 4.23).

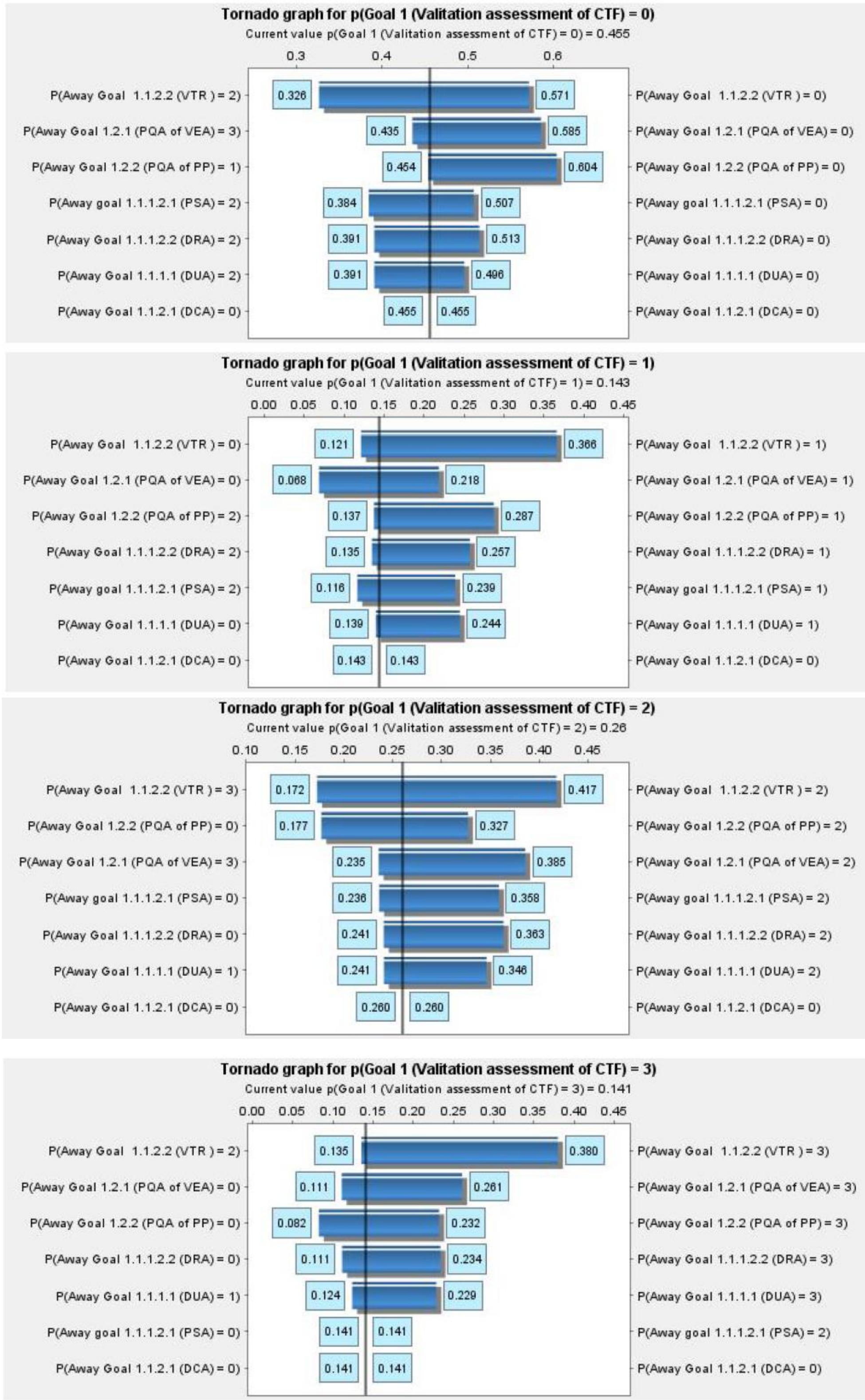


Figure 6.7: Sensitivity analysis of the decision module in Figure 4.23

Figure 6.8 to Figure 6.11 show the tornado plot based on the sensitivity analysis of different grades (grades 0-3) used for assessment of Multiphysics CASL codes for CIPS challenge problem in Figure 5.19. It is evident from Figure 6.11 that the maturity grade '3' is most affected by the maturity of coupled code (Away goal 1.1.2 in Figure 5.19) and PQA of evidence assessment process (Away goal: 1.2.1 in Figure 5.19)

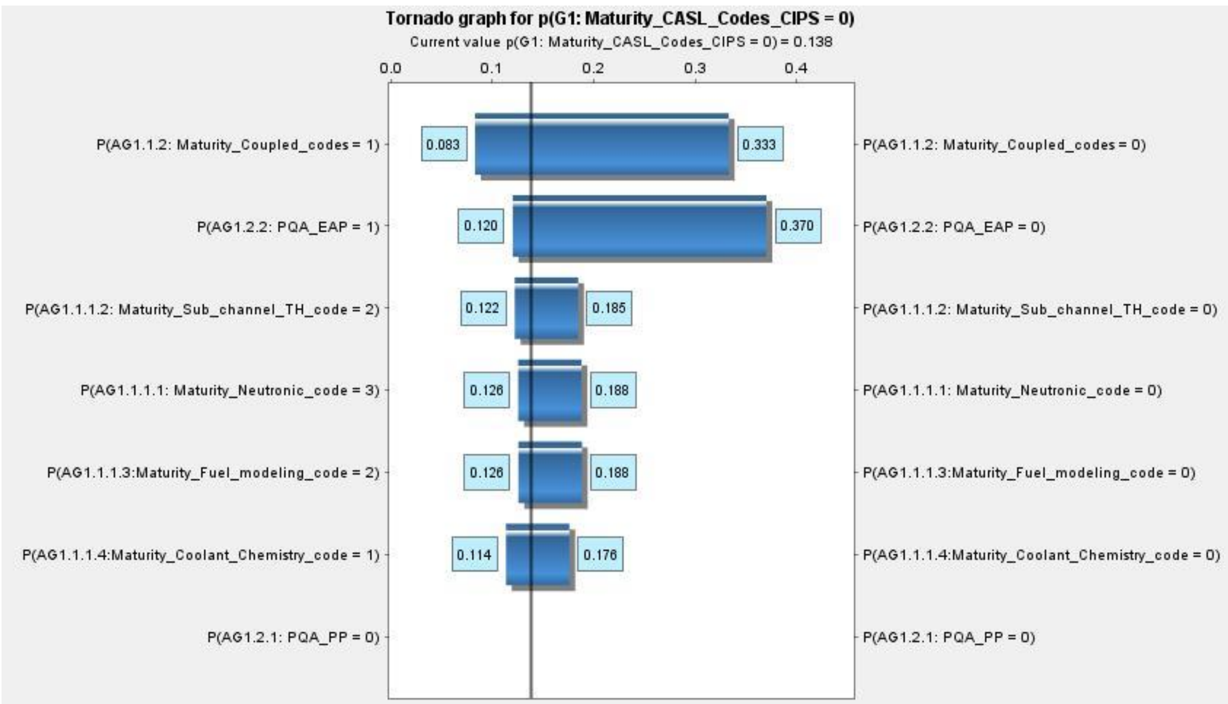


Figure 6.8: Tornado graph for grade '0' for maturity of CASL codes (G1) in Figure 5.19

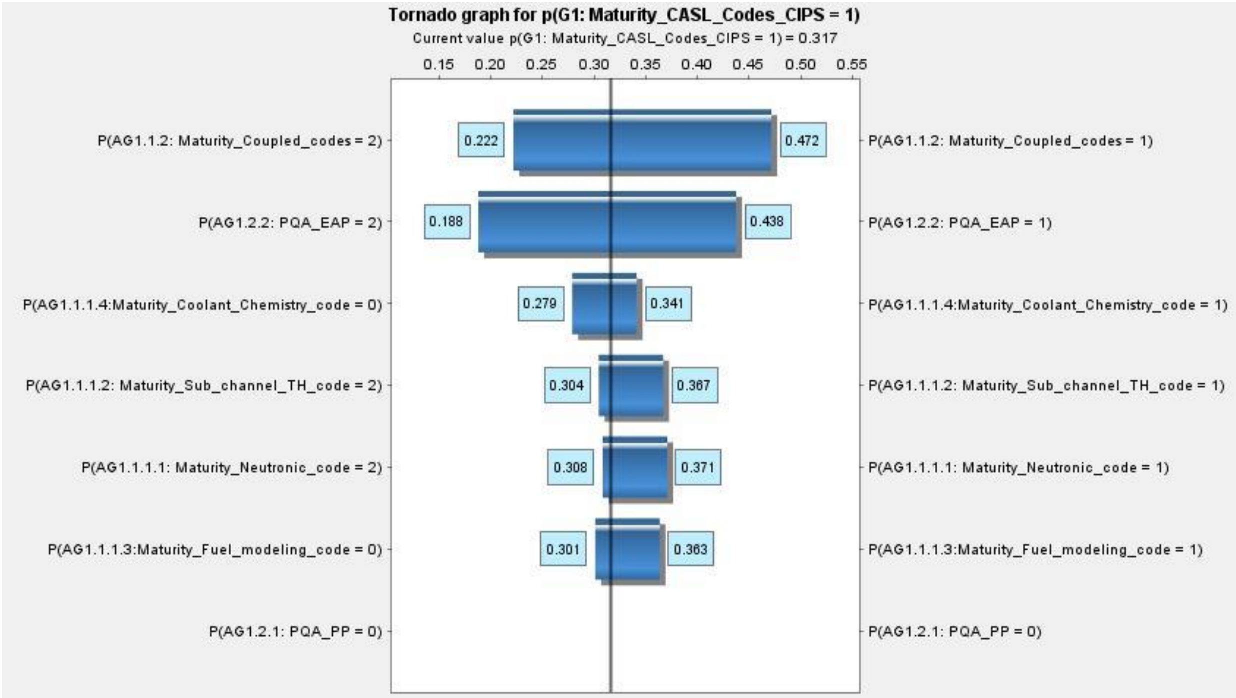


Figure 6.9: Tornado graph for grade ‘1’ for maturity of CASL codes (G1) in Figure 5.19

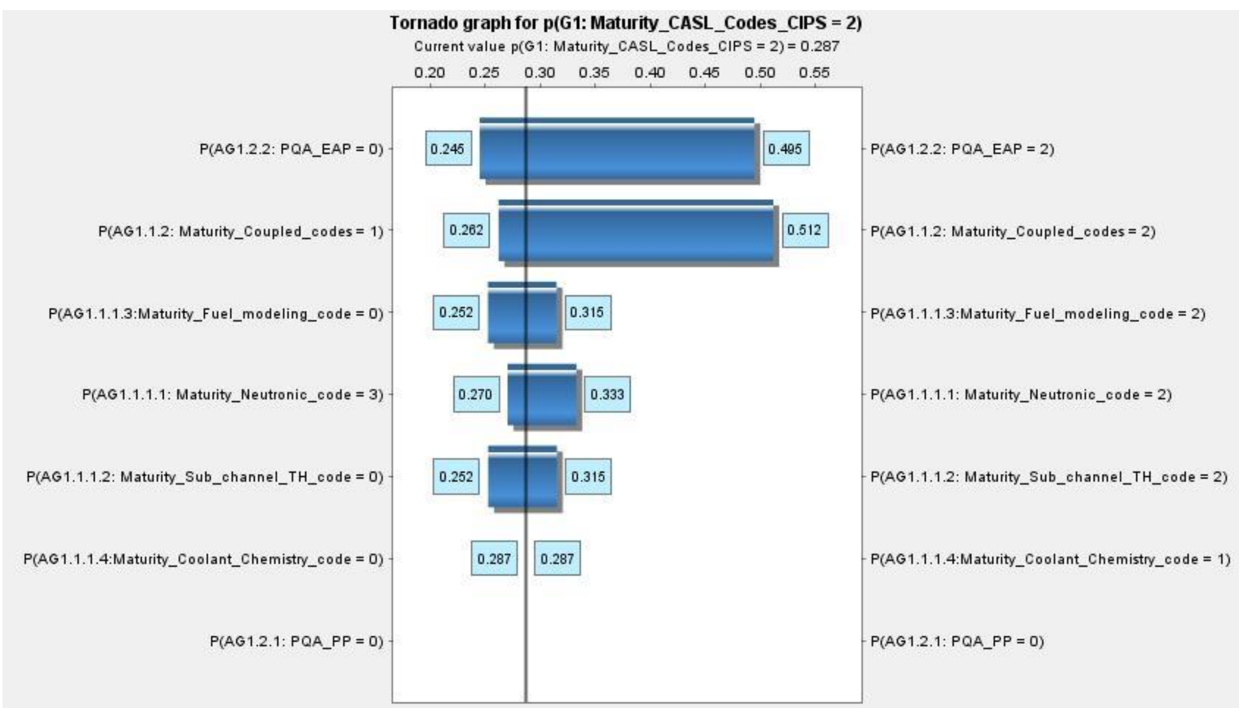


Figure 6.10: Tornado graph for grade ‘2’ for maturity of CASL codes (G1) in Figure 5.19

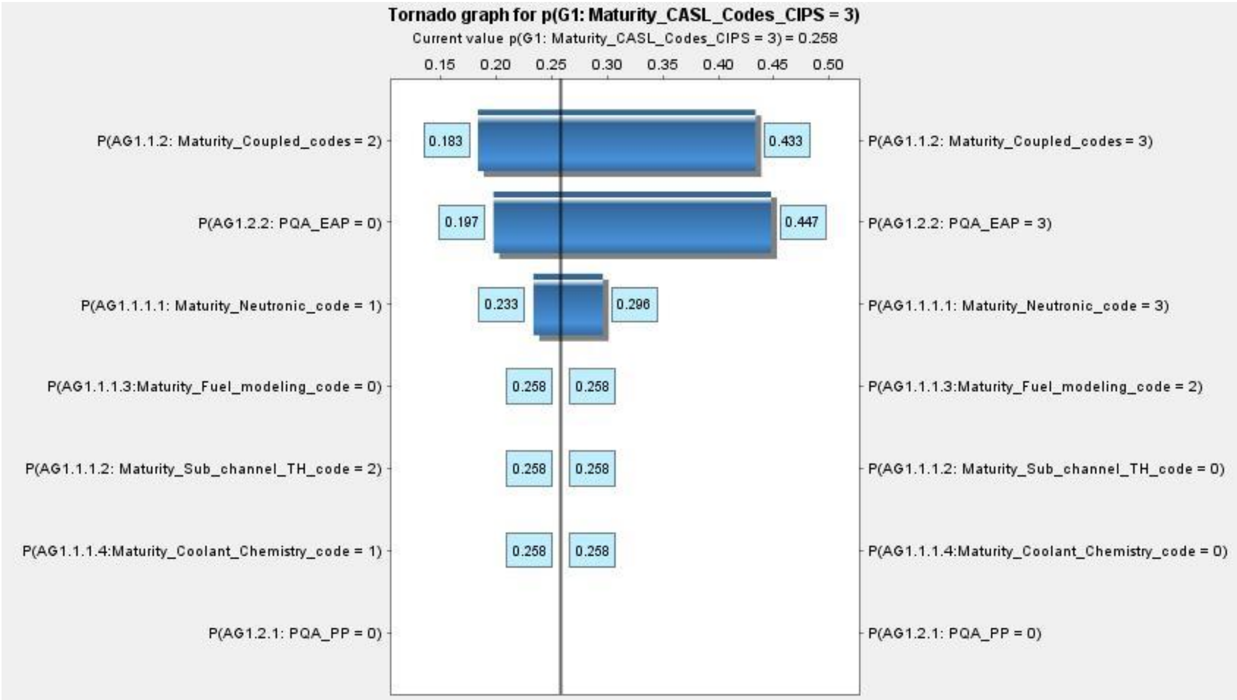


Figure 6.11: Tornado graph for grade ‘3’ for maturity of CASL codes (G1)in Figure 5.19

6.2. Sources of uncertainty in the decision model

We can identify different sources of uncertainty based on the parameters and elements in the decision model that are subject to expert opinion. The input for different parameters in the decision model like weight factor for attributes and utility of maturity levels are decided based on the expert opinion. Each expert may have different perception regarding the importance of attributes and utility of maturity levels. Figure 6.12 shows an example for the utility of maturity levels based on the opinion of two experts. The disparity in the experts' opinion may lead to uncertainty in the decision model. As PIRT/phenomenology pyramid is also based on subjective information, it is also a source of large epistemic uncertainty. Table 6.1 shows different sources of uncertainty in the decision model with examples.

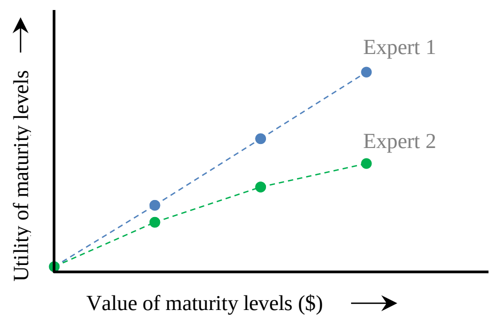


Figure 6.12: Utility of maturity levels based on different expert opinion

Table 6.1: Sources of uncertainty

Type of uncertainty	Example
Decision parameter-based uncertainty	• Weight factors for attributes • Utility of maturity levels
Structure-based uncertainty	• PIRT/Phenomenology pyramid • Structure of the decision model
Other	• Uncertainty in the grades assigned to the evidence

One way to minimize the uncertainty in the decision model is to calibrate the decision model based on the input of the decision maker and decision facilitator after the initial run of the decision model. Psychological scaling [105] can also be used to minimize the uncertainty caused by the differences in the experts' opinion.

CHAPTER 7: CONCLUSIONS AND RECOMMENDATION FOR FUTURE WORK

Modeling and simulation tools are extensively used to support decisions regarding design, development, and safety assessment of nuclear reactors. Therefore, systematic processes and methodologies were developed to assess the credibility of simulation tools for intended applications. Despite comprehensive procedures and guidelines provided in these methodologies, the "adequacy decision" is still left to engineering judgment. Current maturity assessment methodologies, like "Predictive Capability Maturity Model" (PCMM), although comprehensive, provide only high-level guidance.

This work provides a systematic approach that enables clarity and traceability in the maturity assessment process, and facilitates the integration of information for thorough confidence assessment.

The major contribution of this work is the development of a systematic technique for evidence-based quantitative maturity assessment for reliability assessment of a modeling and simulation tool for an intended application. This technique helps in identyfing the major areas of concern in terms of modeling capability, data needs, and quality of assessment process.

This chapter provides a summary of the dissertation, highlights of contributions and recommendation for future work.

7.1. Summary

This dissertation presents a systematic and formalized framework for the assessment of decision regarding the adequacy of a modeling and simulation tool for an intended use (primary focus→ code validation assessment). The framework consists of different elements that encompass structural knowledge representation, evidence classification and characterization, maturity assessment, and refinement.

The proposed framework is developed using an argument modeling technique called Goal structuring notation (GSN). GSN is widely used for the representation of assurance argument in safety cases. We employ GSN to facilitate structural knowledge representation, information abstraction and evidence incorporation in the framework. We also use GSN to develop the skeletal structure for quantitative maturity assessment. The decision schema for the development of the formalized decision model is based on the architecture of PCMM and Analytic hierarchy process (AHP). The number of levels in the decision hierarchy depends upon the required depth and rigor of the analysis. Each attribute and sub-attribute in the decision model is formulated as a claim where the degree of validity of the claim (attribute's assessment) is expressed using different maturity levels (a credibility scale). The GSN representation of the decision model is transformed into a confidence network to provide quantitative maturity assessment using the Bayesian network. Evaluation is performed by comparing the target level for different attributes with their achieved level based on the evidence. A metric based on expected utility theory is proposed to measure the distance between target level and achieved level on a scale of 0 to 1. The capabilities of the framework were demonstrated by two different case studies.

7.2. Contributions

The key contribution of this dissertation are as follows:

(1) **The development of an approach for classification and characterization of evidence for code's maturity assessment.** Code verification, validation, and uncertainty quantification are all confidence-building processes that require

continuous testing, learning, exploration, and documentation. Classification and characterization of evidence help in segregating and filtering important information from codes manual for thorough maturity assessment. The evidence are classified as direct evidence and indirect evidence. Direct evidence supports assessment attribute that directly affects the decision of code's maturity. Indirect evidence supports the assessment of process quality assurance factor.

(2) **The application of Goal structuring notation (GSN) for structural knowledge representation in the maturity assessment process.** In this framework, GSN is used to represent the phenomenology pyramid and decision model. GSN helps in explicitly specifying the strategy of decomposition, assumptions and contextual information in the hierarchical phenomenology pyramid and the decision model. Modular extension in GSN is used to manage large networks in the framework and indicators are used to highlight undeveloped entities or incomplete assessment.

(3) **The application of Bayesian network for evidence-based quantitative maturity assessment of code.** Bayesian network is used to facilitate abstraction of maturity information from lower level attribute to higher level attributes. As expert opinion plays important role in the assessment process, this information is assimilated with the objective data based on evidence using subjective probabilities and causal relation in the Bayesian network.

(4) **The development of a metric based on expected utility theory for comparing maturity of different attributes.** Comparison of attributes' assessment is important for formulating action items for refinement of the framework. It helps in comparing

the assessment result of different decision attributes and identification of major issues related to data, models, and quality of assessment process.

7.3. Recommendation for future work

(1) The input for different parameters in the decision model like weight factor for attributes and utility of maturity levels are decided based on the expert opinion. Each expert may have different perception regarding the importance of attributes and utility of maturity levels. Variation in the opinion of experts may lead to uncertainty. Psychological scaling can be used to minimize these uncertainties. Future work can be focused on the incorporation of physiological scaling models or other technique to minimize the uncertainties in experts' opinion.

(2) The quality of the maturity framework is governed by the level of detail of the decision schema. Complete set of lower level attributes based on data applicability and validation results was developed for code validation assessment. However, other attributes are same as the primary set in PCMM. Future work can develop, and incorporate detail lower level attributes set for all the PCMM attributes.

(3) Currently, the process of transformation from GSN to the Bayesian network is not automated. Future work can be focused on automating the transformation process.

REFERENCES

[1] B. Boyack, R. Duffey, G. Wilson, P. Griffith, G. Lellouche, S. Levy, *et al.*, "NUREG/CR-5249: Quantifying reactor safety margins: application of code scaling, applicability, and uncertainty evaluation methodology to a large-break, loss-of-coolant accident," *Nuclear Regulatory Commission*, 1989.

[2] EMDAP, " Regulatory Guide 1.203: Transient and Accident Analysis Methods," *US Nuclear Regulatory Commission, Office of Nuclear Regulatory Research*, 2005.

[3] W. L. Oberkampf and C. J. Roy, *Verification and validation in scientific computing*: Cambridge University Press, 2010.

[4] NASA, "Standard for Models and Simulations(NASA-HDBK-7009)," *National Aeronautics and Space Administration, Washington, D.C.*, 2008.

[5] AIAA, "AIAA Guide for the Verification and Validation of Computational Fluid Dynamics Simuations " *American Institute of Aeronautics and Astronautics, AIAA-G-007-1998*, 1998.

[6] T. P. Kelly, *Arguing safety: a systematic approach to managing safety cases*: University of York, 1999.

[7] T. L. Saaty, "The analytic hierarchy process: planning," *Priority Setting. Resource Allocation, MacGraw-Hill, New York International Book Company*, p. 287, 1980.

[8] L. A. Zadeh, "Fuzzy sets," *Information and control*, vol. 8, pp. 338-353, 1965.

[9] J. Pearl, *Bayesian networks: A model of self-activated memory for evidential reasoning*: University of California (Los Angeles). Computer Science Department, 1985.

[10] J.-B. Yang and M. G. Singh, "An evidential reasoning approach for multiple-attribute decision making with uncertainty," *IEEE Transactions on systems, Man, and Cybernetics*, vol. 24, pp. 1-18, 1994.

[11] J. S. Kaizer, " NUREG/KM-0006: Fundamental Theory of Scientific Computer Simulation Review," *US Nuclear Regulatory Commission, Office of Nuclear Reactor Regulation*, 2013.

[12] U. K. MoD, "Interim Standard 00-55: The procurement of safety critical software in defence equipment," *UK Ministry of Defence, Directorate of Standardization, Kentigern House 65*, 1991.

[13] M. Ishii, S. Revankar, T. Leonardi, R. Dowlati, M. Bertodano, I. Babelli, *et al.*, "The three-level scaling approach with application to the Purdue University Multi-Dimensional Integral Test Assembly (PUMA)," *Nuclear Engineering and Design*, vol. 186, pp. 177-211, 1998.

[14] N. Zuber, G. E. Wilson, M. Ishii, W. Wulff, B. Boyack, A. Dukler, *et al.*, "An integrated structure and scaling methodology for severe accident technical issue resolution: Development of methodology," *Nuclear Engineering and Design*, vol. 186, pp. 1-21, 1998.

[15] N. Zuber, U. S. Rohatgi, W. Wulff, and I. Catton, "Application of fractional scaling analysis (FSA) to loss of coolant accidents (LOCA): Methodology development," *Nuclear Engineering and Design*, vol. 237, pp. 1593-1607, 2007.

[16] H. Simon, "The architecture of complexity," *Proceedings of the American Philosophical Society*, vol. 106, pp. 467-482, 1962.

[17] M. Griffiths, J. P. Schlegel, T. Hibiki, M. Ishii, I. Kinoshita, and Y. Yoshida, "Phenomena identification and ranking table for thermal-hydraulic phenomena during a small-break

LOCA with loss of high pressure injection," *Progress in Nuclear Energy*, vol. 73, pp. 51-63, 2014.
[18] T. J. Olivier and S. Nowlen, *A Phenomena Identification and Ranking Table (PIRT) exercise for nuclear power plant fire modeling applications*: US Nuclear Regulatory Commission, Office of Nuclear Regulatory Research, 2008.
[19] V. A. Mousseau, N. Dinh, H. S. Abdel-Khalik, B. Williams, R. Smith, B. Adams, *et al.*, "Demonstration of Integrated DA/UQ for VERA-CS on a Core Physics Progression Problem," *Consortium for Advanced Simulation of Light Water Reactors, CASL-I-2014-0158-000*, 2014.
[20] C. Jones, A. Hetzler, N. Dinh, M. Sieger, and P. Athe, "Initial Verification and Validation Assessment for VERA (L2_VVI.VUQ.P15.02_2017_05_10)," *Consortium for Advanced Simulation of Light Water Reactors, CASL-U-2017-1310-000*, 2017.
[21] D. Diamond, C. Edgar, M. Fratoni, H. Gougar, A. Hawari, J. Hu, *et al.*, "Phenomena Identification and Ranking Tables (PIRT) Report for Fluoride High-Temperature Reactor (FHR) Neutronics," U. S. Department of Energy, N.E.U.P, CRMP-2016-08-0012016.
[22] G. E. Wilson and B. E. Boyack, "The role of the PIRT process in experiments, code development and code applications associated with reactor safety analysis," *Nuclear Engineering and Design*, vol. 186, pp. 23-37, 1998.
[23] D. J. Diamond, "NUREG-76750-2006-CP: EXPERIENCE USING PHENOMENA IDENTIFICATION AND RANKING TECHNIQUE (PIRT) FOR NUCLEAR ANALYSIS," *Brookhaven National Laboratories, Upton, NY*, 2006.
[24] T. L. Saaty, "What is the analytic hierarchy process?," in *Mathematical models for decision support*, ed: Springer, 1988, pp. 109-121.
[25] W. L. Oberkampf, M. Pilch, and T. G. Trucano, "SAND2007-5948: Predictive Capability Maturity Model for Computational Modeling and Simulation," *Sandia National Laboratories, Albuquerque, New Mexico*, 2007.
[26] J. Rider, J. R. Kamm, V. G. Weirs, and D. G. Cacui, "SAND2010-234P: Verification, validation and Uncertainty Quantification Workflow in CASL," *Sandia National Laboratories, Albuquerque, New Mexico*, 2010.
[27] B. Boyack, I. Catton, R. Duffey, K. Katsma, G. Lellouche, S. Levy, *et al.*, "Quantifying reactor safety margins part 1: an overview of the code scaling, applicability, and uncertainty evaluation methodology," *Nuclear Engineering and Design*, vol. 119, pp. 1-15, 1990.
[28] ANSI/ANS-10.4-2008, "Verification and Validation of Non-Safety-Related Scientific and Engineering Computer Programs for the Nuclear Industry," *American Nuclear Society*, 2008.
[29] ASME, "Guide for verification and validation in computational solid mechanics," *American Society of Mechanical Engineers, ASME Standard V & V 10-2006*, 2006.
[30] M. J. Steele, "NASA Handbook for Models and Simulations: An Implementation Guide for NASA-STD-7009," *National Aeronautics and Space Administration, NASA-HDBK-7009*, 2013.
[31] V. A. Mousseau and N. Dinh, "CASL Verification and Validation Plan," *Consortium for Advanced Simulation of Light Water Reactors, CASL-U-2016-1116-000*, 2016.
[32] R. G. Hills, W. R. Witkowski, W. J. Rider, T. G. Trucano, and A. Urbina, "SAND2013-8051: Development of a fourth generation predictive capability maturity model," *Sandia National Laboratories report*, 2013.

[33] S. Holtzman, *Intelligent decision systems*: Addison-Wesley Longman Publishing Co., Inc., 1988.
[34] T. L. Saaty, "How to make a decision: the analytic hierarchy process," *European journal of operational research,* vol. 48, pp. 9-26, 1990.
[35] F. Zahedi, "The analytic hierarchy process-a survey of the method and its applications," *interfaces,* vol. 16, pp. 96-108, 1986.
[36] R. W. Youngblood, W. C. Arcieri, S. A. Faridi, and N. P. Kadambi, "NUREG/CR-6833: Formal Methods of Decision Analysis Applied to Prioritization of Research and Other Topics," *US Nuclear Regulatory Commission, Office of Nuclear Reactor Regulation,* pp. 20555-0001, 2003.
[37] D. Kahneman and A. Tversky, "Prospect theory: An analysis of decision under risk," *Econometrica: Journal of the econometric society,* pp. 263-291, 1979.
[38] D. B. Kothe, "CASL: The Consortium for Advanced Simulation of Light Water Reactors," *Bulletin of the American Physical Society,* vol. 55, 2010.
[39] R. K. Salko and M. N. Avramova, "COBRA-TF Subchannel Thermal-Hydraulics Code (CTF) Theory Manual Revision 0," *Consortium for Advanced Simulation of Light Water Reactors, CASL-U-2015-0054-000,* 2015.
[40] B. Collins, T. J. Downar, J. Gehin, A. Godfrey, A. Graham, D. Jabaay*, et al.*, "MPACT Theory Manual Version 2.0.0," 2015.
[41] B. Spencer, J. Hales, D. Gaston, D. Karpeev, R. Williamson, S. Novascone*, et al.*, "BISON contact improvements CASL FY14 report," *Technical Report INL/EXT-14-33285, Idaho National Laboratories, Idaho Falls, ID,* 2014.
[42] M. Jina and M. Short, "L3: MPO.CRUD.P8.02 Two-Phase Fluid Flow Modeling in CRUD using MAMBA-BDM," 2014.
[43] V. Heller, "Scale effects in physical hydraulic engineering models," *Journal of Hydraulic Research,* vol. 49, pp. 293-306, 2011.
[44] F. D'Auria and G. Galassi, "Scaling in nuclear reactor system thermal-hydraulics," *Nuclear Engineering and Design,* vol. 240, pp. 3267-3293, 2010.
[45] W. Wulff, "Scaling of thermohydraulic systems," *Nuclear Engineering and Design,* vol. 163, pp. 359-395, 1996.
[46] A. Petruzzi and F. D'Auria, "Approaches, relevant topics, and internal method for uncertainty evaluation in predictions of thermal-hydraulic system codes," *Science and Technology of Nuclear Installations,* vol. 2008, 2007.
[47] J. B. Reyes, "The Dynamical System Scaling Methodology," in *Proceeding of the 16th - International Topical Meeting on Nuclear Reactor Thermal Hydraulics NURETH -16*, Chicago, 2015.
[48] K. Talus, *Research handbook on international energy law*: Edward Elgar Publishing, 2014.
[49] IAEA, "SSG-23: The Safety Case and Safety Assessment for the Disposal of Radioactive waste," *International Atomic Energy Agency, Specific Safety Guide*, 2012.
[50] T. Govier, *A Practical Study of Argument*. Boston: Wadsworth Pub. Co., 2012.
[51] L. Sun, "Establishing confidence in safety assessment evidence," Department of Computer Science, University of York, 2012.
[52] S. Nair, T. Kelly, and J. L. de la Vara, "An evidential reasoning approach for assessing confidence in safety evidence," in *IEEE 26th International Symposium on Software Reliability Engineering (ISSRE)*, Washigton DC, 2015, pp. 541-552.

[53] L. Sun and T. Kelly, "Safety arguments in aircraft certification," in *Systems Safety 2009. Incorporating the SaRS Annual Conference, 4th IET International Conference on*, 2009, pp. 1-6.
[54] S. E. Toulmin, *The Uses of Argument*. U.K.: Cambridge University Press, 2003.
[55] Adelard, "Assurance and Safety Case Environment, ASCE 4.2," *London: Adelard LLP*, 2016.
[56] K. Attwood, P. Chinneck, M. Clarke, G. Cleland, M. Coates, T. Cockram, *et al.*, "GSN community standard version 1," *Origin Consulting Limited, York (November 2011)*, 2011.
[57] T. Kelly, "A six-step Method for Developing Arguments in the Goal Structuring Notation (GSN)," *Technical report. York Software Engineering*, 1998.
[58] J.-B. Yang and D.-L. Xu, "On the evidential reasoning algorithm for multiple attribute decision analysis under uncertainty," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 32, pp. 289-304, 2002.
[59] S. Nair, N. Walkinshaw, and T. Kelly, "Quantifying Uncertainty in Safety Cases Using Evidential Reasoning," in *Computer Safety, Reliability, and Security: SAFECOMP 2014 Workshops: ASCoMS, DECSoS, DEVVARTS, ISSE, ReSA4CI, SASSUR. Florence, Italy, September 8-9, 2014. Proceedings*, A. Bondavalli, A. Ceccarelli, and F. Ortmeier, Eds., ed Cham: Springer International Publishing, 2014, pp. 413-418.
[60] B.-S. Chen and C.-H. Wu, "Robust optimal reference-tracking design method for stochastic synthetic biology systems: T–S fuzzy approach," *IEEE Transactions on Fuzzy Systems*, vol. 18, pp. 1144-1159, 2010.
[61] Y. Hayashi, E. Tazaki, K. Yoshida, and P. Dey, "Medical Diagnois using Simplified Multi-dimensional Fuzzy Reasoning," in *Systems, Man, and Cybernetics, 1988. Proceedings of the 1988 IEEE International Conference on*, 1988, pp. 58-62.
[62] P. R. Innocent and R. I. John, "Computer aided fuzzy medical diagnosis," *Information Sciences*, vol. 162, pp. 81-104, 2004.
[63] P. J. King and E. H. Mamdani, "The application of fuzzy control systems to industrial processes," *Automatica*, vol. 13, pp. 235-242, 1977.
[64] C.-T. Lin and C. S. G. Lee, "Neural-network-based fuzzy logic control and decision system," *IEEE Transactions on computers*, vol. 40, pp. 1320-1336, 1991.
[65] E. Sanchez, "Fuzzy logic knowledge systems and artificial neural networks in medicine and biology," in *An Introduction to fuzzy logic applications in intelligent systems*, ed: Springer, 1992, pp. 235-251.
[66] F. Steimann and K. P. Adlassnig, "Fuzzy medical diagnosis," *Handbook of fuzzy computation*, p. G13, 1998.
[67] M. Sugeno, "An introductory survey of fuzzy control," *Information sciences*, vol. 36, pp. 59-83, 1985.
[68] K. Tanaka and M. Sugeno, "Stability analysis and design of fuzzy control systems," *Fuzzy sets and systems*, vol. 45, pp. 135-156, 1992.
[69] J. Yaochu and W. Lipo, "Fuzzy Systems in Bioinformatics and Computational Biology," ed: Springer, 2009.
[70] Y. Yuan and M. J. Shaw, "Induction of fuzzy decision trees," *Fuzzy Sets and systems*, vol. 69, pp. 125-139, 1995.
[71] R. Bělohlávek and G. J. Klir, *Concepts and fuzzy logic*: MIT Press, 2011.

[72] F. Herrera, E. Herrera-Viedma, and J. L. Verdegay, "A sequential selection process in group decision making with a linguistic assessment approach," *Information Sciences,* vol. 85, pp. 223-239, 1995.
[73] J. Kacprzyk, "Studies in Fuzziness and Soft Computing," 2000.
[74] M. Negnevitsky, *Artificial intelligence: a guide to intelligent systems*: Pearson Education, 2005.
[75] L. A. Zadeh, "The role of fuzzy logic in the management of uncertainty in expert systems," *Fuzzy sets and systems,* vol. 11, pp. 199-227, 1983.
[76] H.-J. Zimmermann, *Fuzzy sets, decision making, and expert systems* vol. 10: Springer Science & Business Media, 2012.
[77] F. Herrera, "Genetic fuzzy systems: taxonomy, current research trends and prospects," *Evolutionary Intelligence,* vol. 1, pp. 27-46, 2008.
[78] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control," *IEEE transactions on systems, man, and cybernetics,* pp. 116-132, 1985.
[79] T. J. Cleophas, A. H. Zwinderman, and SpringerLink (Online service). (2015). *Machine Learning in Medicine - a Complete Overview*. Available: http://proxying.lib.ncsu.edu/index.php?url=http://link.springer.com/openurl?genre=book&isbn=978-3-319-15194-6
[80] N. Fenton and M. Neil, *Risk assessment and decision analysis with Bayesian networks*: CRC Press, 2012.
[81] S. Imoto, T. Higuchi, T. Goto, K. Tashiro, S. Kuhara, and S. Miyano, "Combining microarrays and biological knowledge for estimating gene networks via Bayesian networks," *Journal of bioinformatics and computational biology,* vol. 2, pp. 77-98, 2004.
[82] R. Jansen, H. Yu, D. Greenbaum, Y. Kluger, N. J. Krogan, S. Chung*, et al.*, "A Bayesian networks approach for predicting protein-protein interactions from genomic data," *Science,* vol. 302, pp. 449-453, 2003.
[83] P. Lucas, *Bayesian networks in medicine: a model-based approach to medical decision making*: na, 2001.
[84] C. J. Needham, J. R. Bradford, A. J. Bulpitt, and D. R. Westhead, "A primer on learning in Bayesian networks for computational biology," *PLoS Comput Biol,* vol. 3, p. e129, 2007.
[85] T. D. Nielsen and F. V. Jensen, *Bayesian networks and decision graphs*: Springer Science & Business Media, 2009.
[86] S. Sadeghi, A. Barzi, and N. Zarrin-Khameh, "Automated medical decision making utilizing bayesian network knowledge domain modeling," ed: Google Patents, 2004.
[87] A. Stassopoulou, M. Petrou, and J. Kittler, "Application of a Bayesian network in a GIS based decision making system," *International Journal of Geographical Information Science,* vol. 12, pp. 23-46, 1998.
[88] F. Taroni, A. Biedermann, S. Bozza, P. Garbolino, and C. Aitken, *Bayesian networks for probabilistic inference and decision analysis in forensic science*: John Wiley & Sons, 2014.
[89] P. Trucco, E. Cagno, F. Ruggeri, and O. Grande, "A Bayesian Belief Network modelling of organisational factors in risk analysis: A case study in maritime transportation," *Reliability Engineering & System Safety,* vol. 93, pp. 845-856, 2008.
[90] P. Weber, G. Medina-Oliva, C. Simon, and B. Iung, "Overview on Bayesian networks applications for dependability, risk analysis and maintenance areas," *Engineering Applications of Artificial Intelligence,* vol. 25, pp. 671-682, 2012.

[91] J. Yu, V. A. Smith, P. P. Wang, A. J. Hartemink, and E. D. Jarvis, "Advances to Bayesian network inference for generating causal networks from observational biological data," *Bioinformatics*, vol. 20, pp. 3594-3603, 2004.
[92] E. Denney, G. Pai, and I. Habli, "Towards measurement of confidence in safety cases," in *Empirical Software Engineering and Measurement (ESEM), 2011 International Symposium on*, 2011, pp. 380-383.
[93] J. Guiochet, Q. A. Do Hoang, and M. Kaaniche, "A Model for Safety Case Confidence Assessment," in *International Conference on Computer Safety, Reliability, and Security*, 2015, pp. 313-327.
[94] R. Hawkins, T. Kelly, J. Knight, and P. Graydon, "A new approach to creating clear safety arguments," in *Advances in Systems Safety*, ed: Springer, 2011, pp. 3-23.
[95] " Assurance and Safety Case Environment, ASCE 4.2," ed. London: Adelard LLP, 2016.
[96] K. A. Maupin and L. P. Swiler, "Validationn metric for deterministic and probablistic data (SANDIA2017-1421)," *Sandia National Laboratories, Albuquerque, New Mexico*, 2017.
[97] N. Dinh, "CIPS Validation Data Plan (INL/EXT-12-25347)," *Idaho National Laboratory, Idaho Falls*, 2012.
[98] N. Dinh, "Validation Data to Support Advanced Code Development " presented at the 15th International Topical Meeting on Nuclear Reactor Thermal Hydraulics, NURETH-14, Pisa, Italy, 2013.
[99] D. Bestion, H. Anglart, J. Mahaffy, D. Lucas, C. Song, M. Scheuerer*, et al.*, "Extension of CFD Codes Application to Two-Phase Flow Safety Problems-Phase 3 (NEA/CSNI/R(2014)13)," *Organisation for Economic Co-Operation and Development, Nuclear Energy Agency, Paris, France*, 2014.
[100] R. K. Salko, T. Blyth, C. A. Dances, J. W. Magedanz, C. Jernigan, J. Kelly*, et al.*, "CTF Validation and Verification," *Consortium for Advanced Simulation of Light Water Reactors, CASL-U-2016-1113-000*, 2016.
[101] R. W. Sterner and R. T. Lahey, "Air/water subchannel measurement of the equilibrium quality and mass flux distribution in a rod bundle," Rensselaer Polytechnic Institute1983.
[102] R. T. Lahey and F. J. Moody, *The thermal-hydraulics of a boiling water nuclear reactor*: Amer Nuclear Society, 1993.
[103] R. K. Salko and M. N. Avramova, "COBRA-TF Subchannel Thermal-Hydraulics Code (CTF) Theory Manual–Revision 0," CASL-U-2015-0054-000, The Pennsylvania State University2015.
[104] X. Cheng and U. Müller, *Review on critical heat flux in water cooled reactors* vol. 6825: FZKA, 2003.
[105] R. Cooke, *Experts in uncertainty: opinion and subjective probability in science*: Oxford University Press on Demand, 1991.

APPENDICES

APPENDIX A: USING FUZZY LOGIC FOR MATURITY QUANTIFICATION

This appendix presents simple examples to illustrate the use of fuzzy logic for quantitative maturity assessment. Maturity assessment is illustrated based on the code validation.

A.1. Example 1: Codifying expert knowledge

This example illustrates how fuzzy logic can be used to codify expert knowledge. Let us assume that we have a choice of four different models, P, Q, R, and S, to simulate an application Z. We have data from an experiment (evidence) that can be used to assess these models. We assume that this experiment is a good representative of the application, i.e. it covers the entire domain of the application and measurement error is also negligible. Comparison of model's prediction with experimental measurement gives bias B (or prediction error) for each model. We assume that this bias lies between 0 to 100% and define the crisp input using the bias B as, $x_B = 1 - B$ where x_B=[0 1]. Next, expert opinion is obtained to grade the maturity of each model on the basis of its model bias value (shown in second column of Table A 1). Using the fuzzy membership function shown in Figure A 1, we can codify the expert knowledge and obtain a measure of the maturity of each model on a scale from 0 to 1. Based on the expert's opinion, we choose the following memebership function for the model bias in this example:

$$\mu_B = exp\left(-\frac{(x_B-1)^2}{0.065}\right), 0 \le x_B \le 1 \qquad (1)$$

Experts use fuzzy quantifiers like " good", "excellent", "unacceptable" to grade an evidence. These quantifiers do not have sharp boundaries, e.g. in Table A 1 both model B and C are graded as "good" by the expert because the difference in their bias value is very small. Fuzzy logic captures this characteristic of expert knowledge using membership functions. Membership function are chosen based on the expert opinion about the model bias value. Model D has been graded as unacceptable by the expert, so we have $\mu_B = 0$ for this model.

Table A 1: Model evaluation example

Model	Bias (B)	Expert opinion	$x_B = 1 - B$	μ_B
P	2%	Very good	0.98	0.99
Q	10%	good	0.90	0.85
R	12%	good	0.88	0.80
S	60%	Unacceptable	0.4	0.00

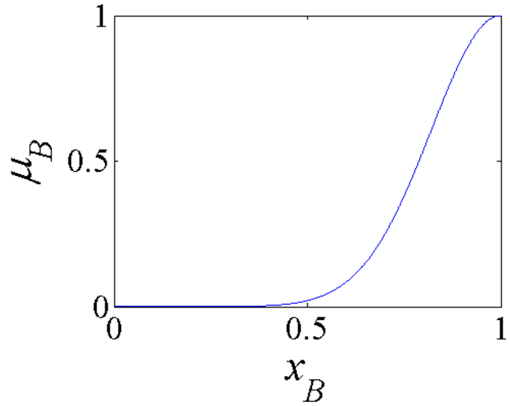


Figure A 1: Membership function for model bias

A.2. Example 2: Maturity assessment for a decision model using Fuzzy logic

The second example is based on the validation example shown in Figure 3.18 in chapter 3. The reiteration of this example using the Fuzzy logic is shown in Figure A 2. In this example, maturity quantification implies a quantitative evaluation of the claim G1, i.e., "*Code x is suitable for predicting the application of interest.*" We use the acronym CA to represent code adequacy, VR to represent validation results, and DA to represent data adequacy. We assume that confidence in code adequacy for this example is dependent on two factors only: (1) Validation result (VR) and (2) Data applicability (DA). Therefore, the FIS for this example consists of two inputs (VR and DA) and one output (CA) (see Figure A 2 for the confidence network).

Figure A 3 shows the membership functions for these variables. The inference rules for this example are shown in Figure A 4. A surface plot for CA corresponding to these rules is shown in Figure A 5 for different value of DA and CR. If DA is evaluated by scaling analysis, we can obtain the crisp input for Data Applicability by, DA=1-SD, where SD represents the scale distortion. The crisp input for the validation result can be obtained in terms of the bias by, VR=1-Bias. The surface plot represents "codified expert knowledge" for the evaluation of code adequacy based on the nature of the two evidence. This surface plot can be modified by changing the membership functions or the rules base or both.

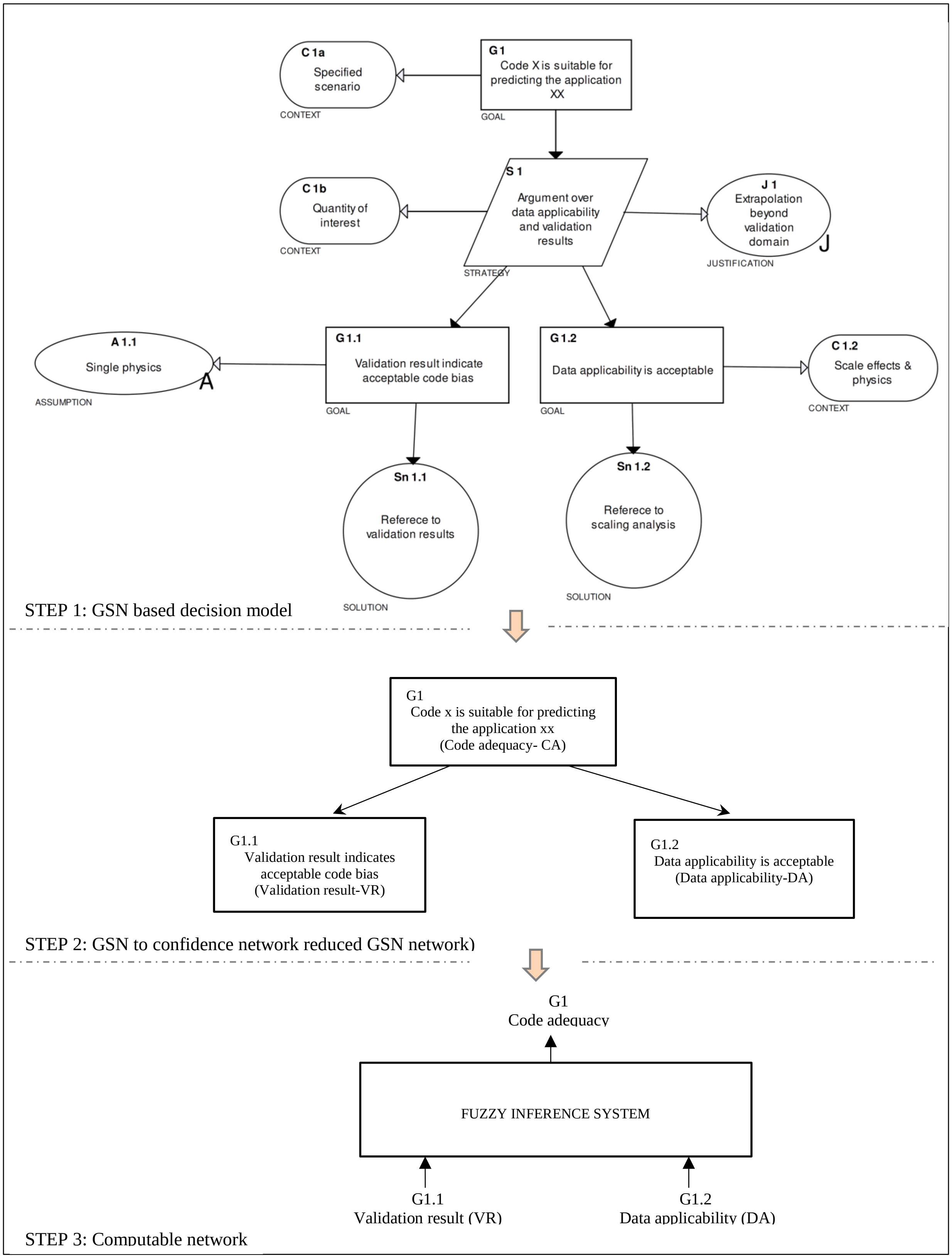


Figure A 2: Transformation of GSN tree to computable network

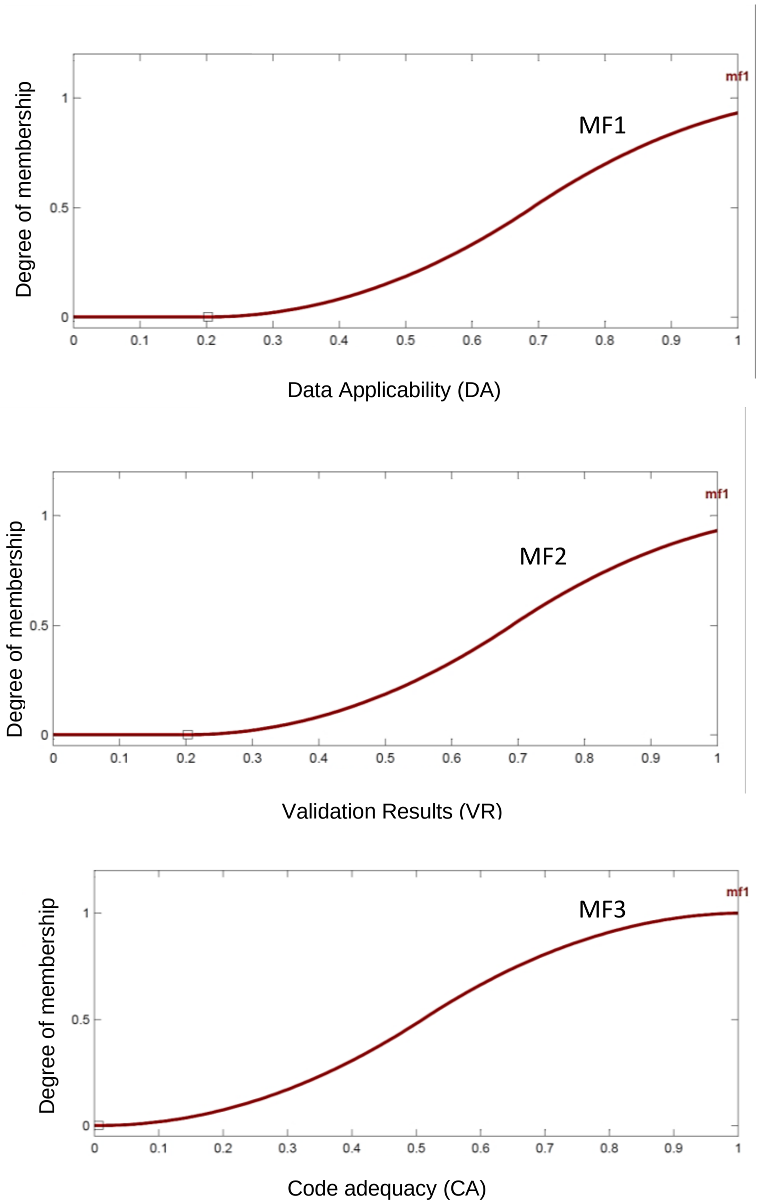


Figure A 3: Membership function for the input and output variable

If (DA is MF1) and (VR is MF2) then (CA is MF3)

Figure A 4: Fuzzy rule base for the fuzzy inference system (FIS)

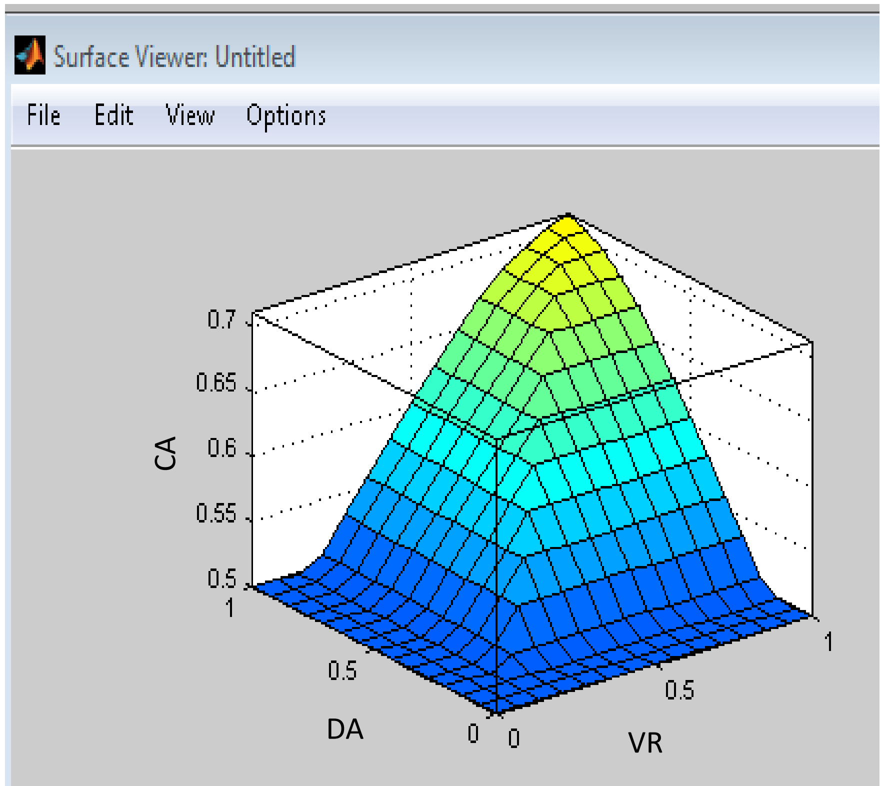


Figure A 5: Surface plot for code adequacy (CA)