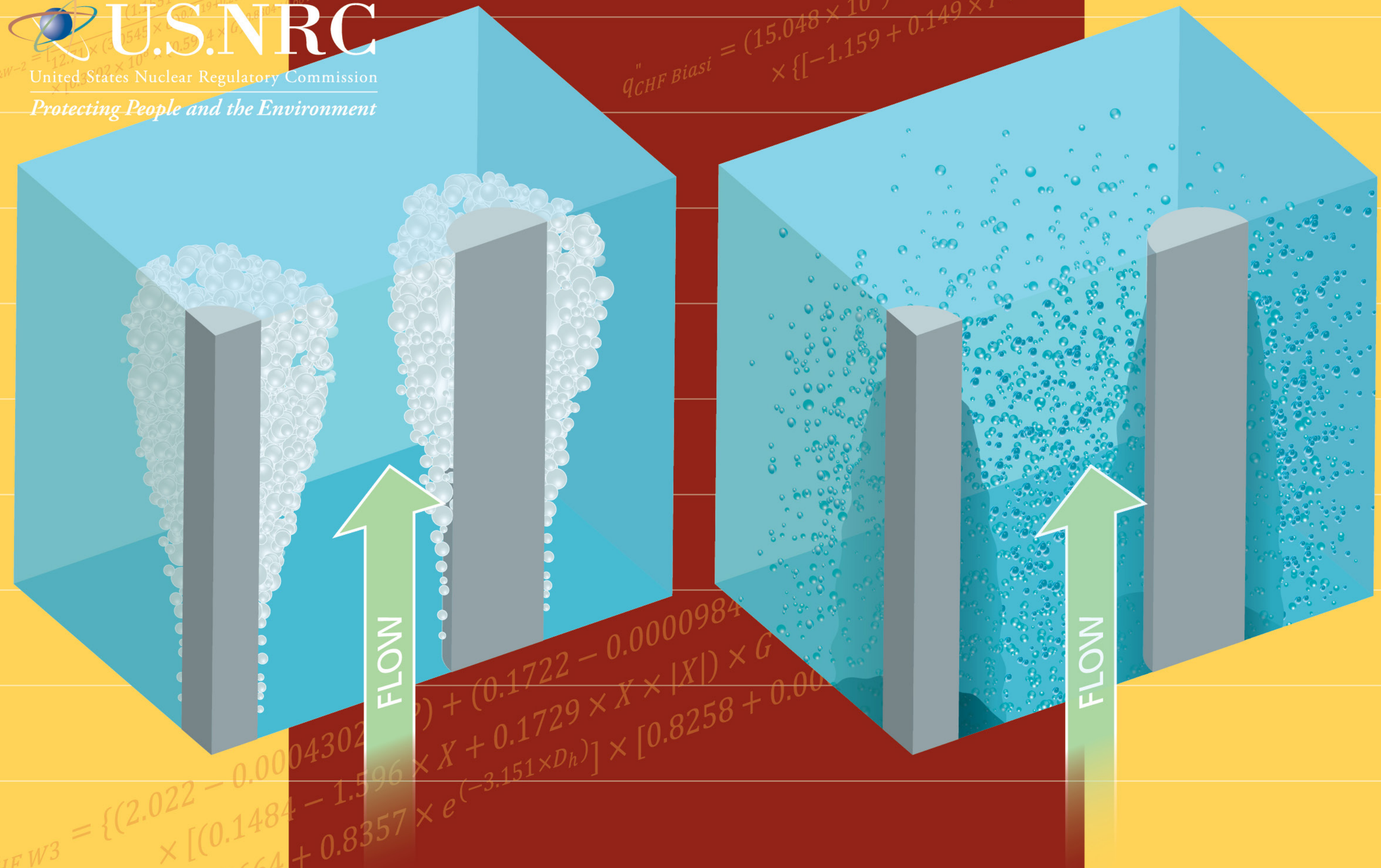
U.S.NRC

United States Nuclear Regulatory Commission

Protecting People and the Environment



CREDIBILITY ASSESSMENT FRAMEWORK FOR CRITICAL BOILING TRANSITION MODELS

A Generic Safety Case to Determine the Credibility of Critical Heat Flux and Critical Power Models

Draft Report for Comment

AVAILABILITY OF REFERENCE MATERIALS IN NRC PUBLICATIONS

NRC Reference Material

As of November 1999, you may electronically access NUREG-series publications and other NRC records at the NRC's Public Electronic Reading Room at http://www.nrc.gov/reading-rm.html. Publicly released records include, to name a few, NUREG-series publications; *Federal Register* notices; applicant, licensee, and vendor documents and correspondence; NRC correspondence and internal memoranda; bulletins and information notices; inspection and investigative reports; licensee event reports; and Commission papers and their attachments.

NRC publications in the NUREG series, NRC regulations, and Title 10, "Energy," in the *Code of Federal Regulations* may also be purchased from one of these two sources.

1. **The Superintendent of Documents**
   U.S. Government Publishing Office
   Mail Stop SSOP
   Washington, DC 20402-0001
   Internet: http://bookstore.gpo.gov
   Telephone: 1-866-512-1800
   Fax: (202) 512-2104

2. **The National Technical Information Service**
   5301 Shawnee Road
   Alexandria, VA 22161-0002
   http://www.ntis.gov
   1-800-553-6847 or, locally, (703) 605-6000

A single copy of each NRC draft report for comment is available free, to the extent of supply, upon written request as follows:

U.S. Nuclear Regulatory Commission
Office of Administration
Publications Branch
Washington, DC 20555-0001
E-mail: distribution.resource@nrc.gov
Facsimile: (301) 415-2289

Some publications in the NUREG series that are posted at the NRC's Web site address http://www.nrc.gov/reading-rm/doc-collections/nuregs are updated periodically and may differ from the last printed version. Although references to material found on a Web site bear the date the material was accessed, the material available on the date cited may subsequently be removed from the site.

Non-NRC Reference Material

Documents available from public and special technical libraries include all open literature items, such as books, journal articles, transactions, *Federal Register* notices, Federal and State legislation, and congressional reports. Such documents as theses, dissertations, foreign reports and translations, and non-NRC conference proceedings may be purchased from their sponsoring organization.

Copies of industry codes and standards used in a substantive manner in the NRC regulatory process are maintained at—

The NRC Technical Library
Two White Flint North
11545 Rockville Pike
Rockville, MD 20852-2738

These standards are available in the library for reference use by the public. Codes and standards are usually copyrighted and may be purchased from the originating organization or, if they are American National Standards, from—

American National Standards Institute
11 West 42nd Street
New York, NY 10036-8002
http://www.ansi.org
(212) 642-4900

Legally binding regulatory requirements are stated only in laws; NRC regulations; licenses, including technical specifications; or orders, not in NUREG-series publications. The views expressed in contractor-prepared publications in this series are not necessarily those of the NRC.

The NUREG series comprises (1) technical and administrative reports and books prepared by the staff (NUREG-XXXX) or agency contractors (NUREG/CR-XXXX), (2) proceedings of conferences (NUREG/CP-XXXX), (3) reports resulting from international agreements (NUREG/IA-XXXX), (4) brochures (NUREG/BR-XXXX), and (5) compilations of legal decisions and orders of the Commission and Atomic and Safety Licensing Boards and of Directors' decisions under Section 2.206 of NRC's regulations (NUREG-0750).

COMMENTS ON DRAFT REPORT

Any interested party may submit comments on this report for consideration by the NRC staff. Comments may be accompanied by additional relevant information or supporting data. Please specify the report number **NUREG/KM-0013** in your comments, and send them by the end of the comment period specified in the *Federal Register* notice announcing the availability of this report.

Addresses: You may submit comments by any one of the following methods. Please include Docket ID **NRC-2019-0043** in the subject line of your comments. Comments submitted in writing or in electronic form will be posted on the NRC website and on the Federal rulemaking website http://www.regulations.gov.

Federal Rulemaking Website: Go to http://www.regulations.gov and search for documents filed under Docket ID **NRC-2019-0043**.

Mail comments to: Office of Administration, Mail Stop: TWFN-7-A60M, U.S. Nuclear Regulatory Commission, Washington, DC 20555-0001, ATTN: Program Management, Announcements and Editing Staff.

For any questions about the material in this report, please contact: Joshua Kaizer, Reactor Engineer, 301-415-1532 or by e-mail at Joshua.Kaizer@nrc.gov.

Please be aware that any comments that you submit to the NRC will be considered a public record and entered into the Agencywide Documents Access and Management System (ADAMS). Do not provide information you would not want to be publicly available.

United States Nuclear Regulatory Commission
Protecting People and the Environment

NUREG/KM-0013

Credibility Assessment Framework for Critical Boiling Transition Models

A generic safety case to determine the credibility of critical heat flux and critical power models

Manuscript Completed:
Date Published:

Prepared by:
J.S. Kaizer
R. Anzalone
E. Brown
M. Panicker
S. Haider
J. Gilmer
T. Drzewiecki
A. Attard (retired, unable to comment on final version)

U.S. Nuclear Regulatory Commission

Office of Nuclear Reactor Regulation

ABSTRACT

Critical boiling transition (CBT) occurs when a flow regime that has a higher heat transfer rate transitions to a flow regime that has a significantly lower heat transfer rate. Models that predict a CBT are a necessary part of reactor safety analysis because they are used to determine plant safety limits. Therefore, the review of CBT models has been a focus of the U.S. Nuclear Regulatory Commission (NRC) since its inception in 1975.

This work presents a generic safety case in the form of a credibility assessment framework that combines aspects of goal structuring notation and maturity assessment. This framework is focused on the credibility assessment of CBT models with specific application to reactor safety analysis. The NRC has performed many such assessments and has generated this framework based on the experience of current and former NRC staff, as well as previous staff reviews as summarized in staff evaluations. This document includes a survey of the important technical and regulatory literature; a detailed technical discussion of CBT models and their application; and a suggested framework for CBT models. This NUREG/KM summarizes the knowledge the NRC staff has developed over the course of 40 years of CBT model and analysis reviews.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

Frameworks, such as presented in this paper, are the result of tremendous effort by numerous individuals. While these individuals and their technical contributions are too numerous to list, the authors offer special thanks to Robert Weisman and Julie Ezell for their legal review and advice which resulted in significant improvement to the document.

ABBREVIATIONS AND ACRONYMS

1-D	one-dimensional
2-D	two-dimensional
3-D	three-dimensional
AOO	anticipated operational occurrence
ASME	American Society of Mechanical Engineers
BWR	boiling-water reactor
CBT	critical boiling transition
CFR	*Code of Federal Regulations*
CHF	critical heat flux
CP	critical power
DNB	departure from nucleate boiling
DNBR	departure from nucleate boiling ratio
G	Goal
GSN	goal structuring notation
LOCA	loss-of-coolant accident
M&S	modeling and simulation
MDNBR	minimum departure from nucleate boiling ratio
NRC	U.S. Nuclear Regulatory Commission
PCT	peak cladding temperature
PWR	pressurized-water reactor
R- or K-factor	relative power factor
SAFDL	specified acceptable fuel design limit
SLMCPR	safety limit minimum critical power ratio
SRP	Standard Review Plan
SSC	systems, structures, and components
V&V	verification and validation

1 INTRODUCTION

Critical boiling transition[1] (CBT) is defined as a transition from a boiling flow regime that has a higher heat transfer rate to a flow regime that has a significantly lower heat transfer rate. For scenarios in which the heat transfer is controlled by the heat flux (such as in nuclear fuel assembly), the reduction in heat transfer rate caused by the CBT results in an increase in the surface temperature in order to maintain the heat flux. If the reduction in the heat transfer rate and resulting increase in surface temperature is large enough, the surface may weaken or melt. In a nuclear power plant, this cladding softening or melting is considered fuel damage.

To ensure that the fuel is not damaged during normal operation or anticipated operational occurrences (AOOs), computer simulations of the fuel are performed to predict the thermal-hydraulic conditions that would occur in the fuel assemblies during various scenarios. The resulting thermal-hydraulic conditions are then input to a CBT model.[2] That CBT model predicts the power which is required for a CBT to occur at the given thermal-hydraulic conditions. Hence the margin to CBT can be obtained by comparing the current power at the specific location in the fuel assembly to the power at which CBT occurs at the same thermal-hydraulic conditions. The U.S. Nuclear Regulatory Commission (NRC) has historically accepted that one way to demonstrate the avoidance of fuel damage during all normal operation and AOOs is to demonstrate that there is margin to a CBT.

Because of the importance of CBT models, a major focus in reactor safety analysis is to determine whether the proposed models can correctly predict CBT. The NRC has reviewed many CBT models over the years and has documented why each model was found acceptable (i.e., able to correctly predict CBT) in the corresponding safety evaluation. The authors of this document have used those safety evaluations along with their own expertise to produce a framework for assessing the credibility of CBT models.

This document includes two main sections. The first section contains a brief background of literature relevant to the assessment of CBT models followed by a discussion of the CBT phenomena and how such phenomena are commonly modeled. The second section describes the development of the credibility assessment framework for CBT models and provides detailed aspects of that framework as well as the evidence[3] commonly used to demonstrate that the criteria in the framework have been satisfied. In total, this document is meant to act as a textbook for those interested in the assessment of CBT models.

1 Many terms have been used to describe these models, including critical heat flux, critical power, critical quality versus boiling length, departure from nucleate boiling, dryout, burnout, and flow boiling crisis.

2 Historically, the models are commonly referred to as correlations because they correlate the CBT phenomenon to other variables in the flow field. However, the term “correlation” has a very specific meaning in statistics; therefore, this document will refer to them as “models.”

3 Evidence as used throughout this document is not intended to mean the rules and legal principles that govern the proof of facts in a legal proceeding. Rather, as used in this document, “evidence” is the available body of facts or information indicating whether a belief or proposition is true or valid.

1.1 Why Use the Term "Critical Boiling Transition"?

Hewitt and Hall-Taylor (1970) discussed a wide range of terms used to describe the phenomenon associated with dryout and critical heat flux (CHF). They noted that the "large diversity of terms tends to be confusing and this diversity reflects a continuing search for a term which is both descriptive and scientifically accurate." They analyzed the most common terms used (burnout, departure from nucleate boiling (DNB), dryout, and CHF); recognized that each term had its own inadequacies and merits; and chose "burnout" as the least unsatisfactory term. Unfortunately, the current literature on the subject does not reflect their choice, which seems to have settled mostly on the term "critical heat flux," although dryout and DNB are still commonly used.

Although CHF is technically independent of any specific phenomena, it is very closely tied to the phenomena of DNB, which occurs when nucleate boiling becomes inadequate to transfer the heat at the fuel surface to the coolant. At that point, the boiling regime begins to depart from nucleate boiling and begins "transition boiling," which is the boiling regime between nucleate boiling and film boiling. The close association between CHF and the phenomenon of DNB is likely due to the fact that CHF is the quantity used to determine whether DNB will occur in a pressurized-water reactor (PWR). However, CHF is typically not the quantity used to determine whether dryout (i.e., the drying out of the thin annular film in contact with the fuel cladding) has occurred in a boiling-water reactor (BWR). Additionally, the heat flux that causes a phenomenon to occur (i.e., the CHF) is different from the phenomenon itself. In technical discussions, the authors found it necessary to separate the phenomenon from any quantity associated with it.

Even considering all of these arguments, the authors of this document, like Hewitt and Hall-Taylor, were hesitant to introduce new terminology and initially decided to use the common term "critical heat flux." However, as the discussion became more detailed and finer distinctions were necessary, the authors reluctantly decided that a different term was necessary and could not be avoided. Therefore, the authors chose to use the term "Critical boiling transition," because it better describes the pertinent phenomena and allows for the necessary distinctions. Because CBT is a new term, we repeat its definition here: CBT[4] is defined as a transition from a boiling flow regime that has a higher heat transfer rate to a flow regime that has a significantly lower heat transfer rate.

1.2 What Is Credibility?

The term "credibility" has seen wide application in the modeling and simulation (M&S) community, specifically in the areas focusing on Verification and Validation (V&V). However, the term is often left undefined. The American Society of Mechanical Engineering's (ASME) V&V 10 Guide for Verification and Validation in Computational Solid Mechanics (2006) did not formally define the term, but did equate it to "*trustworthiness.*" Initially, NASA (2008) discussed the term, but purposefully chose not to define it and instead relied on "the usual sense of the English language." Later, they defined the term as "*the quality to elicit belief or trust in modeling and simulation results*" (NASA 2008[B]). Oberkampf and Roy (2010) do provide a definition for *credibility of computational results* – "results of an analysis that are worthy of belief or confidence," but this definition is not much more detailed than ASME's connection between credibility and trustworthiness. While credibility is intimately linked with trust, the component which is missing from these definitions is how much trust is needed in the specific use of the model. Therefore, the authors of this work have chosen to use a definition based on the work of Kaizer *et al*., (2015)

[4] While CBTs can exist on other surfaces, this work is concerned only with fuel rods used in light water nuclear power plants.

which captures the underlying link to trustworthiness, but maintains awareness of the necessity to make a decision.

Credibility is defined as *the determination that an object (in this particular instance, a model) can be trusted for its intended purpose*. As defined, this is a binary determination. Thus, an object is either deemed credible (i.e., can be trusted for its intended purpose) or not credible (i.e., cannot be trusted for its intended purpose). There are two interesting consequences from this definition of credibility. First, there is no middle ground, all objects must either be credible or not credible. Second, there is no “degree” of credibility. That is, by definition one object cannot be more credible than another. The authors fully acknowledge that some objects may certainly be more trusted than other objects. For example, one individual may have more experience and therefore be more trusted than another individual, or one simulation may be very well vetted and therefore be more trusted than another simulation. However, the credibility of those objects is defined to be binary (i.e., credible, not credible) because decisions themselves are binary (i.e., yes or no).

1.3 <u>What Is a Credibility Assessment Framework?</u>

A credibility assessment framework provides a means to assess whether an object can be trusted for its intended purpose. Such a framework can be thought of as one form of a safety case. A safety case is defined as “a structured argument, supported by a body of evidence that provides a compelling, comprehensible, and valid case that a system is safe for a given application in a given operating environment.”[5] Although various ways exist to provide a safety case (e.g., every safety evaluation produced by the NRC can be thought of as the documentation of a safety case or collection of safety cases), this document makes use of concepts formalized in goal structuring notation (GSN). GSN (GSN Working Group, 2011) “is a graphical argumentation notation that can be used to document explicitly the individual elements of any argument (claims, evidence, and contextual information) and, perhaps more significantly, the relationships that exist between these elements (i.e., how claims are supported by other claims, and ultimately by evidence, and the context that is defined for the argument).” See Denney *et al*. (2011) for an example of GSN.

The framework presented here combines the logic structure of GSN with the evaluation aspects of maturity assessment. Maturity assessment (Kaizer *et al*., 2015) is focused on measuring how “mature” an object is in specific attributes compared to its possible minimum and maximum amount of maturity in those attributes. Maturity assessment frameworks, such as the Predictive Capability Maturity Model (Oberkampf *et al*., 2007) and NASA-STD-7009 (NASA 2008[B]), focus on the evidence that is available and is a means to rank that evidence in a manner useful to a decision maker. For a more detailed description of a maturity assessment and its history, see Oberkampf and Roy (2010).

The credibility assessment framework used in this document is unique in that it combines these two concepts by using the logical structure of goals from GSN and the evaluation of the possible evidence from maturity assessment. The framework is generated from a single main goal. That main goal is then logically decomposed into subgoals. By logical decomposition, we mean the act of generating a set of sub-goals which are logically equivalent to the original goal (i.e.,

[5] This document uses the definition provided by the United Kingdom’s Ministry of Defense (2007). Other U.S. government agencies which have made use of this concept include NASA (2015) and the FDA (2014). The authors’ use of the UK Ministry of Defense definition in this document does not imply USNRC approval of regulatory principles or approaches employed in the UK, nor should the use of the definition be understood to be an NRC endorsement of such principles or approaches as acceptable for use in the US.

necessary and sufficient for the original goal to be met). This decomposition is expressed using GSN notation. Each subgoal can either be further logically decomposed into other subgoals, or if no further decomposition is deemed useful, the subgoal may be considered a base goal and evidence must be provided to demonstrate that the base goal is true. The evidence which is commonly provided is given in a maturity table, where it is ranked from least to most mature. A simple example to illustrate the logic is given below.

The main goal (G) is written as a conclusion, such as "*G - It is safe to drive over the bridge*." Notice that this goal is somewhat ambiguous. What is meant by "safe"? While there is common agreement that it should be "safe" to drive over a bridge, there is disagreement as to what "safe" means in this instance. Such ambiguity is often encountered, but frameworks such as the one provided in this document can be used to define what these ambiguous terms (such as "safe") mean in practice.

The main goal, G, is then logically decomposed into a set of sub-goals, where each sub-goal must be necessary (i.e., if the sub-goal is false, the main goal must also be false) and the set of sub-goals must be sufficient (i.e., if the set of sub-goals is true, the main goal must also be true) to demonstrate that the main goal is true. This simple example has two subgoals: (1) "*The bridge can withstand the weight of my car."* and (2) *"There will not be a natural disaster while I am driving over the bridge*." These goals are given in Figure 1 below.

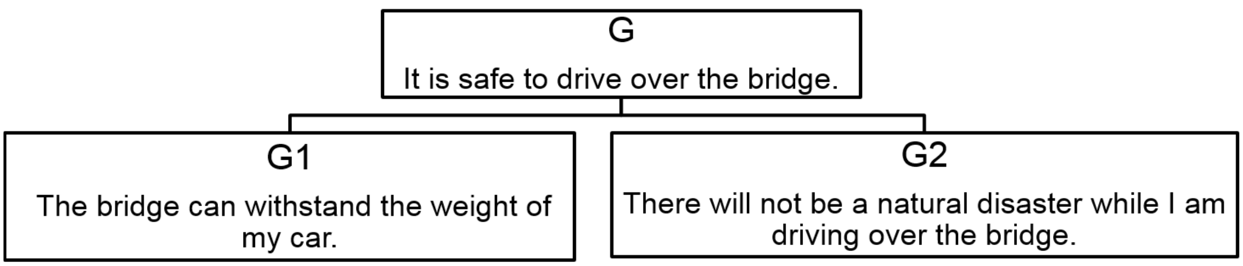


Figure 1 Goals

Each subgoal (e.g., G1 and G2) must either be further decomposed into additional sub-goals, or evidence provided to determine if those sub-goals could be considered true. For this example, no further decomposition was considered. Potential levels of evidence that could be provided to demonstrate that each subgoal is true (i.e., has been met) are given in Figure 2 below.

G

It is safe to drive over the bridge.

G1

The bridge can withstand the weight of my car.

Level	Evidence
0	Don't think about it.
1	Someone has checked it.
2	I drove over it yesterday.
3	Engineering analysis of the bridge demonstrates a significant margin.
4	Current measurements of the bridge along with analysis of the bridge shows that it is structurally sound.
5	A recent bridge inspection showed that the bridge is structurally sound.

G2

There won't be a natural disaster while I am driving over the bridge.

Level	Evidence
0	Don't think about it.
1	In my experience, natural disasters are rare.
2	There hasn't been a natural disaster in this area for at least 20 years.
3	Bridge has been built to withstand any seismic activity. Weather forecast shows no adverse conditions expected.

Figure 2 Framework

The evidence provided is the justification for concluding that the specific base goal is true (i.e., has been met). This evidence is ranked from least to most mature, or providing the least certain justification that the base goal is met to the most certain justification. With higher levels of evidence (e.g., level 3 as opposed to level 1), we can be more certain that the associated base goal is true. Thus, an individual driving over a bridge on his or her daily commute would likely require a very low level of evidence to determine the bridge is credible (i.e., safe to drive across). In all likelihood, the individual may not even consciously think about the credibility of the bridge, or if he or she did, the individual would likely rely on low levels of evidence. However, if the bridge were used to transport heavy haul freight (i.e., oversized loads), a much higher level of evidence would likely be required before the bridge was deemed credible.

The specific pieces of evidence which are considered by this framework are given in Figure 2. If any other evidence (i.e., levels of G1 or G2) are used to demonstrate that the associated goal is true, that evidence should be placed in its appropriate rank in the table. Thus, one could argue that seeing another car drive over the bridge right before they do is evidence that the bridge can withstand the weight of their own car. If this evidence is going to be used, it should be ranked according to the other evidence already in the table (likely falling between levels 2 and 3 and requiring a re-numbering of the table).

Notice that the ambiguity of the word "safe" in the main goal G has now been removed. That is, by saying "*It is safe to drive over the bridge,*" we have not only defined safe as meaning G1 and G2 are true, but we would also state what evidence was given (e.g., Level 3 for G1 and Level 2 for G2). Thus, the ambiguous word "safe" is explicitly defined using the framework.

Additionally, anything not specified in the framework was not considered in determining credibility. Because the framework explicitly establishes the assumptions underlying an assessment, it can be helpful when identifying any areas that may need further consideration (that is, additional subgoals or evidence levels). For example, an individual could argue that our sample framework lacks a sub-goal that accounts for the driving ability of other drivers on the bridge. Another may argue that our first sub-goal should not only consider the weight of our car, but all other vehicles on the bridge at the same time. One of the largest advantages to these frameworks is that others can quickly and easily determine what was and what was not considered. Further, the framework could be updated quickly and easily to account for any omissions.

1.4 Credibility Assessment Framework for Critical Boiling Transition Models

The credibility assessment framework presented in this work is focused on critical boiling transition models. While this framework was generated based on the NRC staff's experience reviewing these models, the framework itself is more broadly applicable to any use of any CBT model. This includes the entire spectrum of possible uses from something as simple as a homework problem to something as significant as reactor safety analysis, and all uses in between. It is important to remember that the appropriate evidence level will change based on the model's intended use. Thus, the level of evidence appropriate for reactor safety analysis will likely be much higher than that which is appropriate for a homework problem.

As this framework is applicable to any use of a CBT model (including, but not limited to, reactor safety analysis), the authors have chosen to use a broader terminology when describing the details of the framework as it can be applied to determining credibility. The process of determining credibility involves two distinct roles: the analyst and the assessor.[6] It is the role of the analyst to generate the model, gather the evidence, and present the argument that the model can be trusted. It is the role of the assessor to determine if the evidence presented is sufficient to justify that the model can be trusted for its intended purpose. In regulatory environments, these roles are usually filled by separate individuals from different organizations, the analyst being the applicant and the assessor being the regulatory agency staff member (e.g., at the NRC, this role is typically called a reviewer). However, in other environments both roles could be performed by individuals from the same organization (i.e., internal peer review), and in some cases could be performed by the same individual (e.g., a homework problem).

[6] The 'assessor' is not a reference to a specific role as defined by other national or international organizations. Instead, the word was chosen solely based on the fact that the person who applies the credibility *assessment* framework is making an assessment, and is therefore an assessor.

2 BACKGROUND ON CRITICAL BOILING TRANSITION

2.1 Literature Survey

This section provides a literature survey of the references considered important for the NRC review of CBT models. Many references associated with CBT phenomena exist; however, the following are of special interest because they are commonly cited in discussions of the models used in nuclear power reactors. For convenience, the references have been separated into technical references (i.e., textbooks, articles, and industry reports) and regulatory references.

2.1.1 Technical References

Tables 1, 2, 3, and 4 list the key technical references for CBT models.

Table 1 Key Textbooks for the Review of CBT Models

Author	Title	Date
Hewitt and Hall-Taylor	Annular Two-Phase Flow	1970
Tong	Boiling Crisis and Critical Heat Flux	1972
Todreas and Kazimi	Nuclear Systems I: Thermal Hydraulic Fundamentals	1990
Lahey and Moody	The Thermal Hydraulics of a Boiling Water Nuclear Reactor	1993
Tong and Tang	Boiling Heat Transfer and Two-Phase Flow	1997

Table 2 Key Papers for the Review of CBT Models

Author	Title	Date
Leidenfrost	On the Fixation of Water in Diverse Fire	1756 (1966)
Tong *et al*.	Influence of Axially Nonuniform Heat Flux on DNB	1965
Macbeth	An Appraisal of Forced Convection Burnout Data	1965–1966
Barnett	A Correlation of Burnout Data for Uniformly Heated Annuli and Its Uses for Predicting Burnout in Uniformly Heated Rod Bundles	1966
Healzer *et al*.	Design Basis for Critical Heat Flux Condition in Boiling Water Reactors	1966
Tong	Prediction of Departure from Nucleate Boiling for an Axially Non-Uniform Heat Flux Distribution	1967
Biasi *et al*.	Studies on Burnout: Part 3—A New Correlation for Round Ducts and Uniform Heating and Its Comparison with World Data	1967
Gellerstedt *et al*.	Correlation of Critical Heat Flux in a Bundle Cooled by Pressurized Water	1969
Hughes	A Correlation of Rod Bundle Critical Heat Flux for Water in the Pressure Range 150 to 725 psia	1970
Piepel and Cuta	Statistical Concepts and Techniques for Developing, Evaluating, and Validating CHF Models and Corresponding Fuel Design Limits	1993
Groeneveld	The 2006 CHF Look-Up Table	2007
Yang *et al*.	Uniform versus Nonuniform Axial Power Distribution in Rod Bundle CHF Experiments	2014
Kaizer	Identification of Nonconservative Subregions in Empirical Models Demonstrated Using Critical Heat Flux Models	2015
Groeneveld	CHF Data Used to Generate 2006 Groeneveld CHF Lookup Tables	2016

Table 3 Industry Reports Associated with CBT Models for PWRs

CBT Model	Title	Date
B&W-2	Correlation of Critical Heat Flux in a Bundle Cooled by Pressurized Water	1970
CE-1	C-E [Combustion Engineering] Critical Heat Flux: Critical Heat Flux Correlation for C-E Fuel Assemblies with Standard Spacer Grids, Part 1—Uniform Axial Power Distribution	1976
XNB DNB	Exxon Nuclear DNB Correlation for PWR Fuel Designs	1983
WRB-1	New Westinghouse Correlation WRB-1 for Predicting Critical Heat Flux in Rod Bundles with Mixing Vane Grids	1984
WRB-2	VANTAGE 5H Fuel Assembly	1985
CE-1 (modified)	C-E Critical Heat Flux: Critical Heat Flux Correlation for C-E Fuel Assemblies with Standard Spacer Grids, Part 2—Non-Uniform Axial Power Distribution	1984
ANFP DNB	Departure from Nucleate Boiling Correlation for High Thermal Performance Fuel	1990
BWU	The BWU Critical Heat Flux Correlations	1996
WRB-2M	Modified WRB-2 Correlation, WRB-2M, for Predicting Critical Heat Flux in 17x17 Rod Bundles with Modified LPD Mixing Vane Grids	1999
BWU Addendum 1	The BWU Critical Heat Flux Correlations Applications to the Mark-B11 and Mark-BW17 MSM Designs	2000
BWU Addendum 2	Application of BWU-Z CHF Correlation to the Mark-BW 17 Fuel Design with Mid-Span Mixing Grids	2002
ABB-NV and ABB-TV	Addendum 1 to WCAP-1 4565-P-A Qualification of ABB Critical Heat Flux Correlations with VIPRE-01 Code	2004
HTP	Departure from Nucleate Boiling Correlation for High Thermal Performance Fuel	2005
BHTP	BHTP DNB Correlation Applied with LYNXT	2005
BWU Addendum 3	The BWU-B11R CHF Correlation for the Mark-B11 Spacer Grid	2005
WSSV and WSSV-T	Westinghouse Correlations WSSV and WSSV-T for Predicting Critical Heat Flux in Rod Bundles with Side Supported Mixing Vanes	2007
ACH-2	The ACH-2 CHF Correlation for the U.S. EPR	2007

CBT Model	Title	Date
ABB-NV (extended) and WLOP	Addendum 2 to WCAP-14565-P-A Extended Application of ABB-NV Correlation and Modified ABB-NV Correlation WLOP for PWR Low Pressure Applications	2008
WNG-1	Westinghouse Next Generation Correlation (WNG-1) for Predicting Critical Heat Flux in Rod Bundles with Split Vane Mixing Grids	2010
WRB-1 and WRB-2	Thermal Design Methodology	2013
KCE-1	KCE-1 Critical Heat Flux Correlation for PLUS7 Thermal Design	2012
ORFEO	The ORFEO-GAIA and ORFEO-NMGRID Critical Heat Flux Correlations	2016

Table 4 Industry Reports Associated with CBT Models for BWRs

CBT Model	Title	Date
GE transient CHF	Loss-of-Coolant Accident and Emergency Core Cooling Models for General Electric Boiling Water Reactors	1971
GEXL	General Electric Thermal Analysis Basis Data, Correlation and Design Application	1977
ANFB	ANFB Critical Power Correlation	1990
R-Factors	R-Factor Calculation Method for GE11, GE12, and GE13 Fuel	1999
D2	10x10 SVEA Fuel Critical Power Experiments and CPR Correlations: SVEA-96+	1999
D1	10x10 SVEA Fuel Critical Power Experiments and CPR Correlations: SVEA-96	2000
GEXL96	GEXL96 Correlation for ATRIUM-9B Fuel	2001
GEXL10	GEXL10 Correlation for GE12 Fuel	2001
GEXL80	GEXL80 Correlation for SVEA96+ Fuel	2004
D4	10x10 SVEA Fuel Critical Power Experiments and CPR Correlation: SVEA-96 Optima2	2005
GEXL97	GEXL97 Correlation Applicable to ATRIUM-10 Fuel	2008
D4 (Modified R-Factor)	SVEA-96 Optima2 CPR Correlation (D4): Modified R-factors for Part-Length Rods	2009
D4 (High and Low Flow)	SVEA-96 Optima2 CPR Correlation (D4): High and Low Flow Applications	2009
GEXL17	GEXL17 Correlation for GNF2 Fuel	2009
SPCB	SPCB Critical Power Correlation	2009
GEXL14	GEXL14 Correlation for GE 14 Fuel	2011
ACE/ATRIUM-10	ACE/ATRIUM-10 Critical Power Correlation	2014
ACE/ATRIUM-10 XM	ACE/ATRIUM 10XM Critical Power Correlation	2014
D5	10x10 SVEA Fuel Critical Power Experiments and New CPR Correlation: D5 for SVEA-96 Optima3	2013
ACE/ATRIUM-11	ACE/ATRIUM-11 Critical Power Correlation	2015

2.1.2 Regulatory References

The regulatory references are separated into the following types:

- Regulations. The *Code of Federal Regulations* (CFR) sets forth regulations that licensees must satisfy.

- Guidance. Following NRC guidance is one way to satisfy the corresponding regulations. Such guidance can be found in NRC Regulatory Guides and NRC publications in specified NUREGS. In addition, the application regulations require an applicant to identify and describe all differences in design features, analytical techniques, and procedural measures proposed for a facility compared to those in NUREG-0800, “Standard Review Plan for the Review of Safety Analysis Reports for Nuclear Power Plants: LWR Edition” (SRP). Previous safety evaluations can also inform the staff’s review of an application.

- Generic Communications. The NRC may choose to send out a general communication on an issue for numerous reasons. Generic communications include administrative letters, bulletins, circulars, generic letters, information assessment team advisories, information notices, regulatory issue summaries, security advisories, and documents for comment.

Table 5 lists the regulatory references associated with CBT models in reactor safety analyses.

Table 5 Regulatory References Associated with CBT Models

Type	Title	Date
Regulations	10 CFR Part 50, "Domestic Licensing of Production and Utilization Facilities," Appendix A, "General Design Criteria for Nuclear Power Plants," General Design Criterion 10, "Reactor Design"	N/A
Regulations	10 CFR 50.36, "Technical specifications"	N/A
Regulations	10 CFR 50.34, "Contents of Applications; Technical Information"	N/A
Regulations	10 CFR Part 50, Appendix B, "Quality Assurance Criteria for Nuclear Power Plants and Fuel Reprocessing Plants"	N/A
Guidance	SRP Section 4.2, "Fuel System Design"	2007
Guidance	SRP Section 4.4, "Thermal and Hydraulic Design"	2007
Generic Communication	Information Notice 2014-01, "Fuel Safety Limit Calculation Inputs Were Inconsistent with NRC-Approved Correlation Limit Values"	2014
Standard	NQA-1 - Quality Assurance Requirements for Nuclear Facility Applications	2015

2.1.2.1 10 CFR Part 50, Appendix A, General Design Criterion 10

General Design Criterion (GDC) 10 in 10 CFR Part 50, Appendix A, is the principal regulation associated with a CBT. This criterion introduces the concept of specified acceptable fuel design limits (SAFDLs). In essence, SAFDLs are those limits placed on certain variables to ensure that the fuel does not fail. One such SAFDL is associated with CBT. Because the decrease in heat transfer following a CBT could result in fuel failure, a SAFDL is used to demonstrate that a CBT does not occur during normal operation and AOOs. Therefore, fuel failure is precluded during normal operation and AOOs.[7]

SRP Section 4.4 includes the following two SAFDLs for use in accounting for the uncertainties involved in developing and using a CBT model (e.g., uncertainties in the values of process parameters, core design parameters, calculation methods, instrumentation) and ensuring that fuel failure is precluded:

[7] Experiencing such a transition may not immediately result in fuel failure. The decrease in heat transfer and subsequent increase in fuel temperature may not be enough to cause the cladding to weaken or melt. Therefore, the point of CBT is considered to be a conservative limit compared to the actual point of fuel damage.

(1) There should be a 95-percent probability at the 95-percent confidence level that the hot fuel rod in the core does not experience a CBT during normal operation or AOOs.

(2) At least 99.9 percent of the fuel rods in the core will not experience a CBT during normal operation or AOOs.

Typically, SAFDL No. 1 is associated with PWRs, and SAFDL No. 2 is associated with BWRs.

Before May 21, 1971, when the GDC took effect, the Atomic Energy Commission (AEC), the predecessor to the NRC, approved construction permits for nuclear power plants based on plant-specific Principal Design Criteria (PDC) that applicants proposed in their construction permit applications as required by the then-extant provisions of 10 CFR 50.34(a). The AEC published proposed General Design Criteria in the Federal Register (32 FR 10213) on July 11, 1967, sometimes referred to as the AEC Draft GDC, which were generally consistent with the PDC previously proposed in applications for construction permits. AEC Draft GDC 6 is the relevant draft GDC and is substantially similar to the current GDC 10. AEC Draft GDC 6 also calls for the reactor core to be designed with appropriate margin to specified limits that preclude fuel damage.

2.1.2.2 10 CFR 50.36

The second regulation associated with a CBT is 10 CFR 50.36, part of which focuses on defining technical specification safety limits. There are multiple limits that are associated with CBT models used during plant operation. These limits can be operating limits, alarms, analysis limits, and safety limits. Generally, only the safety limit and associated limiting conditions for operation (LCOs) and surveillance requirements (SRs) are included in the plant’s technical specifications. The safety limit associated with CBT is typically focused on an accurate quantification of the uncertainty of the CBT model and may also include the quantification of additional uncertainties as well.

2.1.2.3 10 CFR 50.34

The third regulation associated with a CBT is in 10 CFR 50.34, which focuses on defining the information that a licensee must present to ensure safe operation. Specifically, 10 CFR 50.34(a)(4) requires that the Preliminary Safety Analysis Report (PSAR) include determination of the margins of safety during normal operation and AOOs. One of these is the margin to CBT, which verifies that fuel failure is precluded during normal operation and AOOs through analysis.

2.1.2.4 10 CFR Part 50, Appendix B

The fourth regulation associated with a CBT appears in 10 CFR Part 50, Appendix B. It requires licensees to include certain structures, systems, and components (SSCs) in a quality assurance program that satisfies specific criteria. Appendix B, Criterion III, requires that specified design control measure be applied to the design of safety-related SSCs and these measures apply to safety analyses for these SSCs. The CBT model is a key component of the safety analysis subject to 10 CFR Part 50, Appendix B.

2.1.2.5 Other Regulations

Both 10 CFR 50.46 and 10 CFR Part 50, Appendix K, Section I.C.4 focus on the modeling of a nuclear power plant during accident scenarios. While many of these scenarios involve the use of CBT models, a different model may be used than is used to analyze SSC performance during

AOOs. For example, the CBT models used during LOCAs are typically low pressure, conservative models, which are not necessarily fuel design specific. While these models are reviewed by the NRC as a part of any accident evaluation model, they typically are not a major focus during those reviews.

2.2 Critical Boiling Transition Phenomena

A CBT occurs when a flow regime that has a higher heat transfer rate transitions to a flow regime that has a significantly lower heat transfer rate. In nuclear fuel rods, the heat flux from the fuel pellet to the fuel cladding is mostly independent of the heat transferred from the cladding surface to the coolant. As a result, the cladding temperatures will increase until the new heat transfer mechanisms can remove all of the heat from the pellet; the primary mechanism for post-CBT heat transfer will dictate the magnitude of the cladding surface temperature increase. Typically, the post-CBT heat transfer mechanism transfers heat at a much lower rate (i.e., it is less efficient) than the pre-CBT mechanism, and therefore causes a dramatic increase in clad temperature. The temperature increase resulting from CBT could cause the fuel rod cladding surface to weaken or melt and result in fuel failure, therefore it is considered a *critical* transition. Hence, the heat flux at which this transition occurs is known as the *critical heat flux*, the assembly power at which this transition occurs is known as the *critical power*, and the quality at which this transition occurs is known as the *critical quality*.

The difference in the rate of heat transfer associated with the flow regimes before and after the transition is a convenient way to understand the phenomena of CBT. The sections below discuss the two most common critical boiling transitions, DNB and dryout.

2.2.1 Departure from Nucleate Boiling

Departure from nucleate boiling results from a change in the flow regime from nucleate boiling to film boiling and is chiefly a concern in PWRs. During nucleate boiling, the bulk coolant, which is mostly liquid with some vapor, is in intimate contact with the cladding. Vapor is generated as bubbles on the cladding surface at nucleation sites. These bubbles grow on the surface, detach, and flow into the bulk coolant stream. As each bubble leaves the surface, cooler liquid fills the space near the surface that was formerly occupied by the bubble, and the boiling process is repeated. The growth, transport, and collapse of the bubbles increases turbulence close to the wall and causes increased mixing in the thermal boundary layer. Ultimately, this boiling results in extremely high heat transfer rates; therefore, the cladding surface is able to support high heat fluxes at relatively low surface temperatures.

Departure from nucleate boiling occurs when bulk liquid is prevented from coming into contact with the surface. The ultimate cause of the phenomenon is not fully understood but is believed to be bubble crowding that prevents liquid from contacting the surface. Once liquid coolant can no longer contact the surface, heat transfer to the liquid through convection is no longer possible, and the only mechanisms that transfer heat to the bulk liquid coolant are conduction through the vapor and radiation from the surface. At normal cladding temperatures, both of these types of heat transfer mechanisms are relatively inefficient, and the surface's temperature must dramatically increase to remove the heat generated in the pellet. This temperature increase is large enough to cause the surface to become unwettable, thus creating a dry patch. This dry patch may spread axially along the rod and blanket a large majority of the rod in vapor. Thus, the flow regime transitions to film boiling. This rapid increase in surface temperature may also result in fuel failure in a very short period of time.

2.2.2 Dryout

Dryout results from a change in the flow regime from annular flow around the fuel rods to dispersed flow and is mostly a concern in BWRs. In annular flow, a thin liquid film surrounds the cladding, and the bulk flow is mostly vapor with some liquid droplets. Convection transfers heat from the cladding to the annular film, causing some of the liquid in the annular film to evaporate from the film surface and thus adding more vapor to the bulk flow. It is currently believed that evaporation is the only “boiling” that occurs in the annular film boiling regime; no vapor formation occurs at nucleation sites, and no bubbles are generated.

As the coolant flows up the channel, it carries the liquid film up along the cladding surface. This results in entrainment of liquid droplets from the annular film into the bulk coolant, thus reducing the amount of liquid in the film. However, some of the droplets in the bulk coolant are also deposited back onto the film. This deposition will increase the amount of liquid in the film and is a chief concern in the design of grid spacers for BWR assemblies. In summary, as the liquid film flows up the cladding, evaporation and entrainment remove liquid from the film while deposition adds liquid to the film.

Dryout occurs when the annular film disappears completely. Upon reaching dryout, the bulk fluid transitions from annular flow to dispersed flow. In dispersed flow, there is no continuous liquid film on the cladding, and the bulk flow consists of a mixture of vapor and dispersed liquid droplets. Convection occurs between the vapor and the fuel rod. The droplets also act as a heat sink, as they are in the heated vapor and may absorb heat from the vapor as well as impact the heated rod (assuming the rod is still wettable). Generally, radiation is not a significant mode of heat transfer until the surface temperature become much higher (at which point, the rod is typically unwettable). Although the heat transfer is less in dispersed flow than in annular flow, it is still substantial. As a result, the increase in cladding temperature is typically not as dramatic as that resulting from DNB. However, sustained time in dryout will eventually result in fuel failure.

2.2.3 Other Flow Regimes and Transitions

It is important to recognize that the flow regimes inside a reactor core are not precisely defined. Further, potential transitions occur between flow regimes that are not considered “critical” or do not result in a “crisis” because the transition would not significantly reduce heat transfer. For example, different portions of a PWR fuel assembly may be in subcooled nucleate boiling (i.e., boiling which occurs when the bulk of the liquid is sub-cooled and not at saturation), nucleate boiling, and annular flow. Although a shift from nucleate boiling to another regime is technically a departure, it is only considered DNB if the new regime has a significantly lower heat transfer rate.

It is also important to note that the same CBT model will generally be applied in every flow regime in a given reactor type and is not associated with only a single flow regime. Thus, a specific DNB model used in a PWR will not only be used to predict whether the flow regime has transitioned from nucleate to film boiling, but will also be used to predict a transition from subcooled nucleate to film boiling or a transition from annular flow to film boiling, as all of those flow regimes can exist in a PWR assembly.

2.3 Determining When Critical Boiling Transition Occurs

Given certain key parameter values (e.g., flow rate, power, pressure, temperature), a CBT model predicts either the CHF or the critical power (CP) of the assembly that would cause a CBT. This predicted value is then compared to the current heat flux or assembly power to determine the margin to CBT. Typically, CHF models are used in PWRs, whereas CP models are used in BWRs.

2.3.1 Critical Heat Flux Models

Critical heat flux is the cladding surface heat flux that causes a CBT for a given set of local conditions. It is chiefly associated with PWRs and the phenomenon of DNB; however, as stated earlier, CHF models can also predict other CBTs (e.g., the transition from annular flow to film boiling). CHF models are developed through experiments where, under a given set of inlet flow conditions, power is increased until CHF is observed. A computer code is used to calculate the local flow conditions from the boundary conditions of the experiment, and CHF is correlated to those local flow conditions. Thus, when a computer code is used to simulate an AOO, the CHF model can use the local conditions calculated at any location in the core to predict the critical heat flux at that location. The predicted CHF is then compared with the local heat flux to determine the margin to CBT at that location.

2.3.2 Critical Power Models

Critical power is the assembly power that causes a CBT. It is chiefly associated with BWRs and the phenomenon of dryout; however, as stated earlier, CP models can also predict other critical flow transitions. Further, the term “CP model” is something of a misnomer because these models do not generally correlate CP to local conditions (as a CHF model does); instead, they correlate the critical quality (i.e., the quality that causes a CBT) to the boiling length (i.e., distance from the point of initiation of bulk boiling to the location of a CBT).[8] Thus, when a computer code is used to simulate an AOO, the inlet conditions (e.g., power, inlet flow) along with certain local conditions are used to calculate the quality at various axial elevations in the fuel. The quality at each axial elevation is compared to the critical quality at that elevation by assuming that the boiling length is the elevation of the location under consideration. Generally, the critical quality is much greater than the predicted quality; therefore, the assembly power is increased until, at some axial elevation, the critical quality is equal to the predicted local quality. The lowest assembly power at which at least one location equals a quality greater than or equal to the critical quality is known as the CP.

2.3.3 Semi-empirical Modeling

Since 1970, tremendous strides have been made in the generation of CBT models; however, these models are still predominantly semi-empirical (i.e., the models are based more on experimental data than on first-principle physics). Known physical behavior is often used to inform the model’s mathematical form, but empirical coefficients are still needed to ensure accurate model predictions. In effect, this means that, although the model may be informed by physics, they are not treated as theoretical models, but are treated as empirical or data-driven models in that they must be validated with experimental data and should not be used outside of the range of their validation database.

[8] This is not exclusively true because other models are more mechanistic than critical quality/boiling length correlations. Regardless, even these mechanistic boiling transition models do not generally correlate CP directly to fluid conditions.

2.3.4 Conservative vs. Non-Conservative Predictions

For CBT models, "conservative" means that the model will predict a CBT before the actual occurrence of the phenomenon (e.g., at lower powers, at lower flow rates, at lower qualities). Conversely, "non-conservative" means that the model will predict a CBT after the actual occurrence of the phenomenon (e.g., at higher powers, at higher flow rates, at higher qualities).

2.4 Applying a Critical Boiling Transition Model

Unlike many closure models[9] that are developed directly from experimental data, CBT models may[10] call for input that typically cannot be measured directly. In such instances, a thermal-hydraulic computer code is used. This code will calculate the values of key variables needed by the CBT model (e.g., local quality, local mass flux) using some set of field equations and any necessary closure models. Thus, the development and, more importantly, the validation of a CBT model may highly depend on the thermal-hydraulic computer code used and the code options selected. For this reason, using such a CBT model in a different code or in the same code with substantially different code options (e.g., different two-phase closure models) would call for a complete re-validation using those new code or code options.

The application of CBT models can be somewhat confusing. Many closure models, such as Dittus-Boelter (1930), operate as a simple function. The function takes in certain inputs and returns an output. Thus, the validation of such a function would ensure that, given the correct inputs, the function returns the correct output. Although CBT models follow a similar process, the models themselves cannot typically be used or validated in such a simple manner. For example, consider an experiment of a test assembly whose power is increased until a CBT occurs. For this experiment, the inlet flow rate and temperature, axial and radial power shapes, pressure, and assembly power have been measured. However, most CBT models do not use this measured data, but require a different set of data to make a prediction. Hence, the measured data from the experiment is input into a computer code and that code generates the required input data such that the CBT model can make a prediction. The following sections describe how this simple situation would be evaluated using methods commonly applied in a PWR and a BWR.

2.4.1 Applying a Critical Boiling Transition Model in a Pressurized-Water Reactor

In a PWR assembly, a prediction of the CHF would be calculated for each subchannel at each axial elevation. Consider a 5x5 assembly that contains 25 rods, 16 internal subchannels (i.e., between the rods), and 20 external subchannels (i.e., between the rods and the channel wall). Assume that the assembly has a height of 3.65 meters (12 feet) and that the computer code uses an axial nodalization of 7.62 centimeters (3 inches). In total, each subchannel would have 48 axial elevations; given the 36 subchannels (internal plus external), that results in a total of 1,728 nodes, each of which would have its own CHF prediction from the CBT model. Which of the 1,728 predictions should be compared with the single measured CHF from the experiment?

---

[9] Closure models are those additional models needed in order to "close" the problem. They provide the additional relations needed for the number of equations to equal to the number of unknowns so the problem can be solved. They supplement the conservation equations.

[10] Typically, CHF models used in PWRs are subject to this restriction. The subchannel code provides detailed information about the local flow conditions that the CHF model uses to make a prediction. Dryout models are generally less affected because they do not need detailed information about the local flow conditions.

At first glance, comparing the predicted CHF at the location where the CHF was indicated in the test would seem to be the best approach. Suppose that a thermocouple on one of the inside rods (i.e., a rod internal to the 5x5 array and not on the boundary) was the first thermocouple to indicate that a CHF occurred. Further, suppose that this thermocouple is located at an elevation of 2.74 meters (9 feet). This rod would be a member of four subchannels; thus, there may be no way to determine in which of the four surrounding subchannels CHF actually occurred. This does seem to make the problem more tractable because, instead of considering 1,728 nodes, that number is reduced to 4 nodes. However, in such experiments, multiple rods will experience the temperature rise associated with a CBT.[11] While one rod at one axial elevation will achieve such a temperature rise first, the temperature rise used to indicate CHF is somewhat arbitrary in that a slightly different criterion may result in a different CBT value. For example, changing the CBT criterion from a rise of 16.67 degrees Celsius (C) (30 degrees Fahrenheit (F)) to a rise of 11.1 degrees C (20 degrees F) may result in the selection of a different rod in CBT. Suppose there were five thermocouples indicating that CHF was very likely occurring at those locations. That would mean that of the 1728 nodes in the bundle, 20 would need to be considered for the CBT point.

The measured heat flux at the time of CHF could be compared to each of the predicted CHF values in the 20 nodes; however, it is not clear how a single “predicted CHF” value could be objectively chosen. While a ratio of “measured CHF” to “predicted CHF” could be found at each point (the measured value from the experiment and the predicted value from the CBT model) which of these 20 values should be taken as the value from this test? The maximum value, the minimum value, the mean of all 20 values? The usual practice is described below.

It is important to remember that the overall goal of a CBT model is to determine whether a CBT will occur. Thus, the validation process should focus on ensuring that the model appropriately predicts a CBT, and not necessarily that the model predicts CBT at the correct location. Thus, when using the model to make predictions “measured CHF” data will not be available for the reactor assembly under normal operation and AOO conditions. Considering the 5x5 assembly, only 1,728 predictions of CHF for each time step of the scenario will be available. Therefore, those predicted values of CHF are typically compared to the local values of heat flux to determine which of the nodes is closest to CHF using the departure from nucleate boiling ratio (DNBR). The DNBR is defined as the ratio of the predicted CHF of a node to the current heat flux of a node. Equation 1 gives the DNBR.

$$DNBR = \frac{q''_{CHF}}{q''_{Local}} \tag{1}$$

Notice that as long as the node is far from the conditions that cause CHF, the value of DNBR will be greater than 1. As the node approaches those conditions, the DNBR value approaches 1, and when heat flux in the node is equal to the CHF, the DNBR is equal to 1. Given that these simulations are used to demonstrate that CHF does not occur, the DNBR in all of the nodes should always be greater than 1. Further, the node with the smallest DNBR, commonly called minimum departure from nucleate boiling ratio (MDNBR), is the node closest to the conditions that cause CHF.

These concepts of DNBR and MDNBR are used to select a “predicted CHF” value to compare with the “measured CHF” value from the experiment. From the 1,728 nodes, the node that

[11] The temperature rise selected is usually on the order of 11.1 to 27.78 C (20 to 50 degrees F) in under 1 second.

contains the MDNBR could be used as the “predicted” node, and the CHF prediction at this node could be the “predicted CHF”. This may or may not be one of the 20 nodes discussed earlier, but using the CHF from this node as the “predicted CHF” results in a much more representative error of how the CBT model will be applied in practice. While the analyst may know which sub-channel and what elevation CHF occurred at in the experiment, this information is not known in the real-world scenario. Thus, this information should not be used in determining the model’s error. Instead, the “predicted CHF” value should be determined using the same method that will be used when the model is applied in the real-world scenario.

Note that the MDNBR location (and hence the “predicted” CHF value) may change during model development. Thus, as the model changes during its development, different nodal locations in different subchannels would likely be determined to be more the limiting node.

2.4.2 Applying a Critical Boiling Transition Model in a Boiling-Water Reactor

In a BWR assembly, the calculation of the predicted CP would consider each fuel rod in the assembly individually. Consider a 5x5 assembly that contains 25 rods that has a height of 3.65 meters (12 feet) an axial nodalization of 7.62 centimeters (3 inches). Most BWR methods do not model all of the rods and subchannels; instead, they model only a single rod surrounded by a single subchannel of fluid. Modeling all of the subchannels is considered unnecessary because the fuel assembly is contained within a channel; therefore, the water cannot flow between assemblies. To account for the varying thermal-hydraulic conditions at the different locations in the assembly, two different factors are used to “convert” the results of the single rod analysis and make it applicable to the entire assembly.

The first factor is a relative power factor, commonly called the R- or K-factor. The R- or K-factor accounts for the power in a specific rod compared to the powers in the surrounding rods. In the above example, a different R- or K-factor would be calculated for each of the 25 rods depending on each rod’s individual power, which can change over the cycle. The second factor is a thermal-mixing factor, commonly called an additive constant. The thermal-mixing factor accounts for the thermal performance at that specific *xy* location in the assembly. In the above example, a different thermal-mixing factor would be calculated for each of the 25 rods depending on the *xy* location of each rod in the assembly; that factor would not change for that assembly design.

Ideally, the local conditions calculated in the assembly could be directly correlated to the CP. However, this is not the case. A change in power has a dramatic impact on the entire flow field along the length of the assembly, and integral, not local, effects are commonly considered the cause of the CP and its associated phenomenon of dryout.[12] To determine the CP, the mass flow rate, axial and radial power shape, and pressure are fixed. The quality at a given elevation can then be compared to the predicted critical quality from the CBT model given the boiling length (i.e., the length from the start of boiling to the elevation of interest). The power input to the model is increased or decreased until the calculated quality at that location is equal to the critical quality. The corresponding power is the CP.

[12] This consideration of integral effects, as well as the concept of flow memory (Tong 1965), seems to be somewhat of a misnomer. Although what occurs upstream shapes the flow field, CBT occurs at a single location based on the conditions of the local fluid and the heat from the wall. If those local fluid conditions could be modeled perfectly, a consideration of integral effects would not be necessary. However, because of modeling limitations, many of the important parameters of that local fluid cannot be directly modeled; therefore, concepts such as flow memory are useful as modeling simplifications.

Because there are 25 rods, there could be 25 different CPs for each axial elevation. However, because many CBT models correlate the critical quality to the boiling length, it is not necessary to perform calculations below the boiling length. Additionally, it is not necessary to determine the power that would cause a CBT at a certain axial elevation. For example, suppose a CBT occurred on rod 17 at an axial elevation of 10 feet. If an analyst wanted to determine what power would cause a CBT at 8 feet, the power would need to be increased. However, increasing the power to cause a CBT at 8 feet would not make much sense because the goal is to avoid a CBT entirely, and at the current power level, a CBT has occurred. Thus, it is not the power that causes a CBT at every elevation that is important; instead, it is the lowest power that causes a CBT at any elevation at or below the top of the active fuel that is most important.

2.4.3 Applying a Steady-State Model to Transient Conditions

Generally, CBT models are generated with steady-state data (i.e., the test facility reaches a steady state and slowly increases the power until CBT occurs). Information from those data points is then used to generate CBT models. However, when the data are applied in a reactor safety analysis, the CBT model is applied to the transient (i.e., time-varying) conditions occurring during a scenario. Historically, this application of a correlation developed on steady-state data to transient conditions has been considered conservative, and often a few transient tests are performed to demonstrate that the prediction of a CBT model is conservative when it is applied in a transient fashion.

2.5 Addressing Uncertainties and Errors

Many uncertainties and errors are associated with a CBT model. First and foremost, some of these uncertainties have specific meanings and should be defined. In this work, a distinction is made between an error and an uncertainty. The term “error” focuses on the difference between specific predicted values and their corresponding specific “actual” value. For example, the error in a single measurement (absolute error or relative error) is a comparison of the true value to the measured value. The term “uncertainty” focuses on quantifying the variability of a set of values for future predictions. For example, while a prediction is generally a single value, it may be better to think of that prediction as a range of values where that range is defined by the uncertainty in the prediction. The various forms of uncertainties discussed throughout this document are defined as follows:

- Instrumentation uncertainty is associated with a specific instrument used in the experiment. This uncertainty is a result of the underlying precision of the instrument, and is typically provided by the manufacturer of the device in question. Examples include the ±0.50 degrees C (±0.90 degrees F) of a K-type thermocouple or the 1 percent uncertainty of a pressure transducer. Generally, instrumentation uncertainty (future behavior) is approximated through the instrumentation error (past behavior).

- Measurement uncertainty is the total uncertainty associated with recording the measurement from a piece of instrumentation. Although this is often considered to be simply the instrumentation uncertainty, that may be an oversimplification. Uncertainty is often associated with recording the value from the instrument. Data-logging systems typically read in voltages, but not all measurements are provided as a voltage, and these values would need to be converted. Additionally, some uncertainty occurs in the voltage reading of the data-logging system itself. For example, pressure transducers often provide an output between 4 to 20 milliamperes. This output must be converted through a resistor before it can be measured as a voltage. The uncertainty of the resistance in that resistor should be accounted for in the measurement uncertainty because it may not have been accounted for in the instrumentation uncertainty.

- Experimental uncertainty is the total uncertainty associated with recording the value of quantity of interest from an experiment. In many instances, an instrument that measures the quantity of interest may not be available, or even if one is available, that measurement may depend on multiple instruments. For example, the uncertainty associated with the CHF “measurement” would at least need to consider uncertainties associated with the measured power, the manufacturing tolerances of the heater rods (which influence the axial heat and heat flux shape), and the thermocouples used to determine when a CHF event occurs.

- Model error is the difference between the model’s predicted CHF or CP and the actual CHF or CP.

- Model application error is similar to model error, but it accounts for the fact that the CBT model is not used as a standalone equation, but used in a larger calculational framework.

- Validation error is a sample from the population of the model application error. If we consider the model application error as a set which contains the entire population of all possible uses of the model, then the validation error is the sample from that population for which a CHF or CP value was measured in a particular experiment.

- Model uncertainty is associated with the application of the CBT model in a future analysis. This may also be referred to as the predictive capability of the model. This uncertainty quantifies the difference (or ratio) between the power at which a model predicts CBT will occur and the power at which CBT would actually occur. Note that it is not only the uncertainty of how the model predicted the experimental data (i.e., the validation error) but also includes how the model would have predicted other experimental data (i.e., other samples from the model application error) and how that experimental data relates to the real world system of interest of the fuel assembly in a nuclear power plant.

- Plant parameter uncertainties are associated with specific plant parameters, such as flow, power, and pressures. Although these uncertainties do not generally affect the CBT model directly, they are used along with the CBT model to generate the safety limit.

3 CREDIBILITY ASSESSMENT FRAMEWORK

This section discusses the development of a credibility assessment framework for CBT models. As described above, this framework is a generic safety case expressed using concepts from GSN and maturity assessment. This framework was developed based on the experience of members of the NRC technical staff, documented safety evaluations from previous NRC reviews, and various documents found in the open literature. While it was the goal of the authors to have this framework be applicable to all uses of a CBT (i.e., from a homework problem to reactor safety analysis), much of the evidence is based on the evidence that has been historically used for CBT models applied in reactor safety analysis.

The purpose of the framework is summarized as the main goal, *G - The CBT model can be trusted.* Everything which follows is focused on demonstrating that this main goal is true and defines exactly what is meant by the statement "*The CBT model can be trusted*". The main goal is decomposed into the three subgoals in Figure 3 below.

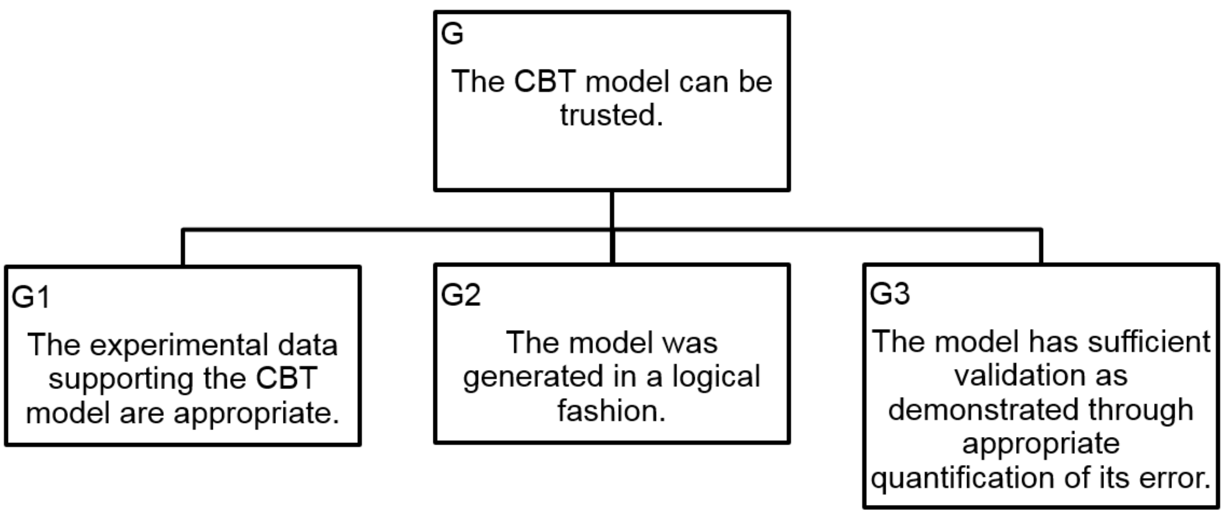


Figure 3 Decomposition of G — Main Goal

As discussed above, the goals (G, G1, G2, G3) are intentionally ambiguous. While there may be no consensus on what is meant by the words "trusted," "appropriate," logical", and "sufficient," most will agree that for a CBT model to be trusted, its experimental data must be appropriate, the model must be logical, and the validation must be sufficient. The further development of the framework through continued decomposition of each goal into sub-goals and specification of the possible levels of evidence, acts to more clearly define these ambiguous terms.

The bulk of this section will focus on the decomposition of all sub-goals into base goals.[13] For each base goal, we provide a discussion of the levels of evidence used for demonstrating that the base goals are true and a discussion of the evidence levels that have been historically used for CBT models in reactor safety analysis.

[13] A goal that is not decomposed further but is supported by evidence.

3.1 G1—Experimental Data

Experimental data are the cornerstone of a CBT model. The data are used to generate the coefficients of the model and validate the model. Additionally, previous experimental data often used influence the form of the model. Therefore, it is essential that experimental data are appropriate. The three subgoals in Figure 4 are used to demonstrate that the experimental data are appropriate.

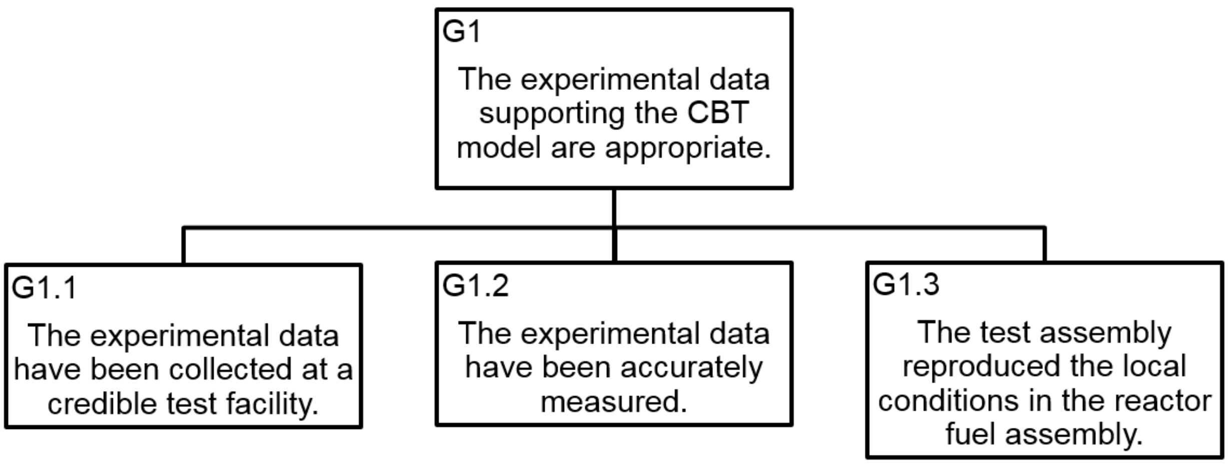


Figure 4 Decomposition of G1—Experimental Data

3.1.1 G1.1—Credible Test Facility

Test facilities that are used to measure CBT primarily focus on measuring key flow parameters that occur during the critical transition. Experimental data has been collected at multiple research facilities and universities over many years (Groeneveld 2007). However, because the time, effort, and resources needed to set up a reliable facility are quite significant, most CBT data used in the nuclear industry have historically come from one of the following facilities:

- Columbia University's Heat Transfer Research Facility (closed in 2003)
- General Electric Company's ATLAS test loop facility in San Jose, CA (closed)
- Stern Laboratories in Hamilton, Ontario (still in use)
- AREVA's KATHY loop in Karlstein, Germany (still in use)
- Westinghouse Electric Corporation's FRIGG and ODEN loops in Västerås, Sweden, for BWRs and PWRs, respectively (still in use)

The two subgoals in Figure 5 are used to demonstrate the credibility of the test facility.

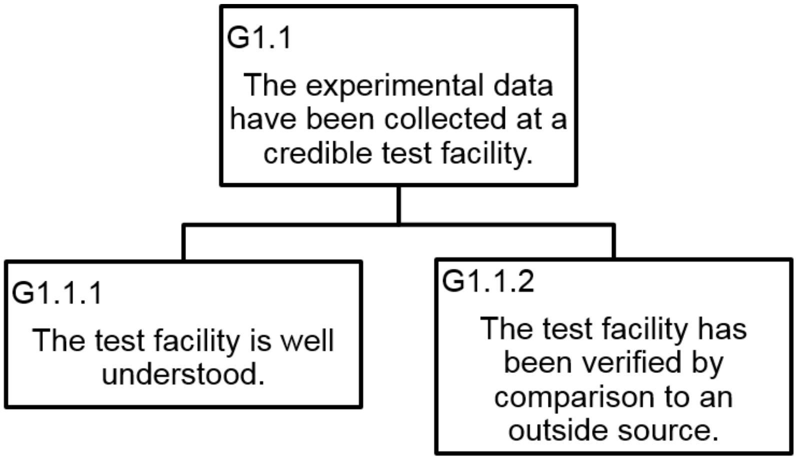


Figure 5 Decomposition of G1.1—Credible Test Facility

No further decompositions of the subgoals were deemed useful. Therefore, the sections below discuss the evidence that could be used to demonstrate that these two base goals (G1.1.1 and G1.1.2) have been satisfied. Additionally, a discussion is provided on the evidence that has been historically used for CBT models applied in reactor safety analysis.

3.1.1.1 G1.1.1—Test Facility Description

The test facility contains the test loop, the control equipment, interconnected piping, and instrumentation needed to perform the experiment. Test loops usually consist of a test section (which contains the simulated test assembly), pressurizer, heat exchangers, pumps, pressure transducers (both absolute and differential), flow meters, and thermocouples. The test assembly contains the simulated fuel rods, which not only supply the power to the test section but also contain the thermocouples that indicate when a CBT occurs.

The description of the test facility must enable the assessor to understand how the facility operates and how the data were obtained. For assessors familiar with CBT testing and for established test facilities, a reference that describes the facility is typically sufficient documentation. In the past, having the assessor visit the test facility and witness testing first hand has greatly increased the assessor's understanding, reducing the total time needed for the assessment, particularly for new assessors and/or new test facilitates. Table 6 gives the evidence commonly provided to demonstrate that this goal has been satisfied.

Table 6 Evidence for G1.1.1—Test Facility Description

G1.1.1	The test facility is well understood.
Level	Evidence
1	A reference that describes the test facility in appropriate detail has been provided. At a minimum, the reference includes loop, test section, and heater rod descriptions.
2	The assessors have visited the test facility. Additionally, a reference that describes the test facility in appropriate detail has been provided.

	At a minimum, the reference includes loop, test section, and heater rod descriptions.

Historical Evidence Levels for Reactor Safety Analysis

Level 1 has been most commonly accepted by the NRC staff, but Level 2 has resulted in increased review efficiency. Because the goal of the reference describing the test facility is to allow the assessor to fully understand the function of the test facility including operation, control, and measurement capabilities, it has often been found to be convenient to have the assessor visit the test facility and witness testing. This is especially true for new assessors unfamiliar with a test facility, but also true for experienced assessors who have not reviewed data from a particular test facility for some period of time. Visiting a test facility and observing testing has been a much more efficient way for the assessor to gain an understanding of the test facility than by reading documentation alone. A significant portion of an assessor's time is spent gaining an understanding of the test facility. The assessor must understand the facility to such an extent that he or she is able to fully understand a complete test run including how the various pieces of equipment interact. Thus, actually visiting the test facility greatly increases the rate of understanding, typically leading to a reduction in the time needed to perform the assessment and fewer questions.

3.1.1.2 G1.1.2—Test Facility Comparison

The test facility description is used as an indicator to determine if the facility is capable of generating accurate data. However, another key piece of evidence is the validation of the test facility itself. One type of validation frequently used is a comparison of the measured CBT data to the results from another credible facility. The justification for the test facilities should be based on factors other than the test facility itself (e.g., comparison to a benchmark, reproduction of data from another facility, or reproduction of known phenomena).

Most facilities in use today have been compared to their older counterparts (for example, many facilities have performed tests to compare to data collected at Columbia University). However, because of the proprietary nature of the test sections, it may be difficult to obtain comparisons to actual CBT data. Therefore, though a new facility would be under the greatest scrutiny in this framework, it may have difficulty meeting this criterion. When comparisons to actual CBT measurements are not possible, the assessor should compare the test facility under evaluation to measured quantities from other experiments (e.g., in the open literature) with similar phenomena. Table 7 gives the evidence commonly provided to demonstrate that this goal has been satisfied.

Table 7 Evidence for G1.1.2—Test Facility Comparison

G1.1.2	The test facility has been verified by comparison to an outside source.
Level	**Evidence**
1	The test facility has been verified by comparison of data obtained at the facility to some benchmarks or some known phenomenological behavior.
2	The test facility has been verified by comparison of data obtained from tests at the facility to data other than CBT data obtained from a credible facility.
3	The test facility has been verified by comparison of CBT data obtained at the facility to CBT data obtained from a credible facility.
4	The test facility has been verified by comparison of CBT data obtained at the facility to CBT data obtained over the same application domain as that of the proposed model at a credible facility.

Historical Evidence Levels for Reactor Safety Analysis

Evidence at Level 2 and Level 3 have been most commonly accepted by the NRC staff. This is largely due to the fact that most test facilitates in operation today are 2nd generation facilitates, and part of their initial testing program was to establish consistency with the data taken from 1st generation facilities. When comparisons to actual CBT measurements are not possible, it is possible for the assessor to consider other measured quantities besides CBT from other experiments with similar phenomena.

3.1.2 G1.2—Accurate Measurements

In order for the test data to be relied upon, the test facility needs to provide accurate measurements of all important experimental parameters, including the measurement of CHF or CP. It is important to note that neither CHF nor CP is a directly measured parameter (like flow rate or pressure); instead, the CHF or CP value is inferred from the assembly power, axial and radial power peaking, and a thermocouple indication that signifies CBT has occurred in the test facility and where in the test section CBT has occurred.

Typically, five experimental parameters are directly measured or controlled.[14] The type of control used for the experimental parameters depends on the type of data being taken. Usually, the desired values are programmed into a computer, and the computer will maneuver the control equipment to the desired state point. Table 8 presents the methods used to measure and control each experimental parameter.

[14] Although the axial heat flux shape is very important for obtaining the local power and may be changed through the exchange of test rods, it is not a measured value during the experiment and, therefore, will be treated in Section 3.1.3 on local conditions.

Table 8 Experimental Parameters Measured or Controlled

Parameter	Method of Measurement	Typical Method of Control
Pressure	Absolute and differential pressure cells on the test section	A pressurizer on the test loop
Power (including radial power peaking)	Reading from rectifiers	Rectifiers that supply power to the simulated fuel rods
Inlet Flow Rate	Flow meter at the inlet	Valve at the inlet or pump speed
Inlet Temperature	Thermocouple at the inlet	Heat exchanger or mixer at the inlet
Rod Temperature Change	Thermocouples inside the simulated fuel rods	N/A (the change in rod temperature is not controlled, but is a response quantity)

The six subgoals in Figure 6 are used to demonstrate the accuracy of the measurements.

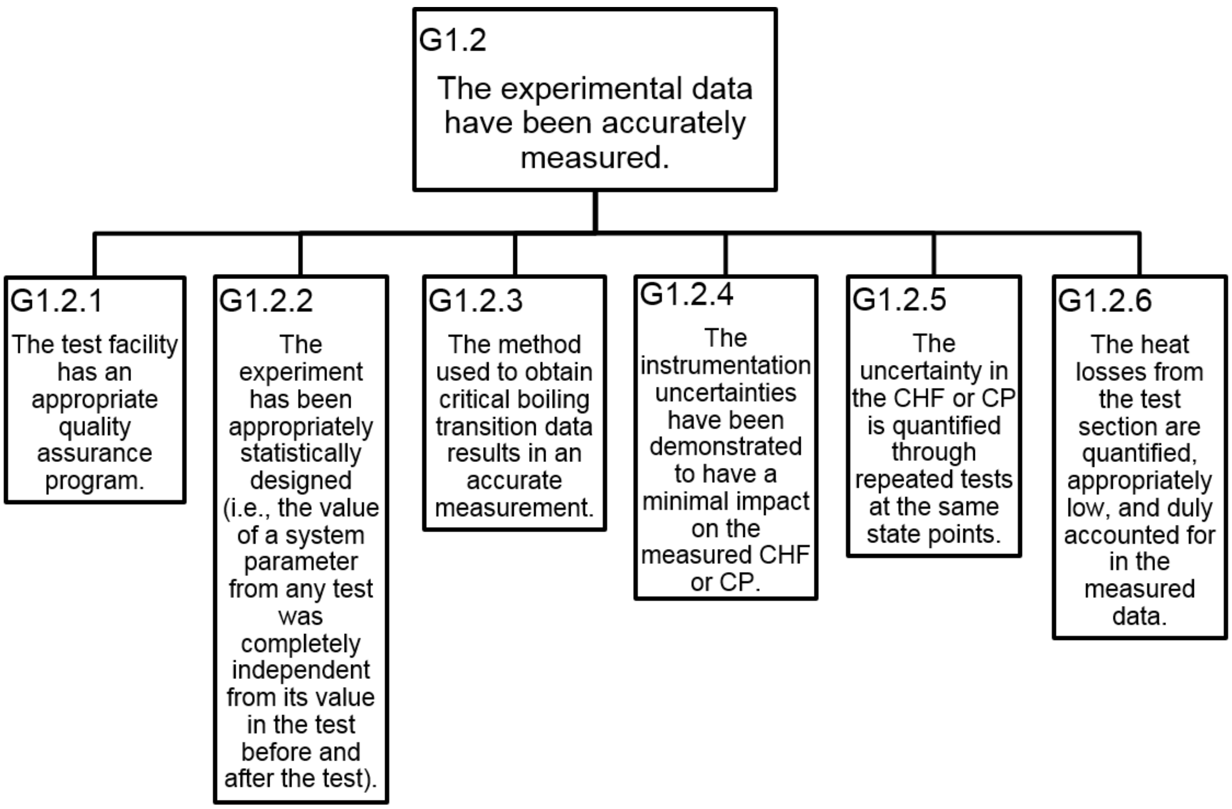


Figure 6 Decomposition of G1.2—Accurate Measurements

No further decompositions of the subgoals were deemed useful. Therefore, the sections below discuss the evidence that could be used to demonstrate that these six base goals have been satisfied. Additionally, a discussion is provided on the evidence that has been historically used for CBT models applied in reactor safety analysis.

3.1.2.1 G1.2.1— Test Facility Quality Assurance (QA) Program

A determination regarding the credibility of a test facility is often assessed by reviewing the quality assurance program applicable to the test facility. Typically, an assessment of a facility's QA program involves determining its compliance with a standard (e.g., ASME's NQA-1, "Quality Assurance Requirements for Nuclear Facility Applications). While different QA standards will have different elements, the following represent some of the issues that should be addressed:

- **Calibrated instrumentation** - Routine calibration of the instrumentation is necessary to ensure that an instrument is resulting in a precise measurement and to quantify any instrumentation error (i.e., accuracy and precision). Generally, the instrumentation's calibration is checked on a routine basis, with the calibration interval set to account for instrument drift over time and drift due to operation. This check should be performed often enough to avoid having to recalibrate the instrumentation after its use. If an instrument does need to be recalibrated after a test, it likely means that the last set of data points taken with that instrument were taken when the instrument was out of calibration. At a minimum, a calibration check should be performed at both the beginning and end of a test campaign. The general assumption is that, if an instrument is within its calibration specification at the beginning and end of a campaign, there is very little chance that it was out of its specification at any time during the campaign. Note that, contrary to the discussion above, the heater rod thermocouples used to detect a CBT are often not calibrated because the absolute value of the temperature is not used. Instead, as previously discussed, a change in temperature over a period of time is used as the criterion for determining that a CBT has occurred. However, the thermocouples used to determine fluid and wall temperatures elsewhere in the test loop should be calibrated. NQA-1, Requirement 12 "Control of Measuring and Test Equipment" provides more details on instrument calibration.

- **Appropriate equipment** – The experimental parameters measured in CBT experiments are provided in Table 8 above. Therefore, instrumentation should be employed to measure these parameters. However, as instrumentation may fail or provide anomalous measurements, a common practice is to employ redundant and diverse instrumentation. Redundant instrumentation is necessary to ensure that (1) instrumentation remains in calibration, and (2) an instrument which suddenly becomes uncalibrated does not greatly impact the resulting experimental data. Further, diverse instrumentation (i.e., use of a different process to perform the measurement) helps achieve a higher degree of confidence that the final measurement is accurate because it reduces the potential for "common cause" failures that could result in inaccurate measurements.

- **Trained personnel** – There are many appropriate ways in which the data could be obtained. It is important that the personnel performing the tests have been trained on the test procedure and test equipment, and are able to follow the test procedure in order to ensure consistent experimental results.

- **Condition of test equipment and the item to be tested** – The test equipment, including the instrumentation, the test section, and all connected piping, should be demonstrated to be in working order. Generally, the bulk of these activities is performed during the shakedown testing, which ensures the test facility is behaving as expected.

- **Suitable environmental conditions** – As CBT tests are often performed in state of the art experimental test facilities, the conditions for both the equipment and the personnel are generally suitable environments.

- **Provisions for data acquisition** – As the data will be used to validate the CBT model, the acquisition of the data are of paramount importance. While there are multiple data acquisition systems that could be used, it is important for specific procedures to be developed and used in order to determine how the data are reduced to the final set of measured values.

Table 9 gives the evidence commonly provided to demonstrate that this goal has been satisfied.

Table 9 Evidence for G1.2.1—Test Facility QA Program

G1.2.1	The test facility has an appropriate quality assurance program.
Level	Evidence
1	A QA program exists that reflects the basic tenets of quality assurance as referenced by a widely accepted international quality organization (e.g., NQA-1).
2	A QA program exists that reflects the basic tenets of quality assurance as referenced by a widely accepted international quality organization (e.g., NQA-1). Documentation is provided that outlines how the design, construction, and test activities were conducted consistent with the QA Program. It is clear that the base expectations of QA were applied.
3	A QA program exists that reflects the basic tenets of quality assurance as referenced by a widely accepted international quality organization (e.g., NQA-1). Documentation is provided that outlines how the design, construction, and test activities were conducted consistent with the QA Program. It is clear that the base expectations of QA were applied. Audit reports properly identify, track, and indicate correction of conditions adverse to quality and are available for inspection.

<u>Historical Evidence Levels for Reactor Safety Analysis</u>

Level 3 has been most commonly accepted by the NRC staff. While the CBT assessor does not typically examine the QA program in the same detail as a QA inspector, previous NRC reviews have shown that understanding the QA program helped the assessor gain an improved understanding of how the data were taken, controlled, reduced, and then used to generate the model. It is important to note that most assessors have typically limited their review to confirming that some type of QA program was in place, rather than providing an extensive review of that program itself.

3.1.2.2 G1.2.2—Statistical Design of Experiment

The goal of the statistical design of the experiment (Box, Hunter, and Hunter, 1978) is to ensure that the testing methods do not introduce any biases into the figure of merit (i.e., the CHF or CP value). Most of the statistical methods used to quantify the uncertainty treat all errors as random. This is equivalent to assuming that each measurement is taken at a randomly determined experimental state point[15] that is completely independent of any measurements taken before or after. However, that is generally not possible for CBT experiments. First, large changes in pressure (and sometimes flow rate) can put tremendous stresses on the test section. Second, changes in flow cause the test section to reach a new thermal equilibrium, which may take a long time. As such, it is often not feasible to dramatically change the flow rate or pressure between test points. Because of these issues, the order in which the test points are taken is typically not random. Table 10 gives the evidence commonly provided to demonstrate that this goal has been satisfied.

Table 10 Evidence for G1.2.2—Statistical Design of Experiment

G1.2.2	The experiment has been appropriately statistically designed (i.e., the value of a system parameter from any test was completely independent from its value in the test before and after the test).
Level	Evidence
1	One or more system parameters were randomized, but no consideration was given to other system parameters.
2	One or more system parameters were randomized, and some consideration was given to all other system parameters.
3	One or more system parameters were randomized, and those parameters that were not randomized between tests were randomized in larger test blocks.
4	All system parameters were completely randomized.

Historical Evidence Levels for Reactor Safety Analysis

Level 3 has been most commonly accepted by the NRC staff. In general, the design of the experiment attempts to randomize the system parameters as much as possible between each test. Since testing is often split into groups (e.g., a set of tests at a single pressure and/or flow rate), parameters are often randomized between test groups. For example, if the pressure were held constant during a group of tests, then the pressures from group to group should be randomized. As much as possible, flow rates are also randomized for a fixed pressure. Because randomization (i.e., independence) is a key assumption in all of the statistics performed on the data and because it is generally not possible to guarantee randomization through the design of the experiment, repeated test points have become a vital part of demonstrating that there are no biases in the test facility.

[15] By state point, we mean the value of each variable that completely determines the state of the system.

3.1.2.3 G1.2.3—Data Fidelity

The method used to obtain CBT data should result in an accurate measurement of CBT. There are typically two different types of tests used in CBT experiments: (1) those used to obtain steady-state data and (2) those used to obtain transient data. It is vital that assessors understand exactly what is occurring in each of these tests. Therefore a careful evaluation of the test constraints, input assumptions, and expected result ranges should be employed.

<u>Measuring a Steady-State Data Point</u>

For steady-state data, the objective is to determine the state point at which CBT occurs. A state point is a coordinate in an n-dimensional space defined by all of the parameters which make up the system. In general, there are two main types of state point: experimental state points and model state points. For an experimental state point, the parameters of interest are those parameters that influence the overall experiment (e.g., system pressure, total power (including radial and axial peaking), inlet flow rate, and inlet temperature). For a model state point, the parameters of interest are those parameters needed by the model to make a prediction of CHF or CP. Depending on the model itself, these generally include global as well as local parameters as well as parameters that are not measured in the experiment (e.g., local mass flux, local quality). For PWRs, the values of parameters that are not measured in the experiment are obtained using a subchannel code that predicts the local flow behavior in the subchannels using the experimental parameters as boundary conditions. In a sense, the subchannel code used can be thought of as the means by which the experimental state point is transformed into a model state point.

It is important to note that a reactor almost never operates at a steady state, especially during an AOO. Because the models are based on steady-state data, the model effectively treats each AOO as if it were made up of a multitude of steady-state state points and determines the heat flux or assembly power that causes a CBT at those individual state points. Multiple previous applications of steady-state models have been demonstrated to be conservative (i.e., a model developed with steady-state data will generally underpredict the heat flux or assembly power that causes a CBT), and it is common for analysts to provide data demonstrating that this conservative assumption remains true for each individual CBT model.

The following standard procedure is used to measure a steady-state data point:

(1) An experimental state point is chosen. As previously discussed, a single value of pressure, power, power shape, inlet flow rate, and inlet temperature is generally chosen. Usually, the initial power is chosen to be somewhat lower than that expected to cause a CBT.

(2) The experimental facility is driven to the state point. Generally, a computer operates the control system to allow for finer control.

(3) Once the initial state point is reached, power is slowly increased while maintaining steady conditions on the other experimental parameters. Some variation in the values of the experimental parameters will exist, but this variation should be kept small and should be accounted for in test procedures. Although steady-state CBT data could be obtained by varying any one of the experimental parameters in an appropriate direction while keeping the others constant (e.g., decreasing the flow rate), such data are usually obtained by slowly increasing the power.

(4) As the power is slowly increased, the rod internal thermocouples are monitored. A CBT is assumed to have occurred if the temperature indicated by one of the thermocouples increases by a specified amount over a specified small period of time or if some maximum temperature is reached.

(5) Once a CBT occurs, power is reduced and the values of the parameters that make up the experimental state point are written to a file. These data, along with the known axial and radial power shape, can then be used to calculate either the CHF or the CP.

Measuring a Transient Data Point

The objective for transient data are to determine the lowest power level at which a specific transient will cause a CBT. In this case, a specific transient is defined through specified time-dependent functions for each experimental parameter. Typically, a computer controls the experimental parameters to ensure that the test achieves the desired behavior of the time-dependent functions. However, not all experimental parameters will vary during the transient (e.g., pressure is almost never varied because of the strain this would place on the test loop). In this sense, steady state can be considered a special type of transient where all time-dependent functions are held constant.

It is also important to note that each AOO is not directly mapped to a specific transient test. Although some AOOs can be mapped into a transient test (e.g., loss of flow), this is not possible with all AOOs. AOOs involving rapid changes in pressure are especially challenging because any rapid change in pressure in the test loop could put the loop at risk. Therefore, additional analysis is usually performed to determine how the transient testing bounds the AOOs.

One of the similarities between transient and steady-state testing is the objective of the test. In each case, the objective is to determine the minimum power at which a CBT will occur for some set of initial and boundary conditions. It is important that the focus is on obtaining the *minimum* power at which a CBT occurs under some set of conditions. Simply finding any power which causes a CBT is not useful as one can always be caused by any sufficiently high power. For example, every conceivable transient will result in a CBT at Graham's number[16] of watts, or even 10^{100} watts (*much* smaller than Graham's number of watts). This does not mean that CBT will occur only at a power of Graham's number because it will obviously occur at much lower powers. Therefore, the objective is to determine the *minimum* power at which a CBT occurs for those initial and boundary conditions. Thus, if those conditions (either steady state or transient) occur in a reactor and if that minimum power is not reached, a CBT would not occur.

The following standard procedure is commonly used to measure a transient data point:

(1) A specific transient is chosen. As previously discussed, time-dependent functions of pressure, power, inlet flow rate, and inlet temperature are generally chosen.

(2) The experimental facility is driven to the state point. Generally, a computer operates the control system to allow for finer control.

[16] Graham's number is one of the largest numbers known in mathematics. It is many orders of magnitude larger than the total number of particles in the observable universe.

(3) Once the initial condition state point is reached, the transient is started. The values of experimental parameters are defined as time-dependent functions that are controlled to within their desired magnitude by the control system.

(4) The rod internal thermocouples are monitored during the transient. A CBT is assumed to have happened if the temperature indicated by one of the thermocouples increases by a specified amount over a specified small period of time.

(5) The magnitude of the initial power can be either increased or decreased, and the same transient can be run again to determine the minimum power at which a CBT occurs. Frequently, the same transient is performed multiple times to determine the *minimum* power.

(6) Once the *minimum* power at which a CBT occurs is known, power is reduced, and the values of parameters that make up the experimental state point are written to a file. These data, along with the known axial and radial power shape, can then be used to calculate either the CHF or the CP for that transient.

Table 11 gives the evidence commonly provided to demonstrate that this goal has been satisfied.

Table 11 Evidence for G1.2.3—Data Fidelity

G1.2.3	The method used to obtain critical boiling transition data results in an accurate measurement.
Level	Evidence
1	A reference has been provided that describes the method used to obtain results from both steady-state and transient tests.
2	A reference has been provided that describes the method used to obtain both steady-state and transient tests. The assessors have examined the reference and believe that it will result in accurate measurements of the CBT for both steady-state and transient tests.
3	A reference has been provided that describes the method used to obtain both steady-state and transient tests. The assessors have examined the reference and believe that it will result in accurate measurements of the CBT for both steady-state and transient tests. Additionally, the assessors have observed the method in practice.

Historical Evidence Levels for Reactor Safety Analysis

Levels 2 and 3 have been most commonly accepted by the NRC staff. An accurate measurement of CBT has three main focuses: (1) Ensuring the state point (i.e., pressure, mass flux, inlet subcooling, power) has been measured and maintained during the entire test run within some small uncertainty; (2) Ensuring that any CBT that would occur is captured in the data; (3) Ensuring that the power at which CBT was recorded was the lowest power that would cause a CBT at that state point. A large part of the review process is spent in gaining an understanding of how the data are taken, reduced, and then used to generate the model. To that end, observing the experiment has been one of the most efficient ways to gain this information.

3.1.2.4 G1.2.4—Instrumentation Uncertainty Impact

Accurate measurements are vital to the success of any experimental program. Therefore, the flow rates, temperatures, pressures, and powers must be measured accurately and precisely, and their associated instrumentation uncertainty must be kept low. Typically, the model's uncertainty does not directly account for instrumentation uncertainties; instead, such uncertainties are treated as part of the randomness of the data. If those uncertainties are reasonably low over the range for which the measurements are taken, this assumption is generally valid. Table 12 gives the evidence commonly provided to demonstrate that this goal has been satisfied.

Table 12 Evidence for G1.2.4—Instrumentation Uncertainty Impact

G1.2.4	The instrumentation uncertainties have been demonstrated to have a minimal impact on the measured CHF or CP.
Level	Evidence
1	The instrumentation uncertainties have been quantified.
2	The instrumentation uncertainties have been quantified and an analysis is used to demonstrate that the uncertainties result in a minimal impact on the measured CHF or CP. OR The instrumentation uncertainties have not been quantified, but repeated test points allow those uncertainties to be captured directly in the CHF or CP value.
3	The instrumentation uncertainties have been quantified and an analysis is used to demonstrate that the uncertainties result in a minimal impact on the measured CHF or CP. This has further been demonstrated by experiments (e.g., repeated test points).

<u>Historical Evidence Levels for Reactor Safety Analysis</u>

Level 3 has been the most commonly accepted by the NRC. While a quantitative analysis of the instrumentation uncertainties on the measured CHF or CP values is possible, it is often more complicated than simply taking additional data points to measure the uncertainty directly. While such an analysis does assume that the instrumentation's uncertainty remains constant over the course of the test, this can usually be confirmed by performing an additional test at the same state point to generate a repeat test point.

3.1.2.5 G1.2.5—Repeated Test Points

The instrumentation uncertainty may be obtained from the instrumentation manufacturer or during calibration. However, the uncertainty on the "measured" CHF or CP at the location of interest (i.e., the experimental uncertainty) cannot be obtained so easily. This uncertainty is a combination of the instrument uncertainty; uncertainties of other input parameters (e.g., axial power shape, selection of the subchannel of interest); and the method used to combine all of the parameters to generate a "measured" CHF or CP at the location of interest.

Because the CHF or CP at the location of interest cannot be directly measured, the experimental uncertainty should be determined by obtaining a "measurement" of CHF or CP at the experimental state point multiple times over the entire test cycle and analyzing the variability in the results. Some variation in the input parameters will occur because obtaining the exact same experimental state point (i.e., pressure, flow rate, and inlet subcooling) is not possible, but this variability should be small compared to the uncertainty in the measured CHF or CP value. A number of repeated test points should be taken at multiple experimental state points and at various times during the test campaign to ensure that the behavior of the test facility has not changed and to provide a quantitative estimate of the uncertainty in the "measured" CHF or CP. The variability in the resulting CHF or CP values should be much lower than the quantified uncertainty of the model. If it is not, this is evidence that there is an error in determining the model's uncertainty. Table 13 gives the evidence commonly provided to demonstrate that this goal has been satisfied.

Table 13 Evidence for G1.2.5—Repeated Test Points

G1.2.5	The uncertainty in the CHF or CP is quantified through repeated tests at the same state points.
Level	Evidence
1	No repeat test points have been taken.
2	One repeat test point was taken over the test campaign. The variability in the resulting CHF or CP value was reasonably low.
3	Multiple repeat test points were taken over the test campaign at various input parameters. The variability in the resulting CHF or CP values was reasonably low.

Historical Evidence Levels for Reactor Safety Analysis

Level 2 and Level 3 have been most commonly accepted by the NRC staff. Aside from satisfying this goal (G.1.2.5), multiple repeat test points (Level 3) can also be used as evidence that the behavior of the test assembly remains consistent over the time frame of the test. The repeated test points may become much more important if other aspects of the behavior of the test assembly are called into question. For example, if there is a geometry change during testing, then the impact of that change could be determined to be minimal if there are an adequate number of repeated test points. The variability from repeat test points is typically small compared to the uncertainty of the CBT model. Additionally, due to the limitations on the statistical design of the experiment, multiple repeat test points are one way to provide evidence that the errors are indeed random and that each experimental state point can be considered independent of every other state point.

3.1.2.6 G1.2.6—Quantified Heat Losses

Along with accurate flow, pressure, temperature, and power measurements, the test section heat losses should also be quantified. Because the CHF or CP is obtained from the power measurement, ignoring the heat losses would result in a "measured" CHF or CP higher than the "actual" CHF or CP value by the amount of heat loss. This would result in a non-conservative measurement.

Typically, test section heat losses are kept very low through active means. In many cases, the test section may sit in a heated water bath to ensure minimum heat loss through the walls. Generally, while the absolute value of the test section heat losses to the surroundings increases as the test assembly power increases, the percentage of the heat losses relative to the test assembly power actually decreases (i.e., the fraction of heat dissipated in the fluid in the test section is lower for higher powered tests). Therefore, the bounding heat losses are generally quantified through a test conducted at a low assembly power. The assessor needs to establish whether the measured CHF or CP data were corrected for the heat losses before the development of the CHF or CP model. If not, the assessor should consider the inherent non-conservatism. Table 14 gives the evidence commonly provided to demonstrate that this goal has been satisfied.

Table 14 Evidence for G1.2.6—Quantified Heat Losses

G1.2.6	The heat losses from the test section are quantified, appropriately low, and duly accounted for in the measured data.
Level	Evidence
1	Heat losses have been quantified and are minimal, but they have not been removed from the power used to calculate the CHF or CP.
2	Heat losses have been quantified and have been removed from the power used to calculate the CHF or CP.

Historical Evidence Levels for Reactor Safety Analysis

Level 1 has been most commonly accepted by the NRC staff. Generally, the percentage of heat loss is calculated for each test. The percentage of heat loss is usually estimated to be greater than that of the actual heat loss measured, but it should still be very low compared to the overall power. Overestimating the heat loss is conservative for the reason given above. While it is generally desirable to minimize heat losses from the test section, it is not strictly necessary as long as the heat losses are measured and accounted for the in power measurement.

3.1.3 G1.3—Reproduction of Local Conditions

The local conditions in the reactor fuel assembly should be reproduced in the test assembly to ensure that experimental data taken in the laboratory apply to the reactor fuel assembly placed in the reactor. The five subgoals in Figure 7 are used to demonstrate the reproduction of local conditions.

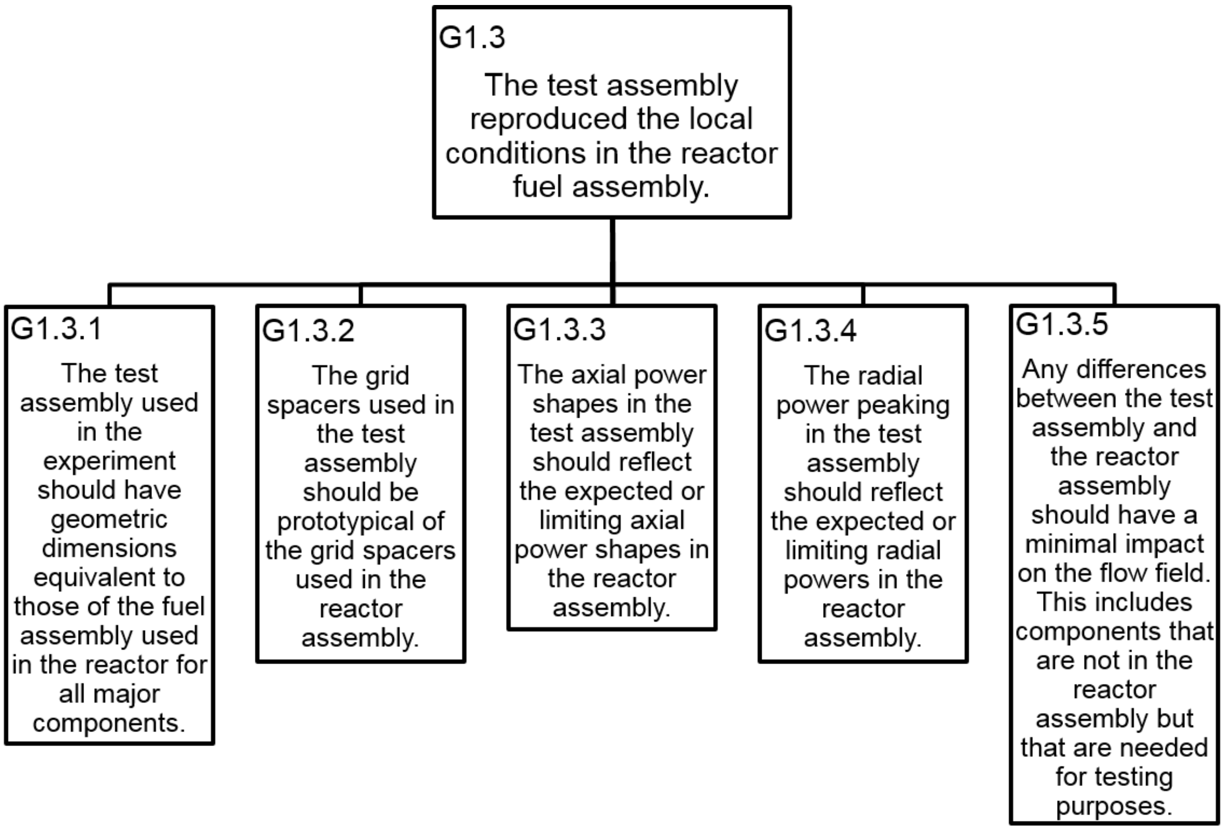


Figure 7 Decomposition of G1.3—Reproduction of Local Conditions

No further decompositions of the subgoals were deemed useful. Therefore, the sections below discuss the evidence that could be used to demonstrate that these five base goals have been satisfied. Additionally, a discussion is provided on the evidence that has been historically used for CBT models applied in reactor safety analysis.

3.1.3.1 G1.3.1—Equivalent Geometric Dimensions

The test assembly provides the structure in which the flow field will be established. The flow field details, many of which will not be measured or directly reproduced in the computer simulation, will directly affect the CBT. Therefore, the flow field in the test assembly should be as similar as possible to the flow field in the reactor fuel assembly.

To ensure a similar flow field, the test assembly is manufactured as a prototypical fuel assembly. This includes the fuel rod pitch and diameter, guide tube rod location and diameter, part-length rod height and axial and radial locations, flow areas, number of grid spacers, distances between grid spacers, grid spacer heights relative to the bottom of the fuel assembly, and total assembly height. Each of these dimensions should be within the design tolerances of the reactor assemblies.

For BWRs, the test assembly is typically full size (e.g., 8x8, 9x9, 10x10) or symmetric (5x5). However, for PWRs, a full-size assembly (e.g., 15x15, 17x17) would require a substantial amount of power. Therefore, smaller 5x5 or 6x6 test assemblies are used. In the early days of CBT testing, 4x4 or smaller assemblies were used, but the unheated channel wall surrounding the test assembly had too large an effect on the interior subchannels. Therefore, 4x4 (and smaller) assemblies are considered too small to provide an adequate representation.[17]

Note that heater rods are potentially subject to large electromagnetic forces caused by the current flowing through them. These forces must be countered or the rods will bend and the subchannel flow area will change during testing. In indirectly heated rods, the direction of the current in adjacent rods can be reversed to counter the electromagnetic forces. However, this is not possible in directly heated rods because the electric potential must be the same in all rods at each grid spacer. Therefore, in order to maintain the sub-channel size in directly heated rod bundles, simple support grids are commonly used. These grids provide structural support and are designed to have minimal impact on the flow field. Often, the grids are only needed in sections of the assembly where there are large spans between mixing vane grids. Table 15 gives the evidence commonly provided to demonstrate that this goal has been satisfied.

Table 15 Evidence for G1.3.1—Equivalent Geometric Dimensions

G1.3.1	The test assembly used in the experiment should have geometric dimensions equivalent to those of the fuel assembly used in the reactor for all major components.
Level	Evidence
1	Many of the components in the test assembly have geometric dimensions equivalent to those of fuel assemblies used in reactors and are within the design tolerance of the fuel assemblies that will be used in the reactor. Any components that do not have equivalent geometric dimensions have dimensions that would result in a conservatively lower prediction of the power or heat flux that causes a CBT.
2	The vast majority of the components in the test assembly have equivalent geometric dimensions that are within the design tolerance of the fuel assemblies that will be used in the reactor. The few components that do not have equivalent geometric dimensions would have a minimal impact on CBT measurements.
3	All components in the test assembly have equivalent geometric dimensions that are within the design tolerance of the fuel assemblies that will be used in the reactor.

[17] This is *not* referred to as the “cold-wall effect,” even though it is due to the impact of the outer cold wall. The term “cold-wall effect” is reserved for the effect of control rod guide tubes and instrument tubes on CHF performance.

Historical Evidence Levels for Reactor Safety Analysis

Level 2 has been most commonly accepted by the NRC staff. For some older CBT models, the heated length was varied to cover a wider range of fuel. While this could be understand to be level 1, it would strongly depend on the importance of the heated length in the CBT model.

While there may be instances in which a CBT model may only achieve Level 1, the demonstration that level 1 is acceptable is challenging as it is difficult to prove that the CBT model would produce conservative predictions under all conditions.

3.1.3.2 G1.3.2—Prototypical Grid Spacers

One of the most important parts of the prototypical assembly is the grid spacer. The spacers ensure that the rods maintain the same pitch as the assembly used in the reactor. The spacers are also the major source of turbulence which acts to increase the heat transfer from the fuel rods. Grid spacers are specifically designed to increase the power or heat flux at which a CBT occurs. In BWRs, the grid spacer is typically designed to increase deposition by directing more of the water droplets entrained in the vapor flow back onto the liquid film. Great care is taken to ensure that the liquid film is not separated (stripped) from the fuel rod in the vicinity of the grid spacer. In PWRs, the grid spacer is typically designed to strip the bubble layer from near the fuel rod surface to reduce bubble crowding and to enhance turbulence and mixing in the subchannel.

Arguably, the design of the grid spacer will have a larger impact on the CBT than any other input parameter. The grid spacers increase the margin to CBT through their increase in turbulence or increase in deposition on the fuel rod. However, the current generation of the computer simulations that make use of CBT models do not directly simulate the impact of the spacers; therefore, the CHF or CP model must capture the spacers' impact. The number of mixing vanes, the shape of the vanes, the location of the vanes in the subchannel, the surface area of the vanes, the angle of the vanes, and the direction of swirl caused by the vanes can all affect the thermal mixing in the fuel assembly subchannel. Therefore, it is vital that the grid spacer used in the test assembly is prototypical when compared to the grid spacer used in the reactor core.

Unfortunately, it is not always possible to use prototypical grid spacers. Therefore, if such grid spacers cannot be used, the grid spacers used in the test section should result in conservative behavior compared to the grid spacers in the reactor core. However, it is very difficult to prove that the one grid spacer will result in conservative behavior under all conditions when compared with another grid spacer. Therefore, demonstrating conservative behavior can be a challenge.

Additionally, fuel assemblies may be comprised of different grid spacer types at different axial elevations. Therefore, the same grid types should appear in the test assembly and at the same elevations as in reactor fuel. Table 16 gives the evidence commonly provided to demonstrate that this goal has been satisfied.

Table 16 Evidence for G1.3.2—Prototypical Grid Spacers

G1.3.2	The grid spacers used in the test assembly should be prototypical of the grid spacers used in the reactor assembly.
Level	Evidence
1	The grid spacers used in the test assembly will result in a conservative under-prediction of the true thermal mixing caused by the grid spacers in the reactor assembly.
2	The grid spacers used are very similar to those that will be used in the reactor assembly but with some slight differences.
3	The grid spacers used are identical to those that will be used in the reactor assembly except for the number of rods (e.g., a 6x6 cutout of a 17x17 assembly).
4	The grid spacers used are identical to those that will be used in the reactor assembly (either identical in size or a symmetric cut of the grid spacer).

Historical Evidence Levels for Reactor Safety Analysis

Level 3 has been most commonly accepted by the NRC staff for PWRs, and Level 4 has been most commonly accepted for BWRs. PWRs typically operate at a higher linear power density, have more rods per assembly, and have fewer assemblies per core. Therefore, it is impractical to test an entire PWR assembly in a test facility because the power needed would be too high. Additionally, PWR methods use a true subchannel analysis and, therefore, model the grid spacers' impact on the local fluid quantities. On the other hand, BWR methods use a simplified subchannel analysis that considers only assembly-averaged flow parameters and, therefore, calls for experimental details on every fuel rod in the assembly. For this reason, BWR tests use full-sized assemblies or representative symmetric sub-assemblies.

Levels 1 and 2 are not common in reactor safety analyses, as even small changes in the grid spacer can have major impacts to the flow field.

3.1.3.3 G1.3.3—Axial Power Shapes

It is important to reproduce the local powers created by the reactor assembly in the test assembly. This is generally done by testing combinations of axial and radial power shapes. Although the fuel rods in the reactor can take on an almost infinite number of axial power shapes, generally only three shapes (cosine, up-skew, and down-skew) are used in testing for BWR models, and three shapes (uniform, cosine, and up-skew) are used in testing for PWR models. Additionally, because of the current experimental designs, the only way to change the axial power shape even in modern CBT testing is to replace the test rods, which is a major undertaking. Every test rod, regardless of whether it is directly or indirectly heated, is constructed to produce a specific axial power shape.

In a directly heated rod, the rod is connected to a power source at the top and bottom, and electricity flowing through the rod itself generates the heat for the test. The axial power shape is manufactured into the rod by adjusting the rod's wall thickness—this impacts the rod's electrical resistance and hence the power produced at different elevations. The outside rod diameter is held constant, and the inside diameter is changed to make the rod's cross-sectional area thicker or thinner. If the rod wall's cross-sectional area is increased by making the wall thicker, the resistivity of that section will decrease, and the power produced per unit length will decrease. Conversely, if the rod wall's cross-sectional area is decreased by making the wall thinner, the resistivity of that section will increase, and the power will increase. Because the highest rod power occurs at the thinnest areas, which are not easy to manufacture, the uncertainty on this peak power (i.e., thickness of the rod) was historically one of the largest uncertainties in the experiment.

In an indirectly heated rod, a heating coil is placed inside the rod and the power shape is controlled by modifying the dimensions of the coil. This coil is then slid into a clad, which acts as the surface of the test rod. Because PWR testing calls for high heat fluxes, PWR testing generally uses directly heated rods. BWR testing may use either directly or indirectly heated rods.

Although any number of axial power shapes could be prescribed in the manufacturing of the rods, typically the rods will have one of four shapes: (1) uniform, (2) cosine, (3) up-skew, or (4) down-skew. Aside from the uniform power shape, each power shape represents a different situation or a different time in the core life. Historically, the uniform power shape was the first power shape used in testing because of the ease of manufacturing (i.e., tubes of a constant wall thickness). However, such a shape always results in a CBT at the very top of the assembly. This situation is considered unphysical (i.e., it does not occur in actual reactors), and questions have recently been raised (Yang *et al*., 2014) on the usefulness of such uniform test data. Consequently, the uniform power shape has been used less frequently in modern CBT testing.

Because early CBT data were based on testing that assumed a uniform power shape, a method was needed to convert the model's predictions so the models could be used for the nonuniform power shapes that occur in reactors. One method used was the Tong factor (Tong *et al*., 1965). Initially, the Tong factor was not a part of the CHF model. Instead, it was used to "correct" the prediction of the CHF model. The factor attempts to adjust the predicted CHF based on the given axial power shape, some information on local conditions, and the elevation under consideration. However, as CHF models have developed, this shape dependence has become more integrated into the model itself.

Ultimately, it is important to ensure that the axial power shapes tested bound all possible power shapes for which the CBT model will be used. One way to demonstrate this is by training a model (i.e., statistically determining its coefficients using regression) with one axial power shape and validating it with another. Table 17 gives the evidence commonly provided to demonstrate that this goal has been satisfied.

Table 17 Evidence for G1.3.3—Axial Power Shapes

G1.3.3	The axial power shapes in the test assembly should reflect the expected or limiting axial power shapes in the reactor assembly.
Level	Evidence
1	Only one axial power shape was used in the test assembly. However, a justification for why the single axial power shape was sufficient is provided.
2	The commonly tested axial power shapes were used in the test assembly. Further, an explanation of why those shapes were appropriate was provided.
3	A number of axial power shapes were used in the test assembly. Further, it was demonstrated that the CBT model was able to make accurate predictions of axial power shapes whose data were not used as training data for the model.

<u>Historical Evidence Levels for Reactor Safety Analysis</u>

Level 2 and Level 3 have been most commonly accepted by the NRC staff. Generally, cosine, up-skew, and down-skew power shapes are used for BWR fuel testing, and cosine and up-skew (and maybe uniform) power shapes are used for PWR fuel testing. Level 1 has been used in the past to confirm a model's behavior on similar fuel or to make a small modification to an existing model but not to qualify a new model. Level 3 is sometimes used as it is often easier to demonstrate through test data that the CBT model is insensitive to axial power shape than to provide other justification.

3.1.3.4 G1.3.4—Radial Power Peaking

It is important to reproduce the local powers experienced by the reactor assembly in the test assembly. Generally, this has been done by testing a combination of axial and radial power shapes. Varying the radial power shape (i.e., radial power peaking) is generally much easier than varying the axial power shape because it can be done by simply supplying more power to select rods in the test assembly and does not necessitate replacing the rods in the assembly.

The importance of the radial power peaking is different for BWR and PWR testing. In PWR testing, the radial power peaking tends to be used to ensure that the CBT occurs away from the outside wall and near the central locations of the test assembly. Because the assembly is only a portion (e.g., 5x5, 6x6) of the entire assembly (e.g., 14x14, 17x17), there is a desire to ensure that the CBT occurs closer to the center of the test assembly and away from any edge effects of the wall, as such a boundary does not exist in an open lattice core. The model predicting a CBT is applied over every subchannel in a fuel assembly, and the resulting predicted CHF is compared to the heat flux from the fuel rods. Although the radial power peaking will affect the heat flux from the fuel rods and consequently the local fluid conditions, the computer code directly simulates all of those impacts.

However, the radial peaking in BWR testing serves a different purpose as a result of how BWR CP correlations are applied. In the current generation of CP correlations, assembly-average thermal-hydraulic conditions and pin powers are used as inputs to the correlation. The margin to dryout in the assembly is then calculated based on the limiting R- or K-factor. R- or K-factors are calculated for each rod based on the pin power distribution of the surrounding rods and the rod additive constant, which is a correlated parameter developed for each rod. Radial power peaking in BWR testing is therefore used to drive different rods into dryout so an additive constant can be determined for each individual rod (or its symmetric partners). This constant accounts for the local thermal-hydraulic conditions in the fluid surrounding the rod in a way that is similar to the subchannel code used in PWR CHF analysis. The testing should be performed over the full range of R- and K-factors expected in the reactor so that the local thermal-hydraulic effects are properly captured in the additive constant.

Because of this difference between PWR and BWR CBT modeling, the criteria for BWRs and PWRs are different. Table 18 gives the evidence commonly provided to demonstrate that this criterion (PWR only) has been satisfied.

Table 18 Evidence for G1.3.4—Radial Power Peaking (PWR)

G1.3.4	The radial power peaking in the test assembly should reflect the expected or limiting radial powers in the reactor assembly.
Level	Evidence
1	Radial power distributions are consistent with those peaking factors expected in reactor fuel.
2	Radial power distributions are higher than those peaking factors expected in reactor fuel.
3	Radial power distributions in the test rods result in a hot subchannel (i.e., a subchannel surrounded by peaked rods that have higher peaking factors than those normally expected in reactor fuel).

<u>Historical Evidence Levels for Reactor Safety Analysis</u>

Level 3 has been most commonly accepted by the NRC staff for PWR fuel. Generally, the hot subchannels are designed toward the interior of the test assembly to ensure the CBT does not occur on an exterior rod, which may be influenced by the channel wall.

Table 19 gives the evidence commonly provided to demonstrate that this criterion (BWR only) has been satisfied.

Table 19 Evidence for G1.3.4—Radial Power Peaking (BWR)

G1.3.4	The radial power peaking in the test assembly should reflect the expected or limiting radial powers in the reactor assembly.
Level	Evidence
1	A wide range of radial power peaking was tested.
2	The testing procedure ensured that each rod experienced dryout in multiple tests over multiple different radial power distributions, thus ensuring the thermal-hydraulic behavior captured in the R- or K-factor and any rod additive constant would be based on the appropriate rod behavior.
3	The testing procedure ensured that each rod experienced dryout in multiple tests over multiple different radial power distributions, thus ensuring the thermal-hydraulic behavior captured in the R- or K-factor and any rod additive constant would be based on the appropriate rod behavior. Additionally, the radial power peaking tested bound the possible radial powers that could be observed during normal conditions and any transients.

Historical Evidence Levels for Reactor Safety Analysis

Level 2 has been most commonly accepted by the NRC staff for BWR fuel. Generally, the tests are focused on peaking each rod in the assembly to ensure a sufficient database for calculating the additive constant. Often, every single rod in the assembly does not need to be peaked because there is some flow symmetry; therefore, only some locations need to be investigated, assuming the assembly behaves symmetrically. If the assembly does not behave symmetrically, more rods in the assembly would need to be peaked to obtain measurements of their performance.

3.1.3.5 G1.3.5—Differences in the Test Assembly

The test assembly used in the experiment and the actual fuel assembly used in the reactor should have few differences, if any. Because much of the important flow behavior of the assembly is not modeled in the computer simulation but captured through the empirical CBT model, the test assembly used to generate that model must be very similar to the actual fuel assembly. However, the two assemblies will likely always have small differences that must be understood and demonstrated to have little-to-no impact. Table 20 gives the evidence commonly provided to demonstrate that this goal has been satisfied.

Table 20 Evidence for G1.3.5—Differences in the Test Assembly

G1.3.5	Any differences between the test assembly and the reactor assembly should have a minimal impact on the flow field. This includes components that are not in the reactor assembly but are needed for testing purposes.
Level	Evidence
1	The main flow features of the test assembly are the same as those of the fuel assembly, with analysis demonstrating that all differences are small.
2	The main flow features of the test assembly are the same as those of the fuel assembly, with experiment demonstrating that all differences are small.
3	The test assembly is identical to a symmetric portion (e.g., 5x5) of the actual fuel assembly.
4	The test assembly is identical to the actual fuel assembly.

Historical Evidence Levels for Reactor Safety Analysis

Level 3 has been most commonly accepted by the NRC staff for PWR fuel because of the reduced fuel assembly size (a 17x17 reactor assembly is a 5x5 or 6x6 test assembly) and the use of support spacers. Level 3 or level 4 is most common for BWRs because the entire assembly (or a very large portion of it) can often be used in the test. Levels 1 and 2 are uncommon, as it is very difficult to justify the use of a CBT model on fuel which is very different from that which was tested.

There are known issues that create deviations between the test assembly and the fuel assembly used in the reactor. For example, in BWR testing, the part-length rods can sometimes prove problematic; therefore, the test assembly may be very similar but not exactly identical to the actual fuel assembly. Because these experiments are very costly and very difficult, differences between the test and fuel assembly are not uncommon. In some past cases, data were discarded because of such differences, and additional testing had to be conducted. In other cases, the differences were small enough that the data were acceptable for use and additional testing was unnecessary. Much is left to the experience and engineering judgment of the assessor and the analyst.

3.2 G2—Model Generation

The statement “The model has been generated in a logical fashion” is intentionally broad because the decision to rely on the model rests mostly on the validation data rather than its method of generation. Additionally, a model could be generated in many ways, and any or every one of those ways could be acceptable. Arguably, it would be possible to guess both the model form and coefficients. If such a model were appropriately validated, showed reasonable physical behavior over the range of its intended use, and had quantified uncertainty, there would be no reason to disallow the use of that model, even though it was based on a guess.

Although any number of methods could be used to generate a CBT model, understanding what method was used and the reasoning behind that method is helpful to the assessor. Therefore, the criteria in this section are less focused on ensuring that a specific method was followed and more focused on ensuring that whatever method was followed is explained and is logical.

The field of machine learning has addressed the general process used to generate a model (and many of the concerns in that process). Therefore, many of the concepts and terms used in that field will be used here. The two subgoals in Figure 8 are used to demonstrate that the model was generated in a logical fashion.

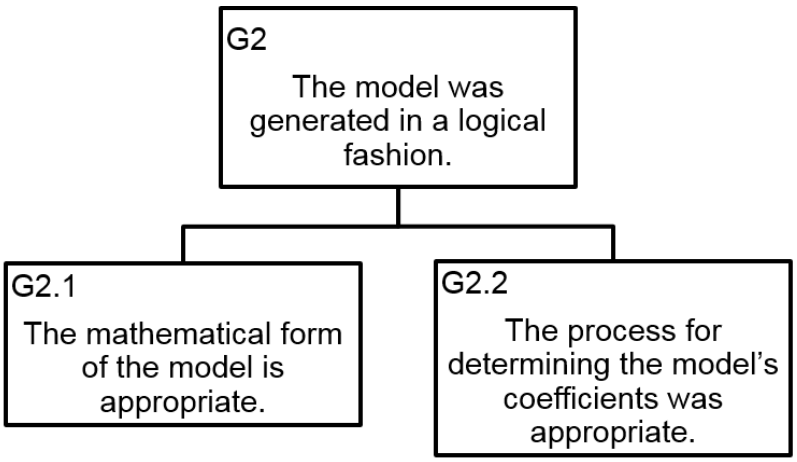


Figure 8 Decomposition of G2—Model Generation

3.2.1 G2.1—The Mathematical Form

The mathematical form of the model must be appropriate in that all relevant parameters appear as variables in the model and the model form itself is reasonable. Typically, the mathematical form of the model is chosen based on an organization’s past experience. The two subgoals in Figure 9 are used to demonstrate that the mathematical form of the model is appropriate.

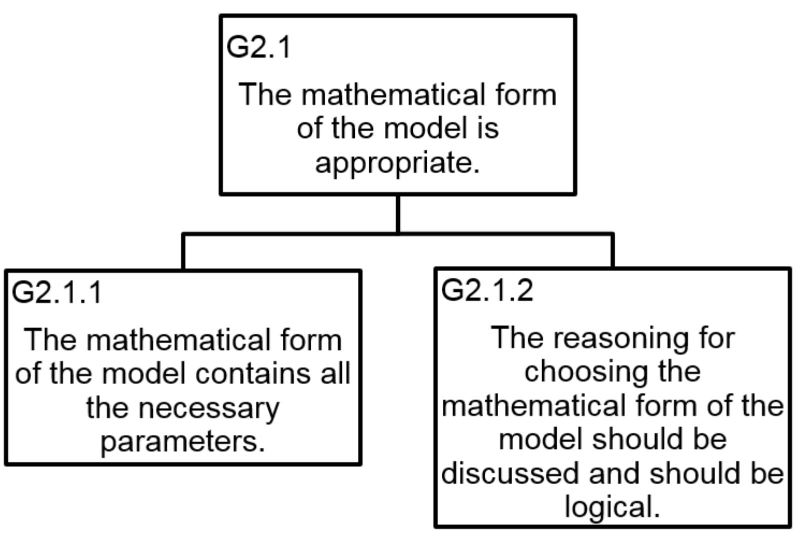


Figure 9 Decomposition of G2.1—The Mathematical Form

No further decompositions of the subgoals were deemed useful. Therefore, the sections below discuss the evidence that could be used to demonstrate that these two base goals have been satisfied. Additionally, a discussion is provided on the evidence which has been historically used for CBT models applied in reactor safety analysis.

3.2.1.1 G2.1.1—Necessary Parameters

CHF models are typically represented as a function of several (5 to 10) parameters, where each variable is generally based on a local parameter in the subchannel. The following are the most common parameters:

- pressure
- local mass flux
- local quality
- inlet enthalpy
- heated hydraulic diameter (to account for any cold-wall effect)
- grid spacing
- other flow or geometry parameters

CP models are also represented by functions of several variables but typically not by local parameters of the subchannel; instead, they are generally based on fuel assembly inlet parameters, including the following:

- pressure
- inlet mass flux
- inlet subcooling
- R- or K-factor (related to local peaking)
- additive constant (related to the flow/enthalpy redistribution of a specific spacer design)
- other flow or geometry parameters

Pressure

Pressure can have a first-order impact on the fluid properties, the flow regime, and thus the predicted CBT. Most AOOs occur at pressures close to the system pressure. The major exception to this is the main steamline break in a PWR, which typically has the lowest pressure of any AOO[18]. Because the pressure encountered during a main steamline break is usually much lower than the normal operating pressure, a specific low-pressure CHF model is often used.

Mass Flux

For PWRs, a local mass flux is used in the calculation of the CHF. This local mass flux is obtained from a subchannel code because PWRs have an open lattice core, and significant mixing between fuel assemblies can occur. For example, it is a common practice in PWR safety analyses to conservatively model the flow entering the hot assembly by reducing it by a small percentage. Because it is an open lattice core, the flow redistributes rather quickly, and this impact is almost negligible after only a few grid spacers. However, at higher axial elevations, the hotter subchannels will generate increased vapor, thus increasing the pressure drop and driving fluid to other subchannels (and potentially into adjacent assemblies). Because the local mass flux calculated by the subchannel code can have a first-order effect on the prediction of the CBT, the code (and all of the selected modeling options) is considered part of the CHF model. Any change

[18]While a main steam line break is formally classified as an accident and not an AOO, many plants analyze them to the stricter standard of an AOO. Limited fuel failure is permitted in a postulated accidents where no fuel failure is permitted in an AOO.

to the code or selection of any different modeling options would warrant revalidation of the CHF model with the new code or modeling options.

For BWRs, the local mass flux is typically not necessary because the fuel assembly is bounded by its channel, and mixing between assemblies does not occur. Therefore, CBT can be correlated to the inlet mass flux. Although a mass exchange occurs between the vapor flow, the liquid droplets, and the fluid film, this exchange is modeled through the calculation of the quality, and the CP model itself captures the entire process.

Local Quality

For PWRs, the local quality has a first-order impact on the CBT. One thing which seems to have a large impact on the local quality is the power shape. Tong's factor (or similar shape factors) accounts for different axial power shapes by reducing (or increasing) the heat flux that is needed to predict a CHF. Tong's factor is supposed to account for the "history" of the flow that would be affected by axial power shape. One theory is that the Tong factor accounts for the radial distance between the heated wall, the void location in the flow, and the void concentration. Although the quality calculated is the total quality of the subchannel, it is quality "near the wall" that would likely have the largest impact on CHF. Thus, a shape factor like Tong's is used to account for this quality distribution in a specific cell of the subchannel. Voids closer to the wall may result in a lower CHF than would voids in the center of the channel because voids at the wall could influence bubble crowding and hence influence the CHF.

For BWRs, the local quality is more of a predictive parameter than a correlating parameter. Many CP models correlate the current boiling length to a critical quality. In such a model, ensuring that CP has not occurred is synonymous with ensuring that the current quality is lower than the critical quality.

Inlet Enthalpy

The inlet enthalpy is used to determine how close the inlet flow conditions are to boiling (e.g., inlet subcooling). If the inlet subcooling is high, boiling will generally occur at higher axial elevations in the fuel assembly, and a higher power will be needed to cause a CBT. Although inlet subcooling can be low, some amount of inlet subcooling is typically necessary or else the start of boiling can occur outside of the fuel assembly and it is not possible to define a boiling length. Models that correlate boiling length to a critical quality inherently assume that the entire boiling length will be in the fuel assembly, which would therefore typically imply that the flow enters the assembly with some subcooling. Even if this assumption is not used, it is usually very difficult to test conditions with zero or negative inlet subcooling (i.e., flow is already boiling).

Inlet subcooling is not as relevant for CHF models as they focus more on local conditions. More generally, PWRs operate with inlet conditions that are much farther from saturation (i.e., more subcooled) than BWRs. However, experimental validation should be used to confirm that the flow at the inlet is subcooled if necessary.

Heated Hydraulic Diameter

Typically, for CHF models, the subchannel heated hydraulic diameter (or a ratio of the heated hydraulic diameter to the true hydraulic diameter) is used instead of the actual hydraulic diameter because of the difference between the behavior of a subchannel surrounded by four rods and the behavior of a subchannel that contains an unheated guide tube. The guide tube is considered a "cold wall"; therefore, its impact is known as the "cold-wall effect." Although a guide tube may change the hydraulic diameter of a subchannel, some guide tubes are of similar size to a fuel rod and, therefore, would have minimal impact on the hydraulic diameter of the channel. However, because the guide tube is unheated, it would have a large impact on the heated hydraulic diameter.

Although it is important to explicitly account for the cold-wall effect in PWRs, it is not directly addressed in BWRs. Generally, the K- or R-factor and the additive constants would account for any impact from the water rods or channel box in a BWR.

Grid Spacing

If the grid spacing (i.e., the distance between two grids) does not vary for a fuel design, obtaining test data at multiple grid spacings is not necessary. However, if the grid spacing can change (e.g., intermediate flow mixers are positioned between some spacer grids), the effect of the distance between all possible combinations of the grids should be accounted for the CBT model.

Typically, CBT occurs just upstream of (i.e., below) a grid spacer. For PWRs, the turbulence is maximized just downstream of (i.e., above) a grid and decreases as the fluid travels further from the grid, reaching a minimum just upstream of the next grid. Therefore, longer spans between grids result in more reduction in turbulence and less mixing, thus increasing the potential for a CBT. For BWRs, the grids direct the droplets entrained in the vapor core to the liquid film on the fuel rod, thus increasing the liquid film thickness. However, as the flow moves downstream from the grid, the additional deposition caused by the grid decreases and the liquid film evaporates and is entrained by the vapor flow. If the deposition rate falls off too quickly or if evaporation or entrainment is too great, the film may dry out before it reaches the next grid where deposition will increase once again.

Additionally, the grids themselves act as fins. Thus, while a CBT would be expected to occur just upstream of a grid, it would be highly unlikely to occur inside a grid because some amount of heat transfer occurs from the rod to the grid and the grid to the coolant. Additionally, the grids themselves are often covered in water, either from the continuous flow field in a PWR or from droplets in a BWR.

R- or K-Factor and Additive Constants (BWR only)

The R- or K-factors and additive constants account for the impacts of various phenomena on CP predictions for each rod position. The additive constants are terms that account for the increase or decrease in mixing at some *xy* location in the grid assembly. These terms are obtained from experimental testing and generally stay fixed for a particular rod *xy* location. The R- or K-factors include the impact of the various power levels of the surrounding rods on the rod in question. These factors and constants have been colloquially termed the "poor man's subchannel code." Instead of simulating a large number of subchannels in the hot assembly, a BWR analysis will simulate only a single rod surrounded by a single subchannel at assembly-averaged conditions. The R- or K-factors are then used, along with the additive constants, to determine the behavior of the rods at each *xy* location in the assembly.

Other Parameters

CBT models may use other parameters. Historically, the heated length has been used, but recent work suggests that this is not the best length parameter to correlate against because the boiling length (i.e., distance from the start of boiling to the current location under consideration) has a larger impact on the CBT (Wieckhorst *et al.*, 2013; Wieckhorst *et al.*, 2015).

Table 21 gives the evidence commonly provided to demonstrate that this goal has been satisfied.

Table 21 Evidence for G2.1.1—Necessary Parameters

G2.1.1	The mathematical form of the model contains all the necessary parameters.
Level	Evidence
1	The model contains all the parameters measured in the experiment.
2	The model parameters include those which have been commonly used in previous models and are considered to be the parameters that have the most significant impacts on a CBT.
3	It is demonstrated from first principles that the model contains all the necessary parameters.

<u>Historical Evidence Levels for Reactor Safety Analysis</u>

Level 2 has been most commonly accepted by the NRC staff. Typically, the CBT model includes a few parameters in addition to those measured in the experiment. Level 3 is considered an ideal situation, and the authors are not aware of a complete first-principle understanding of phenomena associated with a CBT. This is especially true for DNB, for which the phenomenon involved is much more complex than dryout because it involves multiple length scales and a strong dependence on turbulence. It is possible that a claim of thorough understanding of the first principles of CBT could be demonstrated by developing a correlation using very little training data and validating it against a wide variety of conditions.

3.2.1.2 G2.1.2—Reasoning for the Mathematical Form

Currently, there is no known “best” mathematical form for CBT models, which are expressed as multivariate functions because a complete first-principle understanding of the underlying phenomena does not exist. Additionally, because of nonlinear behavior, it may be difficult to separate the impact of the chosen mathematical form and the impact of the chosen values for coefficients. Thus, even identical mathematical forms can behave much differently with different choices of coefficients. Although there is no single “correct” way to generate the mathematical form, the method behind generation of the form should be described to ensure that it is reasonable to the assessor. Additionally, because the validation process will quantify the model’s uncertainty, this criterion focuses on understanding how the mathematical form was generated rather than on ensuring that it was generated in a particular manner. Table 22 gives the evidence commonly provided to demonstrate that this goal has been satisfied.

Table 22 Evidence for G2.1.2—Reasoning for the Mathematical Form

G2.1.2	The reasoning for choosing the mathematical form of the model should be discussed and should be logical.
Level	Evidence
1	The basis of the model’s mathematical form is described.
2	The basis of the model’s mathematical form is described. The description includes the development of the form and justification of the essential elements of the form.
3	A very thorough description of the origins of the mathematical form of the model is provided. This description includes the history of the form, the justifications for using the form, and the process for generating the form.

<u>Historical Evidence Levels for Reactor Safety Analysis</u>

Level 2 has been most commonly accepted by the NRC staff. In many cases, the development of the mathematical model has occurred over the course of many years and has been influenced by numerous factors. Although it is helpful for the assessor to understand this history, and it has previously increased the review efficiency, it is not strictly necessary. Thus, Level 3 and Level 1 are not uncommon.

In general, as long as the model has been validated with data that covers its expected range of use, contains all the necessary parameters, and has a logical form, then the specific form of the model would have a minor impact on model predictions. A model with a logical form will generate relevant predictions over the entire application domain. Trends between data points should be reasonable in that the model should not be discontinuous and the trends should be well-behaved mathematically. Because there are a large variety of mathematical forms that could be chosen, the specific form should not result in unreasonable predictions (e.g., very high, very low, negative, complex numbers) inside the expected domain.

3.2.2 G2.2—Method for Determining Coefficients

The process for determining the values of the model's coefficients should be appropriate. Again, the meaning of "appropriate" in terms of a model's coefficients is vague. Although only a single set of the coefficients would result in the lowest error, as judged by some norm (e.g., the Euclidian norm), minimizing this error is often not the most important criterion when determining the coefficient values. Instead, great care is usually taken to ensure that the model reflects actual physical behavior rather than simply minimizing the error. Thus, many of the coefficients for a model are chosen to ensure that the model has certain desired trends. The three subgoals in Figure 10 are used to demonstrate that the method for determining the coefficients is appropriate.

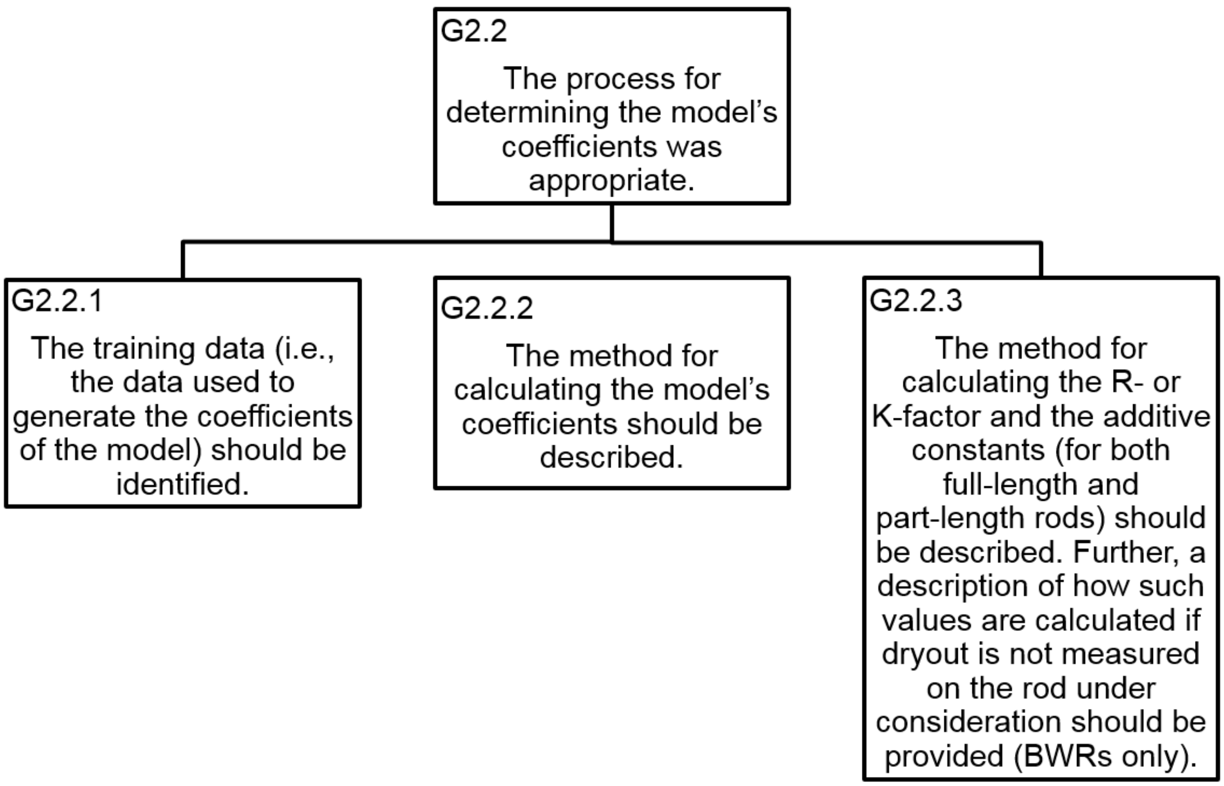


Figure 10 Decomposition of G2.2—Method for Determining Coefficients

No further decompositions of the subgoals were deemed useful. Therefore, the sections below discuss the evidence which could be used to demonstrate that these three base goals have been satisfied. Additionally, a discussion is provided on the evidence that has been historically used for CBT models applied in reactor safety analysis.

3.2.2.1 G2.2.1—Identification of Training Data

The training data are the experimental data used to generate the coefficients of the model. They are distinguished from the validation data, which are the experimental data that are used in the validation process. Ideally, different data should be used for each role. Typically, some large percentage (usually between 70 and 100 percent) of the experimental data will be used as training data. Table 23 gives the evidence commonly provided to demonstrate that this goal has been satisfied.

Table 23 Evidence for G2.2.1—Identification of Training Data

G2.2.1	The training data (i.e., the data used to generate the coefficients of the model) should be identified.
Level	**Evidence**
1	100% of the experimental data are used as training data.
2	Between 90–100% of the experimental data are used as training data.
3	Between 80–90% of the experimental data are used as training data.
4	Between 70–80% of the experimental data are used as training data.
5	Between 60–70% of the experimental data are used as training data.
6	Between 50–60% of the experimental data are used as training data.
7	Between 40–50% of the experimental data are used as training data.
8	Between 30–40% of the experimental data are used as training data.
9	Between 20–30% of the experimental data are used as training data.
10	Between 10–20% of the experimental data are used as training data.
11	Between 0–10% of the experimental data are used as training data.
12	None of the experimental data are used as training data.

Historical Evidence Levels for Reactor Safety Analysis

Levels 1–3 have been most commonly accepted by the NRC staff. As there is no minimum or maximum portion of the data that should be used to train the model, this criterion focuses more on identifying what data are used to train the model rather than on ensuring that a certain amount is (or is not) training data.

In general, all experimental data should be either training or validation data. Thus, if 70 percent of the data are training data, the remaining 30 percent could be used as validation data. Section 3.3 discusses the criteria on the amount of validation data. However, one way to demonstrate the power of a specific model is to have a very small percentage of training data and a very large percentage of validation data.

3.2.2.2 G2.2.2—Calculation of the Model's Coefficients

Again, there is typically no single best way to calculate the model's coefficients. For PWRs, because of the simplicity of the CHF model, the focus is typically on reducing overall error. However, the models for BWRs are generally more complex; therefore, the focus is typically on ensuring that the model has the desired behavior as a function of certain parameters. Whichever method is used, the assessors should understand that method. Table 24 gives the evidence commonly provided to demonstrate that this goal has been satisfied.

Table 24 Evidence for G2.2.2—Calculation of the Model's Coefficients

G2.2.2	The method for calculating the model's coefficients should be described.
Level	Evidence
1	A brief description of the method for calculating the model's coefficients is provided.
2	A detailed description of the method for calculating the model's coefficients is provided.
3	A very thorough description of the method for calculating the model's coefficients is provided. This includes the walkthrough for gathering the experimental data, the data reduction process, and the methods used to generate the coefficients.

<u>Historical Evidence Levels for Reactor Safety Analysis</u>

Level 2 has been most commonly accepted by the NRC staff. The method for calculating the model's coefficients tends to be very detailed. The models are treated as strictly data driven models (i.e., empirical or semi-empirical) in that there is no assumption that the model form contains any ability to predict the physics besides that which it demonstrates through its validation. While it is possible that the model form may be based on equations from first-principle physics, it is not assumed that the model contains any inherent ability to predict the underlying physical mechanisms of the CBT. Therefore, there is no "best practice" in terms of the manner in which the model's coefficients are calculated.

Because the model's uncertainty will be quantified with validation data, choosing the model's coefficients is mostly focused on reducing the model's uncertainty. In the extreme case, the model's coefficients could be guessed and, as long as the model's uncertainty is quantified, the model would still be acceptable for use (all non-linear regressions require a guess of the model coefficients as a starting point). Further, it is common for the model coefficients to be chosen to ensure some known behavior over specific ranges of the model, and not simply to ensure the smallest validation error.

3.2.2.3 G2.2.3—Calculation of Model-Specific Factors and Constants (BWR Only)

The R- or K-factor and additive constants are part of the coefficients of the model itself. However, they are often treated separately from the calculation of other coefficients in the model. They are a very important part of BWR simulations because they allow local fuel rod behavior to be modeled without using detailed local conditions, so their generation should be well understood. Table 25 gives the evidence commonly provided to demonstrate that this goal has been satisfied.

Table 25 Evidence for G2.2.3—Calculation of Model-Specific Factors and Constants

G2.2.3	The method for calculating the R- or K-factor and the additive constants (for both full-length and part-length rods) should be described. Further, a description of how such values are calculated if dryout is not measured on the rod under consideration should be provided (BWRs only).
Level	Evidence
1	A brief description of the method for calculating these values is provided.
2	A detailed description of the method for calculating these values is provided.
3	A very thorough description of the method for calculating these values is provided. This includes a walkthrough for gathering the experimental data, the data reduction process, and the methods used to generate these values.

Historical Evidence Levels for Reactor Safety Analysis

Level 2 has been most commonly accepted by the NRC staff. The method for calculating the R- or K-factor and additive constants tends to be very detailed. It is important for the assessor to understand the process so that he or she can confirm that the behavior modeled in the R- or K-factors and additive constants would result in a reasonable prediction of CBT in a BWR.

3.3 G3—Validation through Error Quantification

Validation is the accumulation of evidence used to assess the claim that a model can predict a physical quantity (Oberkampf and Roy, 2010). Thus, validation is a never-ending process because more evidence can always be obtained to bolster this claim. However, at some point, when the accumulation of evidence is considered sufficient to make the judgment that the model can be trusted for its given purpose, the model is said to be validated. This is not to say that further validation would not be useful but rather that it is believed that the validation currently provided demonstrates that the model can be trusted for its specific use. The authors believe that Anderson and Bates were very wise to begin the first chapter of their book on validation (Anderson and Bates, 2001) with a quote from the National Research Council: “Absolute validity of a model is never determined” (National Research Council, 1990).

Because of the desire to ensure that the model’s prediction is conservative, any bias or uncertainty, or both, in the model’s prediction of CHF or CP should be adequately quantified such that safety analyses can account for it. This process is uncertainty quantification. The first step in this process is to use the experimental data (i.e., the validation data) along with the model’s prediction of that experimental data to calculate the validation error. If the validation error is appropriately distributed through the model’s application domain and if any inconsistencies in the validation error are accounted for, statistics from the validation error can be used to determine the model’s uncertainty. The five subgoals in Figure 11 are used to demonstrate that the model has sufficient validation through the quantification of its error.

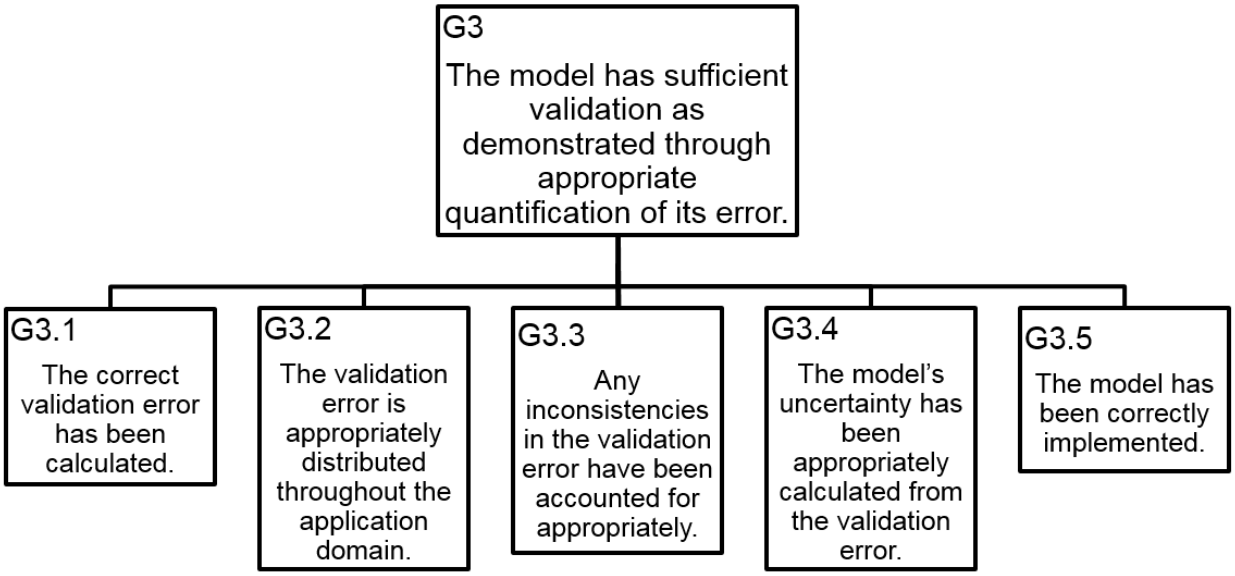


Figure 11 Decomposition of G3—Validation through Error Quantification

3.3.1 G3.1—Calculating Validation Error

Typically, model error is thought of as the difference between the *actual* value that occurs in nature and the *predicted* value of the model. If the model is simple enough or if the experiment is complex enough, the *measured* value from the experiment can be used as the actual value.[19] Ideally, the error could be calculated from the measured value of the instrumentation and the model's prediction under the same conditions. However, this is often oversimplification. Instead, the error of interest should not be the model error but the model application error (i.e., what is the error of using the model in the same manner as it will be applied in the safety analysis).

To clarify, one way to calculate the model error is to measure the heat flux or power at the location of a CBT and consider this the *measured* value and then use the CBT model along with the flow conditions at the time of the CBT to obtain a *predicted* value at that same location. However, a CBT model is generally not applied in this manner. First, it is typical for multiple rods to experience a CBT at the same time. Second, it is typical for the same rod to experience a CBT at different elevations at the same time. Third, the definition for a rod experiencing a CBT is somewhat variable. Generally, the criteria for determining the occurrence of a CBT is some specified change in temperature over a short time span. During testing, a number of thermocouples may register a change just under this amount; therefore, the rods are not considered to have experienced a CBT. However, under this definition it is possible that a CBT may still have occurred. These challenges could make determination of a single *measured* value from an experiment very difficult.

Additionally, the objective is not to ensure that the CBT model can be trusted for predicting the behavior of an experiment for which the heat flux or power that causes a CBT are known; instead, the objective is to determine whether the model can be trusted when applied in a reactor safety analysis where the heat flux or power will be unknown. The interest is not in the *model error* but in

[19] This statement ignores any differences between the measured value of a quantity and the actual value of that quantity; the discussion on instrumentation uncertainties addresses these differences.

the *model application error*. For this reason, the *measured* and the *predicted* values should be related to how a reactor safety analysis applies the model.

For example, the focus of PWR safety analysis is to determine which of the subchannels have the MDNBR value because this subchannel would be the “closest” one to experiencing a CBT. Thus, when a transient is simulated, the MDNBR is obtained, and if that value is greater than some safety limit, CBT is precluded. This method of analysis differs from the experiment in two main ways. First, the experiment determines which *rod* experienced a CBT, but the simulation determines which *subchannel* has the MDNBR. Second, because of how a CBT is defined in the experiment, it is common for more than one rod and even more than one location on the same rod to register as having experienced a CBT, but the simulation produces only one MDNBR value. Because the model is applied using the MDNBR, the *measured* and *predicted* values should be related to the DNBR value.

The term “validation error” was chosen to represent the error of interest for two main reasons. The first reason is to distinguish it from the model error, which is commonly thought of as a difference between the model’s prediction and a measurement. Determining the *measured* and *predicted* values is not as straightforward as many may consider. The second reason is that this error could have been called a “model application error,” but that term was not chosen for a different reason. The model application error is defined as the total population of error of the possible uses of the model inside the expected domain. If an experimental measurement of CBT could be obtained at every point in the expected domain (i.e., an infinite number of points), than that infinite set would be the actual model application error. The validation error is a sample from the model application error population. The validation error is based on the validation data, which only exist at a finite number of points in the expected domain. This distinction is important because one of the key assumptions is that the validation error is a representative sample of the model application error.

Generally, the validation error for a model is either represented as an absolute error (i.e., measured – predicted), or as a relative error (e.g., (measured – predicted)/measured). CBT models in particular use a form of the relative error—measured/predicted is commonly used for PWR validation and predicted/measured is commonly used for BWR validation. Thus, for PWRs, values that are below 1 are non-conservative (i.e., a CBT occurred at heat fluxes below the model’s prediction), and values that are above 1 are conservative (i.e., a CBT occurred at heat fluxes above the model’s prediction). Conversely, for BWRs, values that are below 1 are conservative (i.e., a CBT occurred at powers above the model’s prediction), and values that are above 1 are non-conservative (i.e., a CBT occurred at powers below the model’s prediction).

Table 26 gives the evidence commonly provided to demonstrate that this goal has been satisfied.

Table 26 Evidence for G3.1—Calculating Validation Error

G3.1	The correct validation error has been calculated.
Level	Evidence
1	The validation error is a sample for the population of the model error.
2	The validation error is a sample for the population of the model application error.
3	The model is applied such that the populations of the model error and model application error are identical. The validation error is a sample from this population.

<u>Historical Evidence Levels for Reactor Safety Analysis</u>

Level 2 has been most commonly accepted by the NRC staff. While CBT models are often considered as stand-alone models, they are used as part of larger thermal-hydraulic methodologies. Thus, the error in a CBT model is typically quantified as if it is being used inside the larger methodology (level 2) rather than used as a standalone model (level 1). Level 3 would be ideal as it would mean that model can be treated as a standalone equation.

3.3.2 G3.2—Data Distribution in the Application Domain

The validation error data points should be appropriately distributed throughout the application domain. Consider each of the N input variables used by the model as a dimension (e.g., pressure, mass flux, inlet subcooling). The set of all inputs could be used to generate an N-dimensional "application space," and the "application domain" is the domain in this space over which the model could be applied to predict CHF or CP. Typically, the application domain is defined as an *n*-orthotope which is a two-dimensional (2-D) rectangle, a three-dimensional (3-D) box, or a hyper-rectangle in dimensions greater than 3-D. This shape, the generalization of a rectangle to higher dimensions, is a simplification of the true shape of the application domain and is used because it can be easily defined by N inequalities (corresponding to the number of dimensions in the application space). Using this shape allows a computer program to easily determine whether the current location in the application space is inside or outside of the application domain. For example, the boundaries on the pressure are typically given as follows:

$$P_{Min} \leq P \leq P_{Max} \quad (2)$$

To ensure the model should be used to make a prediction, the computer code will check to ensure that the current pressure is between the minimum and maximum pressure of the application domain.

Defining the application domain as a set of independent inequalities is computationally convenient, but the model may not be valid over that entire domain. Consider the following simplified 2-D domain. Six types of subregions can be defined within the 2-D application space, depending on their proximity to validation error data points and their position relative to the application domain. These six types of subregions, shown in Figure 12, would also exist in application spaces of higher dimensions.

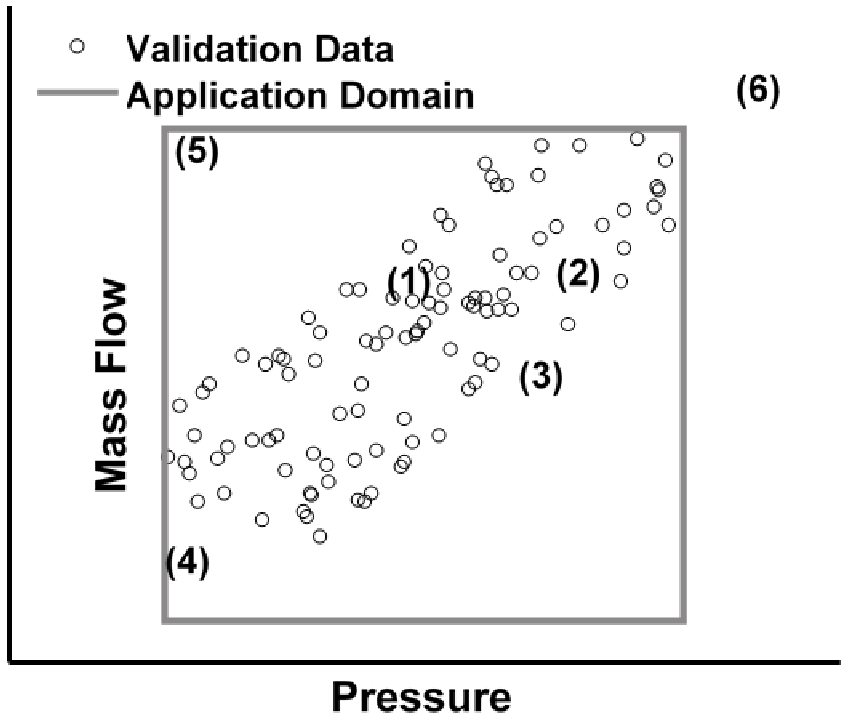


Figure 12   Regions in the Application Domain

<u>Region 1—Well Covered</u>

The first type of region is any region in the application domain that both contains data and is surrounded by data. In this region, the data are not sparse, and the region would be considered "well covered." Although it is tempting to believe that the entire application domain is "well covered," this is only the ideal and is generally not true in practice.

<u>Region 2—Localized Hole</u>

The second type of region is any region in the application domain that contains little to no data but is surrounded by data and thus forms a hole. As the number of dimensions of the application domain increases (i.e., Figure 1 shows a 2-D application domain, but it is common to have domains of six or more dimensions), it is not always clear whether the use of the model in such a region should be considered interpolation or extrapolation. In either circumstance, as long as the region itself is not too big, the use of the model in such regions is generally accepted as justified. Note that there will always be a "hole" between data points because the space is continuous, and the data exist only at discrete points. However, the assessor must exercise judgment about how far apart data are to constitute a localized hole.

<u>Region 3—Edge</u>

The third type of region is any region in the application domain that contains little to no data and is only partially surrounded by data and, therefore, is at an edge. Although uses of the model "near" the bulk of the data would seem reasonable, at some point the region of interest becomes sufficiently distant from the validation error data that the model cannot be considered validated and should not be used in the absence of other justification.

Region 4—Isolated Known Unknown

The fourth type of region is any region in the application domain that contains no data and is somewhat far from any region that does contain data; however, it is a region over which the model can be justified. For example, one common, conservative modeling assumption is to construct CBT models such that the predicted CHF or CP will be 0 at a mass flux of 0. In reality, as the mass flux goes to 0, the predicted CHF will go to a pool-boiling CHF value, which is much higher than 0. Thus, while the region may not have data, the use of the model in the region would be known to be very conservative.

Region 5—Isolated Region

The fifth type of region is any region in the application domain that contains no data and is far from any region that does contain data. In other words, it is an isolated region. Moreover, it is an isolated region in which the model's behavior is unknown. The application domain likely only includes such a region because the choice to represent the domain was a rectangle. The use of the model in such regions of the application domain should be precluded, but that would only be accomplished by defining a more complicated shape for the application domain. In 2-D, this could be easily done. However, many real models are in six or even more dimensions, and the representation of complex shapes in multiple dimensions is very difficult. Although the application domain will likely always be defined as a hyper-rectangle, the domain where the analyst expects[20] to use the model is actually closer to a "hyper-jelly bean" (as described by one engineer).

This is a concern with any higher dimensional model and is one reason why the application domain needs to contain validation error data that span the expected range of use. Although isolated unknown regions are always possible, the best way to ensure that the model will never be used in such a region is to ensure that all conceivable regions where the model will actually be used have data.

It is important to realize that the model's prediction in an isolated unknown region is suspect. On the one hand, the model may have been developed such that it happens to provide reasonable estimates of the CBT in that region. On the other hand, it may predict something completely unphysical in that region such as a negative heat flux or negative power. Model developers tend to understand where these regions exist, and apply the models only in the regions that contain data. However, a new user who is unfamiliar with the model's development process could easily pick up the model, start using it, and wonder why it is making very strange predictions in certain regions.

Region 6—Outside Region

The sixth type of region is any region that is outside the application domain. The computer code using the model will flag the use of the model as improper only in these outside regions. Considering these regions, the six subgoals in Figure 13 are used to demonstrate that the validation error data were appropriately distributed throughout the application domain.

[20] Hence, this document separates the two domains. The Application Domain is the domain over which the model is applied, and is an n-dimensional rectangle. However, the Expected Domain is where the analysts would expect the model to be used, and is subset of the application domain, but generally a much more complex shape that cannot easily be well defined.

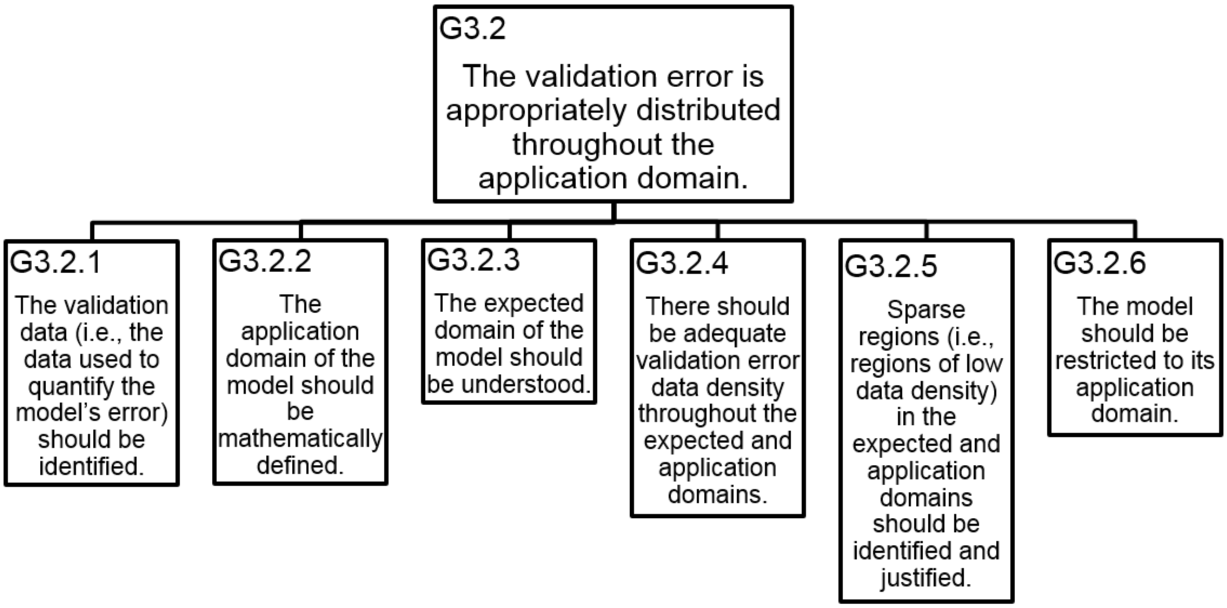


Figure 13 Decomposition of G3.2—Data Distribution in the Application Domain

No further decompositions of the subgoals were deemed useful. Therefore, the sections below discuss the evidence that could be used to demonstrate that these six base goals have been satisfied. Additionally, a discussion is provided on the evidence that has been historically used for CBT models applied in reactor safety analysis.

3.3.2.1 G3.2.1—Identification of Validation Data

The validation data are the experimentally measured values that are used to quantify the model's error. Ideally, these data should be independent from the training data. The model will be used to make predictions about the CBT throughout the application domain. The focus of validation is to quantify the error of those predictions. Although it may seem that use of the training data would be appropriate, the model has already been "tuned" to that data. Thus, quantifying the error of the training data would provide an estimate of "how well the model can predict data that were used to generate the model." This is different from "how well the model can predict data that were not used to generate the model." Because substantially more data points appear in the application domain (an infinite number) than were used to generate the model and because these points are the ones of most interest in future uses of the model, the focus should be on generating an estimate of the error over those points which were not used to generate the model. Thus, experimental data that have not been used to train the model should be held in reserve and used only to validate the model because the model's behavior using these data are indicative of the type of predictions that will be made in its future uses.

However, in many instances, the validation data and the training data are one and the same. There are methods in machine learning that can be applied to determine whether the selection of the training data affects the resulting uncertainty, such as random subsamples and k-folds. In each of these methods, the data are randomly separated into subsets of training and validation data. The training data are used to develop the coefficients of the model, and the validation data are used to determine the overall uncertainty of the model. Then, the process is repeated with a different randomly-selected data set assigned to training and the remaining data assigned to validation. Processes like these can provide reasonable estimates of the impact of using the same training data as validation data. Even for well-formed models, using the same dataset for training and validation can increase uncertainty by 2 to 3 percent. This increase is small but far from negligible, and it may be higher or lower depending on the circumstances. Table 27 gives the evidence commonly provided to demonstrate that this goal has been satisfied.

Table 27 Evidence for G3.2.1—Identification of Validation Data

G3.2.1	The validation data (i.e., the data used to quantify the model's error) should be identified.
Level	Evidence
1	Validation data have been identified, and they are the same as the training data.
2	Validation data have been identified, and they are the same as the training data. To quantify this impact, a method such as k-folds or random subsamples has been used.
3	The validation data are independent from the training data.

Historical Evidence Levels for Reactor Safety Analysis

Level 3 (specifically, a 70/30 or 80/20 split between the training and validation data) has been most commonly accepted by the NRC staff. In a sense, this is similar to performing a single k-folds calculation or a single random calculation of subsamples. Level 2 has been used in the past, but the model's error in predicting data that were not used to generate the model will almost always be greater than its error in predicting data that were used to generate the model. Therefore, using the same data for training data as validation data often involves additional work, such as k-folds or random subsamples. If only Level 1 is achieved, a bias may need to be added to account for the fact that the resulting error is likely lower than actually expected.

3.3.2.2 G3.2.2—Defining the Application Domain

The application domain should be defined such that the computer code applying the model is able to determine whether the model should be used for a given set of input parameters. Generally, this is done using inequalities such as those given in the following expressions:

$$P_{Min} \leq P \leq P_{Max} \tag{3}$$

$$G_{Min} \leq G \leq G_{Max} \tag{4}$$

Defining the application domain in such a manner results in a hyper-rectangle, which contains many regions in which no data exist. Although a more accurate method of defining the application domain could be used to only specify the region which contains data, such alternative methods do not currently exist. Table 28 gives the evidence commonly provided to demonstrate that this goal has been satisfied.

Table 28 Evidence for G3.2.2—Defining the Application Domain

G3.2.2	The application domain of the model should be mathematically defined.
Level	Evidence
1	The application domain has been mathematically defined as a hyper-rectangle.
2	The application domain has been mathematically defined as a shape other than a hyper-rectangle to better capture its true shape.
3	The application domain has been mathematically defined in terms of a maximum allowable distance from validation error data.

Historical Evidence Levels for Reactor Safety Analysis

Level 1 has been most commonly accepted by the NRC staff. Because application domains defined as hyper-rectangles often contain many regions which are technically part of the application domain, but contain no data and are far from where the plant operates, there is generally a desire for the analyst to not only specify the application domain, but also to understand the expected domain.

3.3.2.3 G3.2.3—Understanding the Expected Domain

The application domain is defined as the domain in the N-dimensional input space over which the model *could* be applied. However, that domain is different from the domain over which the model is *expected* to be applied. The “expected domain” is defined as the domain in the N-dimensional input space over which the model will *likely* be applied because it corresponds to state points that occur during normal operation or AOOs. Unlike the application domain, which is mathematically defined so that a computer can determine whether the model is being used outside of that domain, the expected domain is generally not formally defined due to its complex shape. For example, if the application domain is represented as a box in a series of 2-D plots of one input parameter versus another input parameter, the expected domain would be represented by some region in each box.

Ideally, as knowledge progresses, the application domain would become closer to the expected domain, and both domains would contain only regions with data. Table 29 gives the evidence commonly provided to demonstrate that this goal has been satisfied.

Table 29 Evidence for G3.2.3—Understanding the Expected Domain

G3.2.3	The expected domain of the model should be understood.
Level	**Evidence**
1	Each parameter in the model is considered separately.
2	2-D plots (parameter versus parameter) that contain the locations of the validation error data and the expected range of those parameters during normal operation and AOOs are provided. The expected ranges are well covered by validation error data (N parameters = $\frac{N \cdot (N-1)}{2}$ plots).
3	Another method that considers more than two parameters at a time (e.g., 3-D plots) is used.
4	A method that considers all N parameters at the same time is used.

Historical Evidence Levels for Reactor Safety Analysis

Level 2 has been most commonly accepted by the NRC staff. Although used in the past, Level 1 completely ignores any correlations between the input parameters themselves. For example, if there is only low-pressure data at low mass flows and high-pressure data at high mass flows, the model's prediction in the region that has both low-pressure and high mass flow would not have any associated validation error data. However, to determine whether this situation exists, the data should be plotted with more than one input parameter at once (i.e., at least a 2-D plot).

Just as Level 1 reduces the problem to a single dimension in the N-dimensional input space, Level 2 reduces the problem to two input dimensions. Both dimensional reductions cause the loss of information, but the information loss caused by reductions from N dimensions to 2-D is believed to be less significant. Ideally, all N dimensions could be considered at the same time, but the authors are not currently aware of a method for doing so.

3.3.2.4 G3.2.4—Validation Error Data Density in the Expected Domain

The expected domain should have adequate data density to ensure adequate coverage for future uses of the model. Typically, the regions with the most data will be those regions in which the plant will be close to normal operating conditions. However, these regions are not necessarily the same as the regions in which the plant would be closest to experiencing a CBT. Thus, although certain regions are expected to have a very high density of validation error data, the entire expected domain should be well covered. Note that the entire application domain will likely not be covered, due to the practice of representing the application domain as a hyper-rectangle. While it is not necessary for the entire application domain to be well covered with validation data, it is necessary for the expected domain to be well covered. Table 30 gives the evidence commonly provided to demonstrate that this goal has been satisfied.

Table 30 Evidence for G3.2.4—Validation Error Data Density in the Expected Domain

G3.2.4	There should be adequate validation error data density throughout the expected and application domains.
Level	**Evidence**
1	Each input parameter is considered independently from all others. Few regions have sparse data, and the model’s use in those regions can be justified. Thus, the problem is treated as N number of 1-D spaces.
2	A set of two input parameters are considered in combination. The data density is sufficient, only a few regions of sparse data exist, and the model’s use in those regions can be justified. All possible combinations of sets of two input parameters are considered. Thus, the problem is treated as $\binom{N}{2}$ number of 2-D spaces.
3	A set of three input parameters are considered in combination. The data density is sufficient, only a few regions of sparse data exist, and the model’s use in those regions can be justified. All possible combinations of sets of input parameters are considered. Thus, the problem is treated as $\binom{N}{3}$ number of 3-D spaces.
4	All input parameters are considered in combination. The data density is sufficient, only a few regions of sparse data exist, and the model’s use in those regions can be justified. Thus, the problem is treated as a single N-D space.

<u>Historical Evidence Levels for Reactor Safety Analysis</u>

Level 2 has been most commonly accepted by the NRC staff. Again, Level 1 is considered insufficient because it ignores any correlations between the input parameters themselves. Level 3 would require the use of 3D plots, and such plots are difficult to represent in a 2D document (i.e., on a sheet of paper). For this reason, assessors have previously found it important to obtain the data used to correlate and validate the model—this data can be used to generate 3-D plots, which can be examined in detail on a computer. In addition, the number of dimensions observed at the same time can be increased to four dimensions by using a color gradient on the points. Higher dimensional plots are possible, but understanding such plots as the number of dimensions grows becomes difficult.

As of yet, there are no precise limits on data density. Even the concept of data density is difficult to define precisely, as the volume over which the data density would be determined contains dimensions that cannot be easily combined in a meaningful way. Therefore, the density in each region is generally judged to be sufficient if it is similar to previous densities from past approved models. It is expected that there will be a large cluster of points around the normal operating conditions and fewer points at the extremes of the expected domain.

Finally, levels 1-3 are graphical methods that rely on qualitative judgment. Level 4 considers some method that is quantitative, but the authors are not aware of any such method that currently exists.

3.3.2.5 G3.2.5—Sparse Regions

As discussed above, there may be sparse regions in the application domain for a variety of reasons. Usually, sparse regions appear in the application domain because of the method chosen to describe the domain (e.g., as a hyper-rectangle). However, these regions in the application domain may not be a part of the expected domain. Such regions in the application domain but not in the expected domain should be identified, but further justification is not necessary, as the model is not expected to be used in those regions. However, any sparse region that lies within the expected domain would need further justification as the model would be expected to be used in that region. Table 31 gives the evidence commonly provided to demonstrate that this goal has been satisfied.

Table 31 Evidence for G3.2.5—Sparse Regions

G3.2.5	Sparse regions (i.e., regions of low data density) in the expected and application domains should be identified and justified.
Level	Evidence
1	There are many sparse regions in the expected domain.
2	There may be some sparse regions in the application domain. There may be some sparse regions in the expected domain.
3	There may be some sparse regions in the application domain. There may be some sparse regions in the expected domain, but the use of the CBT model in these regions is justified.
4	There may be some sparse regions in the application domain. There are no sparse regions in the expected domain.
5	There are no sparse regions in either the application or the expected domain.

<u>Historical Evidence Levels for Reactor Safety Analysis</u>

Level 3 has been most commonly accepted by the NRC staff. There may be sparse regions that are at the edges of the model's intended use (e.g., low mass flux or high mass flux), though additional justification is usually provided for these regions. As discussed above, there are numerous ways to justify the use of a model in a sparse region. The most common are: (1) demonstrating that the model is conservative in the region, (2) demonstrating that it is not possible for the fuel assembly to operate in the region, and (3) demonstrating that the region is not, in fact, a sparse region. However, there are often instances in which the model does need to be used in a region that is sparse (or at least has a very low data density). In these instances, a bias applied to the model in the region in question may address the sparseness of the data without unnecessarily negatively impacting the model's predictions in other parts of the application domain. In the higher dimensional spaces that are typical of most real application domains, the issue of sparse regions becomes more difficult to understand and define.

3.3.2.6 G3.2.6—Restricted to the Application Domain

Restricting the CBT model to its application domain is important. There are a variety of ways in which this restriction can be placed and upheld on the computer code using the CBT model. Table 32 gives the evidence commonly provided to demonstrate that this goal has been satisfied.

Table 32 Evidence for G3.2.6—Restricted to the Application Domain

G3.2.6	The model should be restricted to its application domain.
Level	Evidence
1	The computer code does not check whether the model is being used outside of its application domain. Instead, the code analyst ensures that the model was used only inside of its application domain when reviewing the code output.
2	If the computer code attempts to use the model outside of its application domain, the code's output marks it as a warning; however, the simulation continues to run.
3	If the computer code attempts to use the model outside of its application domain, the code's output marks it as an error; however, the simulation continues to run.
4	If the computer code attempts to use the model outside of its application domain, the code's output marks it as an error, and the simulation immediately quits running.

<u>Historical Evidence Levels for Reactor Safety Analysis</u>

Levels 3 and 4 have been most commonly accepted by the NRC staff. Level 1 would present human-factors issues and should not be used if more than a few simulations are needed in the particular application. Level 2 could also be present human-factors issues because users may not recognize the severity of the application domain violation. In general, appropriate evidence for G3.2.6 depends on the QA program the simulation is performed under and whether the restriction to the application domain is the responsibility of that QA program or of the computer code itself.

3.3.3 G3.3—Inconsistency in the Validation Error

Statistics from the validation error will be used as estimates of parameters from the population of the model application error in order to quantify the uncertainty of the CBT model. This assumes that the model application error can be described as a single population with the same distribution and parameters (e.g., mean, variance) over the entire application domain and that the validation error is a representative sample of this distribution.

As discussed by Box, Hunter, and Hunter (1978) one of the key assumptions in the data are the assumption of independence. If the model application error is dependent on its location in the application domain, it would be a collection of many populations, not a single population. Piepel and Cuta (1993) argue that the validation error would not likely be from a single population; instead, it would contain subregions in the application domain where the validation error would be from different populations. Although the authors agree that this is likely the case, the assumption of a single underlying population and independence should be reasonable as long as the validation error is consistent and no obvious non-conservatisms exist.[21] The three subgoals in Figure 14 are used to demonstrate that any inconsistencies in the validation error have been appropriately addressed.

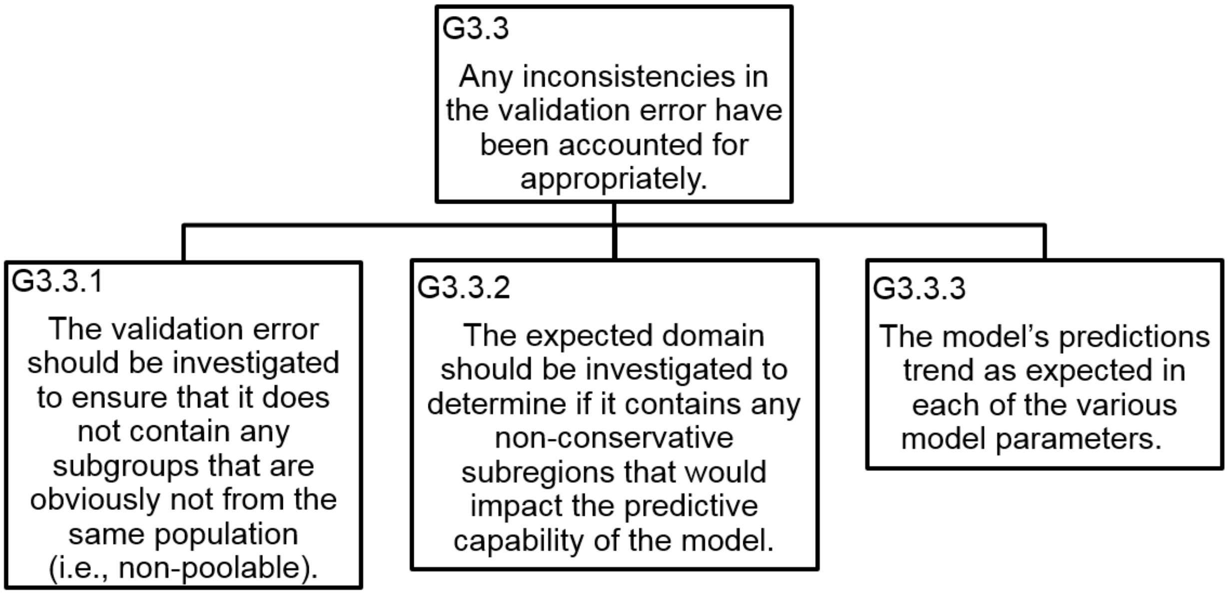


Figure 14 Decomposition of G3.3—Inconsistencies in the Validation Error

No further decompositions of the subgoals were deemed useful. Therefore, the sections below discuss the evidence which could be used to demonstrate that the three base goals have been satisfied. Additionally, a discussion is provided on the evidence which has been historically used for CBT models applied in reactor safety analysis.

3.3.3.1 G3.3.1—Identifying Non-poolable Data Sets

The validation error is typically made up of multiple sets of data. The validation error data may be taken at low pressures, high flows, different axial power shapes, slightly different geometries, and so on. Analysts generally assume that all of this data are poolable, i.e., that all of the data can be treated as if they came from a single underlying population. If this is true, then the validation error, which is based on the validation data, may be a representative sample from this larger population, and therefore a good estimate of the behavior of the total population of model application error.

However, there are a number of reasons why the validation error may not be a representative sample of the overall population of the model application error. First and foremost, the validation error itself may represent several different populations. That is, pooling all of the validation errors from each data set into a single validation error may be incorrect. For example, the CBT model may make much better predictions at low pressures than at high mass fluxes. Pooling the data would obscure this difference. The assumption of poolability should be tested by identifying key data subsets in the validation error data set and by determining whether those data sets are indeed from the same population.

[21] In this sense, "non-conservative" means that the prediction of the CBT model over predicts the CHF or CP value by an amount greater than that accounted for by any uncertainty applied to the model. For CHF correlations, this would typically be the 95/95 value.

Although statistical tests can be performed to determine whether two subsets are from the same population (i.e., have the same distribution shape, the same mean, the same variance), caution should be used. Incorrectly determining that the data sets are from different populations when, in fact, they are from the same population is common. This is known as a Type 1 error or a “false positive.” The probability of Type 1 errors increases with each additional test performed. For example, using the common significance value of 5 percent, the probability of obtaining a false positive after one test is only 5 percent. However, if 14 tests are performed with the same significance value, the probability of obtaining at least one false positive is over 50 percent. Thus, even if the data are from the same population, performing 14 tests will more than likely result in the conclusion that the data are from different populations. Therefore, these tests should be applied only when necessary.

For PWRs, data should be separated (at a minimum) by axial power shape and by subchannel type (rod and guide tube) as these are the main data sets that have been shown to be non-poolable. If any of the sets are not poolable, the model’s uncertainty should be derived from the limiting data set. For BWRs, data should be separated (at a minimum) by axial power shape. Thus, it should be determined whether all power shapes are poolable data sets. If they are not poolable, the model’s uncertainty should be derived from the limiting data set.

The following statistical tests are commonly used during this process:

- Analysis of variance, commonly known as ANOVA, for equality of means
- T-test, for equality of means
- F-test, for equality of variances
- Chi-square test, for equality of variances
- D’Agostino’s test, for normality
- Wilks-Shapiro test, for normality
- Anderson-Darling test, for normality

If the validation error is made from (i.e., contains) data sets that are not poolable, the most limiting or most conservative data set should be chosen if using a single value to quantify the model’s uncertainty. Table 33 gives the evidence commonly provided to demonstrate that this goal has been satisfied.

Table 33 Evidence for G3.3.1—Identifying Non-poolable Data Sets

G3.3.1	The validation error should be investigated to ensure that it does not contain any subgroups that are obviously not from the same population (i.e., non-poolable).
Level	Evidence
1	No subgroups were analyzed for poolability.
2	All relevant subgroups were investigated, and there was statistical evidence that the groups were from different populations. Therefore, the statistics from the limiting subgroup data set were used to determine the model’s uncertainty.
3	All relevant subgroups were investigated, and there was no statistical evidence that the groups were from different populations. The statistics from the combined data sets were used to determine the model’s uncertainty.

<u>Historical Evidence Levels for Reactor Safety Analysis</u>

Level 2 has been most commonly accepted by the NRC staff. If Level 1 were presented, it would generally call for additional work and justification to be acceptable in reactor safety analysis, as it is very common for models to have different predictive behavior over their application domain. Level 3 is also common, but not as often achieved, as there is usually a subgroup which is slightly more limiting than the others.

3.3.3.2 G3.3.2—Identifying Non-conservative Subregions

Another key assumption is the assumption of statistical independence of the data (i.e., independent and identically distributed or iid) in the expected domain. As Piepel and Cuta (1993) point out, the model’s uncertainty will likely vary over the expected domain. Therefore, an effort is made to determine whether any obvious non-conservative subregions can be identified in the validation error. The absence of such a subregion does not prove that statistical independence exists; however, the authors are not aware of any other means to make such a determination.

Historically, non-conservative subregions have been identified by reviewing plots of the validation error versus the various input parameters (e.g., pressure, mass flux, quality). The lack of a visual trend in these plots was the justification that the model’s uncertainty did not vary over the application domain. However, Kaizer (2015) points out that this visual one-dimensional (1-D) plotting method ignores dependences among the various parameters and that non-conservative subregions in the expected domain can be missed. Therefore, he proposed another method that can be used to analyze data in up to three dimensions at a time. Although this proposed method has a visual component to identify suspected non-conservative regions, it uses a statistical test to determine whether the subregion is, in fact, non-conservative.

Because this method is limited to three dimensions, only the most important input parameters are typically investigated together. For PWRs, those three parameters are typically the mass flux, pressure, and local quality. For BWRs, those three parameters are typically the mass flux, pressure, and inlet temperature (or subcooling). Other combinations should also be investigated as necessary.

Proving that non-conservative subregions do not exist is not the objective. Such proof would call for taking a very large number of data points. Given the limited data available in the validation error data set, the only statement that can be confirmed is that no obvious non-conservative subregion has been identified. However, if a non-conservative subregion is found, the model uncertainty in that region would need to be increased to reflect the model's predictive capability in that region.

Table 34 gives the evidence commonly provided to demonstrate that this goal has been satisfied.

Table 34   Evidence for G3.3.2—Identifying Non-conservative Subregions

G3.3.2	The expected domain should be investigated to determine if it contains any non-conservative subregions that would impact the predictive capability of the model.
Level	Evidence
1	Plots of each model input parameter versus the validation error (i.e., predicted versus measured or measured versus predicted) are provided. This visual method (e.g., the 1-D method) demonstrates that there are no trends in the validation error with any input parameter.
2	Plots of each model input parameter versus the validation error (i.e., predicted over measured or measured over predicted) are provided. This visual method (e.g., the 1-D method) demonstrates that there are no trends in the validation error with any input parameter. Additionally, a method similar to the one proposed by Kaizer (2015) is used to demonstrate that there are no obvious non-conservative subregions in the application domain.
3	A method further refined from the one proposed by Kaizer (2015) is used. Such a method is able to consider all N-dimensions at the same time and does not call for the user to visually identify any suspected non-conservative subregions.

<u>Historical Evidence Levels for Reactor Safety Analysis</u>

Level 1 has historically been most commonly accepted by the NRC staff. However, recent reviews have used Level 2. The method discussed in Level 2 has revealed multiple non-conservative subregions that required additional analysis or testing. Level 3 is ideal as it would be a completely objective; however, the authors are not currently aware of any such method.

3.3.3.3 G3.3.3—Appropriate Trends

Certain trends common in CHF and CP models could be expected in future models. Generally, these trends can be seen by analyzing the plots of CHF or CP versus each of the various model parameters. This includes both an examination of all the data at once and an examination of only selected portions of the data (e.g., CHF at nominal pressures with decreasing mass flux). Depending on the situation, the measured and predicted CHF and CP may need to be analyzed separately. Table 35 gives the evidence commonly provided to demonstrate that this goal has been satisfied.

Table 35 Evidence for G3.3.3—Appropriate Trends

G3.3.3	The model's predictions trend as expected in each of the various model parameters.
Level	Evidence
1	Plots of the validation error (i.e., predicted over measured or measured over predicted) versus each model input parameter are provided.
2	Plots of the measured or predicted CBT values versus each model input parameter are provided. All trends are as expected.
3	Plots of the measured and predicted CBT values versus each model input parameter are provided. All trends are as expected.

Historical Evidence Levels for Reactor Safety Analysis

Level 1 has been most commonly accepted by the NRC staff. The trends should not only be smooth and continuous, but also conform to known behavior of the associated phenomena. It is often helpful to compare the trends of the current model with trends from previously approved models. Generally, further details of this criterion are investigated only if inconsistent behavior in the expected domain is suspected.

3.3.4 G3.4—Calculating Model Uncertainty

In CHF models used in PWRs and CP models used in BWRs, the model uncertainty is obtained from the validation error. However, the means of calculating the model uncertainty and its application to reactor safety analysis vary greatly.

Departure from Nucleate Boiling Ratio Limit Used in Pressurized-Water Reactors

For CHF models used in PWRs, the model's uncertainty is applied in the DNBR limit. That limit is used to ensure that there will be *at least a 95-percent probability at the 95-percent confidence level that the hot fuel rod in the core does not experience a DNB or CBT condition during normal operation or AOOs*. This DNBR limit is solely dependent on the CHF model's performance and is independent of any conditions at the plant (e.g., the loading pattern).

The DNBR limit is a statistical limit derived from the validation error. The validation error (usually represented as a ratio of the measured CHF to the model predicted CHF) is assumed to be a representative sample from the population of the model application error. Therefore, the 95th percentile of the population of the model application error is estimated using the 95/95 value from the validation error. In other words, the 95th percentile of the validation error is estimated using a process that will overestimate the percentile 95 percent of the time (e.g., Owen's method (Owen, 1963) and Wilks' method (Wilks 1941; Wilks 1943)). This 95/95 value is then used as the DNBR limit and bounds the uncertainty of the CHF model.

For example, if the measured versus predicted values are normally distributed, the DNBR limit could be determined to be the 95/95 value calculated, as prescribed by Owen, with the k-value obtained from Owen's tables. Equation 5 is used to calculate the 95/95 value:

$$l_{95/95} = \mu - k \cdot \sigma \tag{5}$$

Where $l_{95/95}$ is the 95/95 value, μ is the mean of the measured to predicted values, k is a factor from Owen's tables, and σ is the standard deviation of the measured to predicted values. Generally, the DNBR limit is simply the reciprocal of the 95/95 value; however, a conservative bias is usually added. Generally, this bias (b) may simply come from rounding up the DNBR limit to a number with 3 significant figures (i.e., 1.133 becomes 1.14), but additional biases may also be added to account for other non-conservatism in the model (for more details see Information Notice 2014-01 (U.S. Nuclear Regulatory Commission, 2014)). Equation 6 gives the resulting DNBR limit:

$$DNBR\ Limit = \frac{1}{l_{95/95}} + b \tag{6}$$

Safety Limit Minimum Critical Power Ratio Used in Boiling-Water Reactors

For CP models used in BWRs, the safety limit minimum critical power ratio (SLMCPR) reflects the model's uncertainty. That limit is used *to ensure at least 99.9 percent of the fuel rods in the core will not experience a CBT during normal operation or AOOs*. Unlike the DNBR limit, the SLMCPR does not depend solely on the CP model's performance, but instead also depends on some conditions at the plant (e.g., the core design).

A separate methodology is used to determine the SLMCPR, and the uncertainty in the CP model is an input to that methodology. Usually, this uncertainty is represented by the standard deviation of the model's prediction of the experimental data[22] (i.e., the standard deviation of the validation error). This standard deviation is used to capture the model's uncertainty. If the mean of the validation error is greater than one or if the sample is not normal, than the model's uncertainty is increased by artificially increasing the standard deviation before it is used in the SLMCPR methodology. Mean values of less than one are generally not credited in determining the SLMCPR.

Conservative Calculation of the Model Statistics

The model's uncertainty is quantified using statistics from the validation error. Those statistics are estimates of the parameters from the population of the model application error. Thus, the statistics of the validation error should be calculated in such a manner that they bound the true model application error. The three subgoals in Figure 15 are used to demonstrate that the validation error has been appropriately quantified.

[22] It should be noted that in PWRs, the validation error is given as the ratio of the measured value to the predicted value, but in BWRs the validation error is usually given as the ratio of the predicted value to the measured value.

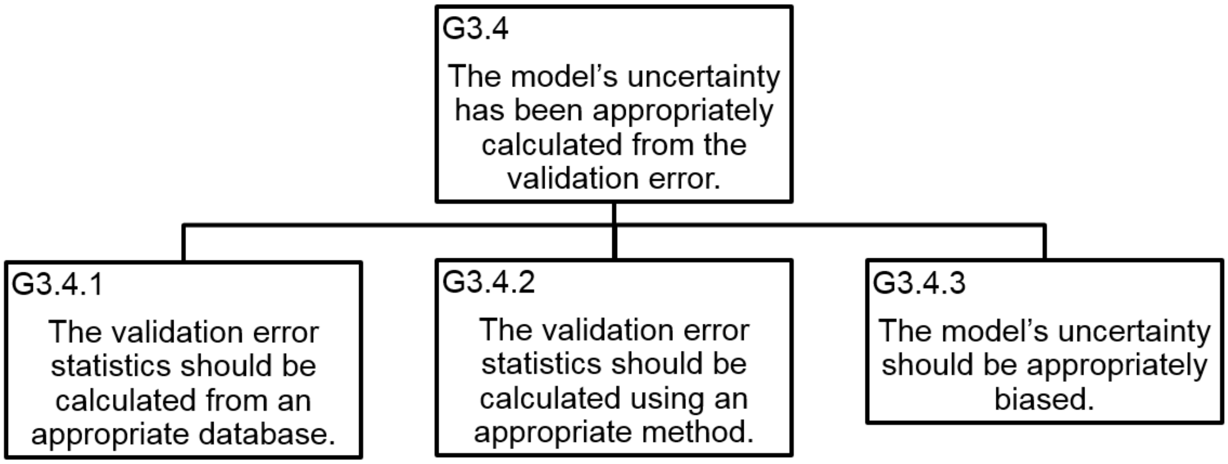


Figure 15 Decomposition of G3.4—Quantification of the Model's Error

No further decompositions of the subgoals were deemed useful. Therefore, the sections below discuss the evidence which could be used to demonstrate that the three base goals have been satisfied. Additionally, a discussion is provided on the evidence which has been historically used for CBT models applied in reactor safety analysis.

3.3.4.1 G3.4.1—Error Database

It may not be appropriate to use the entire validation error database to calculate the model's statistics, especially if the expected domain has non-poolable data sets or non-conservative subregions. Therefore, the assessor should confirm that the statistics used to generate the validation error are from an appropriate sample of data. Table 36 gives the evidence commonly provided to demonstrate that this goal has been satisfied.

Table 36 Evidence for G3.4.1—Error Database

G3.4.1	The validation error statistics should be calculated from an appropriate database.
Level	Evidence
1	The model's uncertainty was calculated using the entire database of validation error.
2	The model's uncertainty was calculated using a subset of the validation error, which resulted in a more conservative calculation.
3	The model's uncertainty was calculated from the limiting subset of the validation error, which resulted in a more conservative calculation.

<u>Historical Evidence Levels for Reactor Safety Analysis</u>

Level 1 has been most commonly accepted by the NRC staff, but it generally assumes the data are poolable and does not contain any non-conservative subregions. Level 2 is often provided if it appears that a subset may be more limiting, but there is no definitive proof. Generally, if definitive proof exists that a specific subset if most limiting, then the uncertainty is often calculated from only the data in that subset (Level 3).

3.3.4.2 G3.4.2—Validation Error Statistics

The method used to calculate the validation error statistics should be appropriate. This generally means ensuring that the assumptions of any method used are fulfilled (e.g., if Owen's method is used to calculate the 95/95 value, the distribution of the validation error should be normally distributed). Statistical methods may call for the data (i.e., the validation error) to (1) have the same mean and variance (i.e., homoscedasticity), (2) be from the same distribution, (3) be from a normal distribution, and (4) be independent and identically distributed (i.e., iid data). If populations within the data do not have the same mean or variance, a conservative mean or variance can be chosen to bound the model uncertainty. If the data are not normally distributed, a nonparametric method (such as the Wilks method) can be used to calculate the model uncertainty. However, if the data are not independent and identically distributed, the model's predictive capability would vary depending on the location in the application domain, and the model's uncertainty would have to account for this variability. Table 37 gives the evidence commonly provided to demonstrate that this goal has been satisfied.

Table 37 Evidence for G3.4.2—Validation Error Statistics

G3.4.2	The validation error statistics should be calculated using an appropriate method.
Level	Evidence
1	The data used to calculate the model's uncertainty appear to be independent and identically distributed. The method used to calculate the statistics is a parametric method. Although the necessary preconditions of such a method were not satisfied, assumptions could be made to ensure that the resulting uncertainty was conservative.
2	The data used to calculate the model's uncertainty appear to be independent and identically distributed, and one of the following applies: • The method used to calculate the statistics is a parametric method. The assumptions of such a method were demonstrated to be true (i.e., there is no reason to believe they are false) through statistical testing. • The method used to calculate the statistics is a nonparametric method.

<u>Historical Evidence Levels for Reactor Safety Analysis</u>

Level 2 has been most commonly accepted by the NRC staff. However, Level 1 has been used when the resulting statistics could be justified to be conservative.

3.3.4.3 G3.4.3—Model Uncertainty Bias

After the model's uncertainty is calculated, it is commonly biased in a conservative direction. For example, a vendor may want to use a three-digit number as the DNBR limit. Thus, if the DNBR limit were calculated as 1.2301, it would be "rounded" up to 1.24 (which is equivalent to the addition of a bias of 0.0099). However, sometimes a bias is added to account for an uncertainty that the model does not address. Table 38 gives the evidence commonly provided to demonstrate that this goal has been satisfied.

Table 38 Evidence for G3.4.3—Model Uncertainty Bias

G3.4.3	The model’s uncertainty should be appropriately biased.
Level	Evidence
1	The model needed a large bias (> 1%).
2	The model needed a small bias (< 1%).
3	The model needed no bias, or the only biasing was due to rounding.

<u>Historical Evidence Levels for Reactor Safety Analysis</u>

Level 3 has most commonly been accepted by the NRC staff, but Level 2 has also been commonly accepted. A larger bias (i.e., greater than 1 percent) indicates that some uncertainty in the model was not accounted for, which is generally avoided. In addition, because such biases are generally applied based on engineering judgment, not experimental data, the bias itself is subjective. Although situations arise that warrant the use of large biases, it is far from desirable because there is often little justification for choosing the specific bias instead of a larger or smaller value.

3.3.5 G3.5—Model Implementation

Once the model’s uncertainty has been quantified by experimental data, the model can be applied in a reactor safety analysis. However, the implementation of the model in the analysis should be consistent with its use during validation.

For some CBT models, this may mean that the same computer code is used in both the validation and the application of the models. Although certain inputs to the CBT model (e.g., pressures, flow rates, power) would be expected to change depending on the situation in which the model was used, those inputs may depend less on the situation and more on which closure models were selected in the computer code that exercises the CBT model. In those situations, it may be possible to change the inputs to the CBT model without changing the inputs to the computer code itself (e.g., plant conditions) but merely by changing the closure models chosen. Therefore, it is important to ensure that if the inputs to the CBT model depend on closure models in the computer code that implements the CBT model, the same closure models are used in both the validation and the application of the CBT model.

The reason for this is that the CBT model was validated with those closure models being applied, and the uncertainty was quantified using only that set of closure models. The CBT model could be used with another set of closure models, but the uncertainty would need to be quantified again (i.e., determine the new validation error with the new closure models). Re-validation of the model and re-quantification of the uncertainty is not necessarily a major exercise. The experimental data already exist, and a new data set of validation errors can be obtained using the changed model, code, or options. For the framework discussed here, only Criteria G3.2.1 and G3.2.2 would likely need to be confirmed because the evidence supplied to justify all other criteria would likely remain the same; however, this should be borne out by the analysis.

If the model's prediction in the changed code is similar to its prediction in the previous code, the evidence used to justify these two criteria may even remain the same. The three subgoals in Figure 16 are used to demonstrate that the model has been correctly implemented.

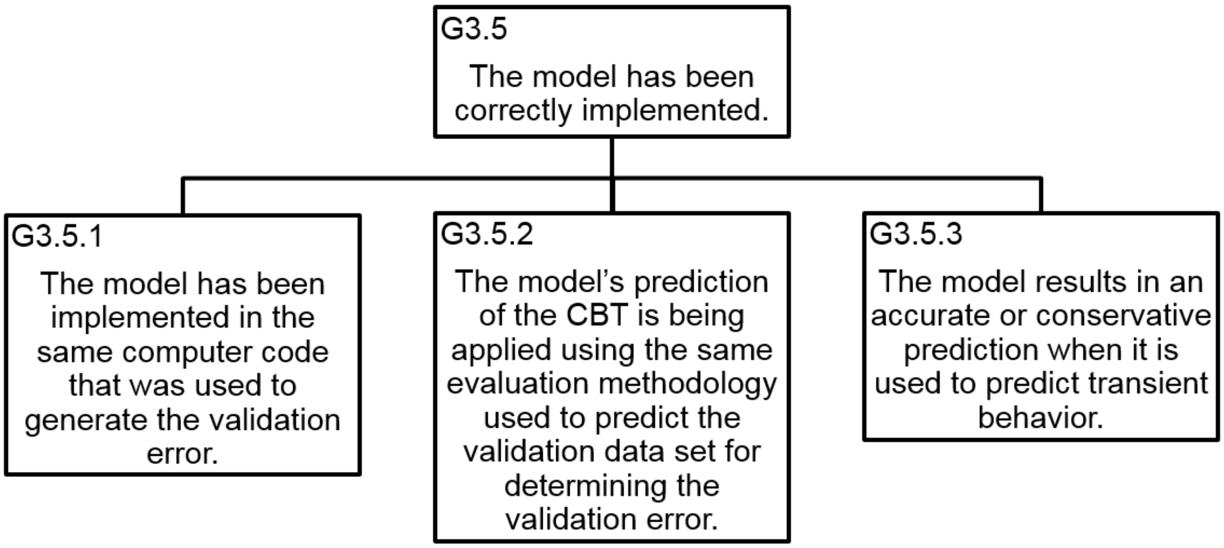


Figure 16 Decomposition of G3.5—Model Implementation

No further decompositions of the subgoals were deemed useful. Therefore, the sections below discuss the evidence which could be used to demonstrate that these three base goals have been satisfied. Additionally, a discussion is provided on the evidence which has been historically used for CBT models applied in reactor safety analysis.

3.3.5.1 G3.5.1—Same Computer Code

The computer code and the options used to specify the closure models and any other functionality of that computer code should be the same. This is a much larger concern in PWRs because a subchannel simulation contains many more uses of the field equations and closure models. The direct modeling of the thermal-hydraulic response of the assembly should be consistent from the validation to the application of a CBT model. Table 39 gives the evidence commonly provided to demonstrate that this goal has been satisfied.

Table 39 Evidence for G3.5.1—Same Computer Code

G3.5.1	The model has been implemented in the same computer code that was used to generate the validation error.
Level	Evidence
1	The model has been implemented in a computer code very similar to the one that was used to generate the validation error.
2	The same computer code with the same closure models and code options that was used to generate the validation error will be used to perform any reactor safety analysis.

Historical Evidence Levels for Reactor Safety Analysis

Level 1 has been most commonly accepted by the NRC staff for BWRs and Level 2 for PWRs. Many CBT models used for BWRs calculate the critical power of an assembly, and therefore the computer code used generally does not calculate complex local thermal-hydraulic phenomena, but rather more general parameters like assembly quality. As these formulations rely on few closure models, it is possible to use the same CBT model in multiple BWR analysis codes. However, because PWR analysis is performed at the sub-channel level, there are multiple closure models used. Those closure models calculate the local parameters that are used by the CBT model. Because the CBT model predictions could be changed by changing the closure models, changing a computer code generally involves re-analyzing the validation data with the new code for the PWRs.

3.3.5.2 G3.5.2—Same Evaluation Methodology

It is important not only to use the same computer code to implement the CBT model but also to implement the model in the same manner. Although the comparison of the "measured" values to the "predicted" values is the basis for the validation, as discussed above, this comparison is generally not as simple as comparing the CHF or CP at the location in the test assembly that experienced a CBT to the predicated value at that location; therefore, a distinction was drawn between *model error* and *model application error*. Another way of referring to the same evaluation methodology is to ensure that the manner in which the model will be used (i.e., model application error), is consistent with how the validation error was determined.

Section 3.3, which defines validation error, discusses the reasoning for this. Here, the authors will only reiterate that the goal of using the CBT model is to ensure that a CBT does not occur, not to ensure that, if a CBT does occur, the model predicts the exact location where it occurs. If the model is able to identify the location, that can be evidence that the model is well correlated with the physics of the assembly, but it is not a requirement and, moreover, may not be useful when determining whether the model is appropriate.

Table 40 gives the evidence commonly provided to demonstrate that this goal has been satisfied.

Table 40 Evidence for G3.5.2—Same Evaluation Methodology

G3.5.2	The model's prediction of the CBT is being applied using the same evaluation methodology used to predict the validation data set for determining the validation error.
Level	Evidence
1	The model is implemented using a very similar evaluation methodology.
2	The model is implemented using the same evaluation methodology.

Historical Evidence Levels for Reactor Safety Analysis

Level 2 has been most commonly accepted by the NRC staff. Level 1 could be used if an analysis demonstrated that the changes would not affect the model's uncertainty.

3.3.5.3 G3.5.3—Transient Prediction

Like many other thermal-hydraulic models, many CBT models are developed using data taken under steady-state conditions but applied in a transient[23] simulation. Although this is a common practice, it should be justified, especially for models that contain integrals over space or time. This is generally more of a focus in BWRs than it is in PWRs and is ultimately demonstrated through transient tests. Those tests generally use time-varying inputs for power, flow, subcooling, or a combination of these parameters. The goal is to demonstrate that there are no transients at which a CBT occurred but was not predicted and, secondarily, to demonstrate that there were no tests in which CBT was predicted (i.e., should have occurred) but did not occur.

Again, the goal of these tests is to demonstrate how well the model predicts whether CBT will occur; therefore, these transient tests should be conducted close to conditions that cause a CBT. Tests that are run too far from those conditions in either direction (i.e., either a test that was very far from a CBT actually occurring such as a very low-power test or a test in which a CBT must occur such as a very high-power test) would not be useful.

Table 41 gives the evidence commonly provided to demonstrate that this goal has been satisfied.

Table 41 Evidence for G3.5.3—Transient Prediction

G3.5.3	The model results in an accurate or conservative prediction when it is used to predict transient behavior.
Level	Evidence
1	No experimental justification is provided.
2	Some experimental justification is provided.
3	Statistically significant experimental justification is provided.

<u>Historical Evidence Levels for Reactor Safety Analysis</u>

Level 1 or Level 2 have been commonly accepted by the NRC staff for PWRs, whereas Level 2 or Level 3 have been commonly accepted for BWRs. The reason for the additional testing is likely due to the manner in which CBT is modeled differently in BWRs and PWRs. In PWRs, the CBT is based on local sub-channel parameters. Historically, it has been shown that CBT models which are generated with steady state data will accurately or conservatively predict CBT during a transient. This same assumption about CBT models being made with steady state data being conservative for transient data are also made for models used in BWRs. However, while it is possible to confirm this assumption through testing on a BWR test assembly, such testing would be very difficult on a PWR test assembly.

For Level 3, by “statistically significant,” the authors mean that there were enough conservative predictions from transients (i.e., those in which the CBT model was correct) to account for any situations in which the CBT model may have been non-conservative. For example, if the CBT model was non-conservative in a single test but conservative in only eight tests, its predictive

[23] In this context, transient means “time varying.”

capability would be in question, as eight tests is generally considered to be too small a number to determine any statistical significance.

4 SUMMARY AND CONCLUSION

This work presents a generic safety case that can be used to determine the credibility of CBT models. This safety case was generated through the experience of many experts at the NRC, previously written safety evaluations, and documents in the open literature. This document captures the knowledge and experience of multiple NRC staff members over many years. The document presents a background on CBT including a literature survey, a description of the underlying phenomena, and how those phenomena are commonly modeled. The document also presents a credibility assessment framework, which combines the structure from GSN with the capability of maturity assessment. The elements of the framework provided in this document have been applied in multiple reviews at the NRC and have decreased total review time, increased review consistency, and increased review efficiency.

5 REFERENCES

[1] Hewitt, G.F., and N.S. Hall-Taylor, *Annular Two-Phase Flow*, Pergamon Press, Oxford, United Kingdom, 1970.

[2] Tong, L.S., and Y.S. Tang, *Boiling Heat Transfer and Two-Phase Flow,* Second Edition, Taylor & Francis Group, Washington, DC, 1997.

[3] ASME. *Guide for verification and validation in computational solid mechanics*. ASME V&V 10-2006, New York, NY, American Society of Mechanical Engineers, 2006.

[4] NASA. "Towards a Credibility Assessment of Models and Simulations," *49^{th} AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*. AIAA-2008-2156, 2008.

[5] NASA[B]. *Standard for Models and Simulations*. NASA-STD-7009, Washington, DC, National Aeronautics and Space Administration, 2008.

[6] Oberkampf, W.L., and C.J. Roy, *Verification and Validation in Scientific Computing*, Cambridge University Press, Cambridge, United Kingdom, 2010.

[7] Kaizer, J.S., A.K. Heller, and W.L. Oberkampf, "Scientific Computer Simulation Review," *Reliability Engineering and System Safety*, 138:210–218, 2015.

[8] Ministry of Defense, Defense Standard 00-56, "Safety Management Requirements for Defense Systems, Part 1: Requirements," Issue 4, 2007.

[9] Denney, E.W., and Pai, G.J., "Safety Case Patterns: Theory and Applications", NASA/TM–2015–218492, Feburary 2015.

[10] Food and Drug Administration, "Infusion Pumps Total Product Life Cycle - Guidance for Industry and FDA Staff", December 2, 2014.

[11] Goal Structure Notation (GSN) Working Group, GSN Community Standard Version 1, Origin Consulting Limited, York, United Kingdom, 2011.

[12] Denney, E., G. Pai, and I. Habli, "Towards Measurement of Confidence in Safety Cases," *Proceedings of the 5^{th} International Symposium on Empirical Software Engineering and Measurement*, pp. 380–383, September 2011.

[13] Oberkampf, W.L., M. Pilch, and T.G. Trucano, *Predictive Capability Maturity Model for Computational Modeling and Simulation*. SAND2007-5948, Albuquerque, NM, Sandia National Laboratories, 2007.

[14] Tong, L.S., *Boiling Crisis and Critical Heat Flux*, U.S. Atomic Energy Commission, 1972.

[15] Todreas, N.E., and M.S. Kazimi, *Nuclear Systems I: Thermal Hydraulic Fundamentals*, Taylor & Francis Group, Washington, DC, 1990.

[16] Lahey, R.T., and F.J. Moody, *The Thermal Hydraulics of a Boiling Water Nuclear Reactor—Second Edition*, American Nuclear Society, La Grange Park, IL, 1993.

[17] Leidenfrost, J.G., "On the Fixation of Water in Diverse Fire," *A Tract about some qualtiies of common water*, 1756, translated, Wares, C., *International Journal of Heat Mass Transfer*, 9:1153, 1966.

[18] Tong, L.S., H.B. Currin, P.S. Larsen, and O.G. Smith, "Influence of Axially Nonuniform Heat Flux on DNB," *AIChE Chemical Engineering Symposium Series*, 62(64):35–40, 1965.

[19] Macbeth, R.V., "An Appraisal of Forced Convection Burnout Data," *Proceedings of the Institute of Mechanical Engineers*, 1965–1966.

[20] Barnett, P.G., "A Correlation of Burnout Data for Uniformly Heated Annuli and Its Use for Predicting Burnout in Uniformly Heated Rod Bundles," Atomic Energy Establishment Winfrith (AEEW-R) 463, 1966.

[21] Healzer, J.M., J.E. Hench, E. Janssen, and S. Levy, "Design Basis for Critical Heat Flux Condition in Boiling Water Reactors," APED-5, September 1966.

[22] Tong, L.S., "Prediction of Departure from Nucleate Boiling for an Axially Non-Uniform Heat Flux Distribution," *Journal of Nuclear Energy*, 21:241–248, 1967.

[23] Biasi, L., G.S. Clerici, S. Garribba, R. Sala, and A. Tozzi, "Studies on Burnout: Part 3—A New Correlation for Round Ducts and Uniform Heating and Its Comparison with World Data," *Energia Nucleare* 14:530–536, 1967.

[24] Gellerstedt, J.S., R.A. Lee, W.J. Oberjohn, R.H. Wilson, and L.J. Stanek, "Correlation of Critical Heat Flux in a Bundle Cooled by Pressurized Water," *Two-Phase Flow and Heat Transfer in Rod Bundles,* American Society of Mechanical Engineers, New York, NY, 1969.

[25] Hughes, E.D., "A Correlation of Rod Bundle Critical Heat Flux for Water in the Pressure Range 150 to 725 psia," IN-1412, Idaho Nuclear Corporation, Idaho Falls, ID, 1970.

[26] Piepel, G.F., and J. M. Cuta, "Statistical Concepts and Techniques for Developing, Evaluating, and Validating CHF Models and Corresponding Fuel Design Limits," *SKI Technical Report,* 93:46, 1993.

[27] Groeneveld, D.C., J. Shan, A.Z. Vasic, L.K.H. Leung, A. Durmayaz, J. Yang, S.C. Cheng, and A. Tanase, "The 2006 CHF Look-Up Table," *Nuclear Engineering and Design*, 237(15):1909–1922, 2007.

[28] Yang, B., J. Shan, J. Gou, H. Zhang, A. Liu, and H. Mao, "Uniform versus Nonuniform Axial Power Distribution in Rod Bundle CHF Experiments," *Science and Technology of Nuclear Installations*, Volume 2014, 2014.

[29] Kaizer, J.S., "Identification of Nonconservative Subregions in Empirical Models Demonstrated Using Critical Heat Flux Models," *Nuclear Technology*, 190:65–71, 2015.

[30] Groeneveld, D.C., "CHF Data Used to Generate 2006 Groeneveld CHF Lookup Tables," NUREG/KM-0011, U.S. Nuclear Regulatory Commission, Washington, DC, 2017.

[31] Babcock & Wilcox, "Correlation of Critical Heat Flux in a Bundle Cooled by Pressurized Water," BAW-10000, Babcock & Wilcox, Lynchburg, VA, March 1970, Agencywide Documents Access and Management System (ADAMS) Accession No. ML082490748 (*Proprietary Information, Nonpublicly Available*).

[32] Combustion Engineering (C-E), "C-E Critical Heat Flux: Critical Heat Flux Correlation for C-E Fuel Assemblies with Standard Spacer Grids, Part 1—Uniform Axial Power Distribution," CENPD-162-P-A, C-E, Stamford, CT, September 1976, ADAMS Accession No. ML083010357 (*Proprietary Information, Nonpublicly Available*).

[33] Exxon Nuclear Company, "Exxon Nuclear DNB Correlation for PWR fuel Designs," XN-NF-621(P)(A), Revision 1, Richland, WA, September 1983, ADAMS Accession No. ML16265A315 (*Proprietary Information, Nonpublicly Available*).

[34] Motley, F.E., K.W. Hill, F.F. Cadek, and J. Shefcheck, "New Westinghouse Correlation WRB-1 for Predicting Critical Heat Flux in Rod Bundles with Mixing Vane Grids,"

WCAP-8762-P-A, Westinghouse Electric Company, Pittsburgh, PA, July 1984, ADAMS Accession No. ML080630433 (*Proprietary Information, Nonpublicly Available*).

[35] Westinghouse Electric Company, "VANTAGE 5H Fuel Assembly," WCAP-10444-P-A, Pittsburgh, PA, September 1985, ADAMS Accession No. ML080650257 (*Proprietary Information, Nonpublicly Available*).

[36] C-E, "C-E Critical Heat Flux: Critical Heat Flux Correlation for C-E Fuel Assemblies with Standard Spacer Grids, Part 2—Non-Uniform Axial Power Distribution," CENPD-207-P-A, Stamford, CT, December 1984, ADAMS Accession No. ML16260A362 (*Proprietary Information, Nonpublicly Available*).

[37] AREVA, "Departure from Nucleate Boiling Correlation for High Thermal Performance Fuel," ANF-1224(P)(A), and Supplement 1 to ANF-1224(P)(A), Lynchburg, VA, April 1990.

[38] Farnswoth, D.A., and G.A. Meyer, "The BWU Critical Heat Flux Correlations," BAW-10199P-A, Revision 0, Framatome Technologies, Lynchburg, VA, February 1996.

[39] Smith III, L.D., M.W Lloyd, Y.X. Sung, and W.J. Leech, "Modified WRB-2 Correlation, WRB-2M, for Predicting Critical Heat Flux in 17x17 Rod Bundles with Modified LPD Mixing Vane Grids," WCAP-15025-P-A, Westinghouse Electric Company, Pittsburgh, PA, April 1999, ADAMS Accession No. ML081610106 (*Proprietary Information, Nonpublicly Available*).

[40] Farnsworth, D.A., and G.A. Meyer, "The BWU Critical Heat FIux Correlations Applications to the Mark-B11 and Mark-BW17 MSM Designs," BAW-10199P-A, Addendum 1, Framatome Cogema Fuels, Lynchburg, VA, December 2000, ADAMS Accession No. ML003777245 (*Proprietary Version, Nonpublicly Available*).

[41] Farnsworth, D.A., and G.A. Meyer, "Application of BWU-Z CHF Correlation to the Mark-BW 17 Fuel Design with Mid-Span Mixing Grids," BAW-10199P-A, Addendum 2, Framatome ANP, Lynchburg, VA, September 2002, ADAMS Accession Nos. ML022560552 (*Proprietary Version, Nonpublicly Available*) and ML022560550 (*Nonproprietary Version, Publicly Available*).

[42] Sung, Y.X., P.F. Joffre, and P.A. Hilton, "Addendum 1 to WCAP-1 4565-P-A Qualification of ABB Critical Heat Flux Correlations with VIPRE-01 Code," WCAP-14565-P-A Addendum 1-A, Westinghouse Electric Company, Pittsburgh, PA, August 2004, ADAMS Accession Nos. ML042610371 (*Proprietary Version, Nonpublicly Available*) and ML042610368 (*Nonproprietary Version, Publicly Available*).

[43] AREVA, "Departure from Nucleate Boiling Correlation for High Thermal Performance Fuel," EMF-92-153(P)(A), Revision 1, January 2005, ADAMS Accession Nos. ML051020019 (*Proprietary Version, Nonpublicly Available*) and ML051020017 (*Nonproprietary Version, Publicly Available*).

[44] Farnsworth, D., and K.R. Greene, "BHTP DNB Correlation Applied with LYNXT," BAW-10241(P)(A), Revision 1, AREVA, Lynchburg, VA, July 2005, ADAMS Accession Nos. ML052500092 (*Proprietary Version, Nonpublicly Available*) and ML052500075 (*Nonproprietary Version, Publicly Available*).

[45] Farnsworth, D., "The BWU-B11R CHF Correlation for the Mark-B11 Spacer Grid," BAW-10199P-A, Addendum 3, Framatome ANP, Lynchburg, VA, November 2005, ADAMS Accession No. ML070170690 (*Proprietary Version, Nonpublicly Available*) ML042990354 (*Nonproprietary Version, Publicly Available*).

[46] Joffre, P.F., Y.R. Chang, R. Kapoor, Y.X. Sung, L.D. Smith III, and P.A. Hilton, "Westinghouse Correlations WSSV and WSSV-T for Predicting Critical Heat Flux in Rod Bundles with Side Supported Mixing Vanes," WCAP-16523-P-A, Westinghouse Electric Company, Pittsburgh, PA, August 2007, ADAMS Accession Nos. ML072570633 (*Proprietary Version, Nonpublicly Available*) and ML072570327 (*Nonproprietary Version, Publicly Available*).

[47] Farnsworth, D.A., and R.L. Harne, "The ACH-2 CHF Correlation for the U.S. EPR," ANP-10269P-A, AREVA, Lynchburg, VA, December 2007, ADAMS Accession Nos. ML080790191 (*Proprietary Version, Nonpublicly Available*) and ML080790193 (*Nonproprietary Version, Publicly Available*).

[48] Joffre, P.F., R. Kapoor, Y.X. Sung, and P.A. Hilton, "Addendum 2 to WCAP-14565-P-A Extended Application of ABB-NV Correlation and Modified ABB-NV Correlation WLOP for PWR Low Pressure Applications," WCAP-14565-P-A, Addendum 2-P-A, Westinghouse Electric Company, Pittsburgh, PA, April 2008, ADAMS Accession Nos. ML081280713 (*Proprietary Version, Nonpublicly Available*) and ML081280712 (*Nonproprietary Version, Publicly Available*).

[49] Joffre, P.F., Y.X. Sung, R. Mathur, L.D. Smith III, and P.A. Hilton, "Westinghouse Next Generation Correlation (WNG-1) for Predicting Critical Heat Flux in Rod Bundles with Split Vane Mixing Grids," WCAP-16766-P-A, Westinghouse Electric Company, Pittsburgh, PA, February 2010, ADAMS Accession Nos. ML100850532 (*Proprietary Version, Nonpublicly Available*) and ML100850528 (*Nonproprietary Version, Publicly Available*).

[50] Mitsubishi Heavy Industries, Ltd., "Thermal Design Methodology," MUAP-07009-P-A, Tokyo, Japan, August 2013, ADAMS Accession Nos. ML13284A072 (*Proprietary Version, Nonpublicly Available*) and ML13284A069 (*Nonproprietary Version, Publicly Available*).

[51] Korea Hydro & Nuclear Power Company, "KCE-1 Critical Heat Flux Correlation for PLUS7 Thermal Design," APR1400-F-C-TR-12002-P, Gyeongju, South Korea, November 2012, ADAMS Accession Nos. ML13018A158 (*Proprietary Version, Nonpublicly Available*) and ML13018A147 (*Nonproprietary Version, Publicly Available*).

[52] AREVA, "The ORFEO-GAIA and ORFEO-NMGRID Critical Heat Flux Correlations," ANP-10341(P), Lynchburg, VA, August 2016, ADAMS Accession Nos. ML16238A076 (*Proprietary Version, Nonpublicly Available*) and ML16238A078 (*Nonproprietary Version, Publicly Available*).

[53] Slifer, B.C., and J.E. Hench, "Loss-of-Coolant Accident and Emergency Core Cooling Models for General Electric Boiling Water Reactors," NEDO-10329, Equation C-32, General Electric Company, San Jose, CA, April 1971.

[54] General Electric Company, "General Electric BWR Thermal Analysis Basis (GETAB): Data, Correlation and Design Application," NEDO-10958-PA, San Jose, CA, January 1977, ADAMS Accession Nos. ML092820214 (*Proprietary Version, Nonpublicly Available*) and ML102290144 (*Nonproprietary Version, Publicly Available*).

[55] AREVA, "ANFB Critical Power Correlation," ANF-1125(P)(A), and Supplements 1 and 2, Richalnd, WA, April 1990, ADAMS Accession No. ML081820434 (*Proprietary Information, Nonpublicly Available*).

[56] General Electric Hitachi Nuclear Energy, "R-Factor Calculation Method for GE11, GE12, and GE13 Fuel," NEDC-32505P-A, Revision 1, Wilmington, NC, July 1999, ADAMS Accession No. ML060520637 (*Proprietary Information, Nonpublicly Available*).

[57] Harris, W.R., and Y.Y. Yung, "10x10 SVEA Fuel Critical Power Experiments and CPR Correlations: SVEA-96+," CENPD-389-P-A, ABB Combustion Engineering Nuclear Power, Inc., Windsor, CT, September 1999, ADAMS Accession Nos. ML993470286 (*Proprietary Version, Nonpublicly Available*) and ML993420024 (*Nonproprietary Version, Publicly Available*).

[58] Harris, W.R., and Y.Y. Yung, "10x10 SVEA Fuel Critical Power Experiments and CPR Correlations: SVEA-96," CENPD-392-P-A, Revision 00, CE Nuclear Power, LLC, Monroeville, PA, September 2000, ADAMS Accession Nos. ML003767392 (*Proprietary Version, Nonpublicly Available*) and ML003767366 (*Nonproprietary Version, Publicly Available*).

[59] Harrington, R., and J.G.M. Anderson, "GEXL96 Correlation for ATRIUM-9B Fuel," NEDC-32981P, Global Nuclear Fuel, Wilmington, NC, September 2001, ADAMS Accession Nos. ML003755947 (*Proprietary Version, Nonpublicly Available*), ML012490537 (*Safety Evaluation*, *Proprietary Version, Nonpublicly Available*), and ML012670193 (*Nonproprietary Version, Publicly Available*).

[60] Harrington, R., and J.G.M. Anderson, "GEXL10 Correlation for GE12 Fuel," NEDC-32464P, Revision 2, Global Nuclear Fuel, Wilmington, NC, September 2001, ADAMS Accession No. ML012760512 (*Proprietary Information, Nonpublicly Available*).

[61] Harrington, R., "GEXL80 Correlation for SVEA96+ Fuel," NEDC-33107P-A, Revision 1, Global Nuclear Fuel, Wilmington, NC, October 2004, ADAMS Accession Nos. ML043210062 (*Proprietary Version, Nonpublicly Available*) and ML043210058 (*Nonproprietary Version, Publicly Available*).

[62] Harris, W., M. Majed, G. Norback, and Y.Y. Yung, "10x10 SVEA Fuel Critical Power Experiments and CPR Correlation: SVEA-96 Optima2," WCAP-16081-P-A, Westinghouse Electric Company, Pittsburgh, PA, March 2005, ADAMS Accession No. ML051260213 (*Proprietary Information, Nonpublicly Available*) ML003676083 (*Nonproprietary Version, Publicly Available*).

[63] Global Nuclear Fuel, "GEXL97 Correlation Applicable to ATRIUM-10 Fuel," NEDC-33383P, Revision 1, Wilmington, NC, June 2008, ADAMS Accession No. ML082070090 (*Proprietary Information, Nonpublicly Available*) and ML082070088 (*Nonproprietary Version, Publicly Available*).

[64] Norback, G., and W. Harris, "SVEA-96 Optima2 CPR Correlation (D4): Modified R-factors for Part-Length Rods," WCAP-16081-P-A, Addendum 2-A, Westinghouse Electric Company, Pittsburgh, PA, February 2009, ADAMS Accession Nos. ML072200243 (*Proprietary Version, Nonpublicly Available*) and ML072200242 (*Nonproprietary Version, Publicly Available*).

[65] Norback, G., and W. Harris, "SVEA-96 Optima2 CPR Correlation (D4): High and Low Flow Applications," WCAP-16081-P-A, Addendum 1-A, Westinghouse Electric Company, Pittsburgh, PA, March 2009, ADAMS Accession Nos. ML091060144 (*Proprietary Version, Nonpublicly Available*) and ML091060143 (*Nonproprietary Version, Publicly Available*).

[66] Global Nuclear Fuel, "GEXL17 Correlation for GNF2 Fuel," NEDC-33292P, Revision 3, Wilmington, NC, June 2009, ADAMS Accession No. ML091830641 (*Proprietary Information, Nonpublicly Available*) and ML091830624 (*Nonproprietary Version, Publicly Available*).

[67] AREVA, "SPCB Critical Power Correlation," EMF-2209(NP)(A), Revision 3, Richland, WA, September 2009, ADAMS Accession Nos. ML093650230 (*Proprietary Version,*

Nonpublicly Available) and ML093650235 (*Nonproprietary Version, Publicly Available*) and ML111290532 (*Nonproprietary Version, Publicly Available*).

[68] Global Nuclear Fuel, "GEXL14 Correlation for GE 14 Fuel," NEDC-32851P-A, Revision 5, Wilmington, NC, April 2011, ADAMS Accession No. ML111290535 (*Proprietary Information, Nonpublicly Available*).

[69] AREVA, "ACE/ATRIUM-10 Critical Power Correlation," ANP-10249(P)(A), Revision 2, Richland, WA, March 2014, ADAMS Accession Nos. ML14175A226 *(Part 1 of 3, Proprietary Version, Nonpublicly Available*); ML14175A227 (*Part 2 of 3*, *Proprietary Version, Nonpublicly Available*); ML14175A228 (*Part 3 of 3*, *Proprietary Version, Nonpublicly Available*), and ML14175A229 (*Nonproprietary Version, Publicly Available*).

[70] AREVA, "ACE/ATRIUM 10XM Critical Power Correlation," ANP-10298(P)(A), Revision 1, Richland, WA, March 2014, ADAMS Accession Nos. ML14183A739 (*Part 1 of 3*, *Proprietary Version, Nonpublicly Available*); ML14183A743 (*Part 2 of 3, Proprietary Version, Nonpublicly Available*); ML14183A748 (*Part 3 of 3*, *Proprietary Version, Nonpublicly Available*); and ML14183A734 (*Nonproprietary Version, Publicly Available*).

[71] Bergmann, U., M. Hemlin, K. Bergman, and J-M. Le Corre, "10x10 SVEA Fuel Critical Power Experiments and New CPR Correlation: D5 for SVEA-96 Optima3," WCAP-17794-P, November 2013, ADAMS Accession Nos. ML13333A276 (*Proprietary Version, Nonpublicly Available*) and ML13333A275 (*Nonproprietary Version, Publicly Available*).

[72] AREVA, "ACE/ATRIUM-11 Critical Power Correlation," ANP-10335PP, Richland, WA, February 2015, ADAMS Accession Nos. ML15062A552 *(Part 1 of 2*, *Proprietary Version, Nonpublicly Available*); ML15062A555 (*Part 2 of 2*, *Proprietary Version, Nonpublicly Available*); and ML15062A554 (*Nonproprietary Version, Publicly Available*).

[73] *U.S. Code of Federal Regulations*, "General Design Criteria for Nuclear Power Plants," Appendix A to "Domestic Licensing of Production and Utilization Facilities," Part 50, Chapter I, Title 10, "Energy."

[74] *U.S. Code of Federal Regulations*, "ECCS Evaluation Models," Appendix K to "Domestic Licensing of Production and Utilization Facilities," Part 50, Chapter I, Title 10, "Energy."

[75] *U.S. Code of Federal Regulations*, "Contents of Applications; Technical Information," Section 34 of "Domestic Licensing of Production and Utilization Facilities," Part 50, Chapter I, Title 10, "Energy."

[76] *U.S. Code of Federal Regulations*, "Quality Assurance Criteria for Nuclear Power Plants and Fuel Reporcessing Plants," Appendix B to "Domestic Licensing of Production and Utilization Facilities," Part 50, Chapter I, Title 10, "Energy."

[77] U.S. Nuclear Regulatory Commission, "Fuel System Design," Section 4.2 of NUREG-0800, "Standard Review Plan for the Review of Safety Analysis Reports for Nuclear Power Plants: LWR Edition," Revision 3, March 2007, ADAMS Accession No. ML070740002.

[78] U.S. Nuclear Regulatory Commission, "Thermal and Hydraulic Design," Section 4.4 of NUREG-0800, "Standard Review Plan for the Review of Safety Analysis Reports for Nuclear Power Plants: LWR Edition," Revision 2, March 2007, ADAMS Accession No. ML070550060.

[79] U.S. Nuclear Regulatory Commission, “Fuel Safety Limit Calculation Inputs Were Inconsistent with NRC-Approved Correlation Limit Values,” Information Notice 2014-01, February 21, 2014, ADAMS Accession No. ML13325A966.

[80] ANSI/ASME NQA-1, “Quality Assurance Program Requirements for Nuclear Power Plants,” American National Standards Institute/American Society of Mechanical Engineers, New York, NY, 2015

[81] Dittus, F.W., and L.M.K. Boelter, “Heat Transfer in Automobile Radiators of the Tublar Type”, University of California Publications in Engineering, 2, pp 443-461, 1930.

[82] Box, G.E.P., W.G. Hunter, and J.S. Hunter, *Statistics for Experimenters*, John Wiley & Sons, Ltd., New York, NY, 1978.

[83] Wieckhorst, O., S. Opel, R. Harne, and F. Filhol, “Challenges in CHF Correlation Development,” LWR Fuel Performance Meeting (TopFuel 2013), September 15–19, 2013.

[84] Wieckhorst, O., H. Gabriel, R. Harne, M. Anghelescu, and O. Martinie, “ORFEO—A CHF Correlation for GAIA, AREVA’s Advanced PWR Fuel Assembly Design,” *16th International Meeting on Nuclear Reactor Thermal Hydraulics (NURETH-16)*, Chicago, IL, August 30–September 4, 2015.

[85] Anderson, M.G., and P.D. Bates (Eds.), *Model Validation: Perspectives in Hydrological Science*. John Wiley & Sons, Ltd., Chichester, United Kingdom, 2001.

[86] National Research Council, *Ground Water Models; Scientific and Regulatory Applications*, The National Academies Press, Washington, DC, p. 303, 1990.

[87] Owen, D.B., “Factors from One-Sided Tolerance Limits and the Variables Sampling Plans,” SCR-607, Sandia Corporation, Albuquerque, NM, 1963.

[88] Wilks, S.S., “Determination of Sample Sizes for Setting Tolerance Limits,” *The Annals of Mathematical Statistics*, 12(1), 91–96, 1941.

[89] Wald, A., “An Extension of Wilks’ Method for Setting Tolerance Limits,” *The Annals of Mathematical Statistics*, 14(1), 43–55, 1943.

APPENDIX A LISTING OF ALL GOALS

GOAL	The critical boiling transition model can be trusted.
G1	The experimental data supporting the critical boiling transition model are appropriate.
G1.1	The experimental data have been collected at a credible test facility.
G1.1.1	The test facility is well understood.
G1.1.2	The test facility has been verified by comparison to an outside source.
G1.2	The experimental data have been accurately measured.
G1.2.1	The test facility has an appropriate quality assurance program.
G1.2.2	The experiment has been appropriately statistically designed (i.e., the value of a system parameter from any test was completely independent from its value in the test before and after the test).
G1.2.3	The method used to obtain critical boiling transition data results in an accurate measurement.
G1.2.4	The instrumentation uncertainties have been demonstrated to have a minimal impact on the measured critical heat flux or critical power.
G1.2.5	The uncertainty in the critical heat flux or critical power is quantified through repeated tests at the same state points.
G1.2.6	The heat losses from the test section are quantified, appropriately low, and duly accounted for in the measured data.
G1.3	The test assembly reproduced the local conditions in the reactor fuel assembly.
G1.3.1	The test assembly used in the experiment should have geometric dimensions equivalent to those of the fuel assembly used in the reactor for all major components.
G1.3.2	The grid spacers used in the test assembly should be prototypical of the grid spacers used in the reactor assembly.
G1.3.3	The axial power shapes in the test assembly should reflect the expected or limiting axial power shapes in the reactor assembly.
G1.3.4	The radial power peaking in the test assembly should reflect the expected or limiting radial powers in the reactor assembly.
G1.3.5	Any differences between the test assembly and the reactor assembly should have a minimal impact on the flow field. This includes components that are not in the reactor assembly but that are needed for testing purposes.
G2	The model was generated in a logical fashion.

G2.1		The mathematical form of the model is appropriate.
	G2.1.1	The mathematical form of the model contains all the necessary parameters.
	G2.1.2	The reasoning for choosing the mathematical form of the model should be discussed and should be logical.
G2.2		The process for determining the model's coefficients was appropriate.
	G2.2.1	The training data (i.e., the data used to generate the coefficients of the model) should be identified.
	G2.2.2	The method for calculating the model's coefficients should be described.
	G2.2.3	The method for calculating the R- or K-factor and the additive constants (for both full-length and part-length rods) should be described. Further, a description of how such values are calculated if dryout is not measured on the rod under consideration should be provided (boiling-water reactors only).
G3		The model has sufficient validation as demonstrated through appropriate quantification of its error.
G3.1		The correct validation error has been calculated.
G3.2		The validation error is appropriately distributed throughout the application domain.
	G3.2.1	The validation data (i.e., the data used to quantify the model's error) should be identified.
	G3.2.2	The application domain of the model should be mathematically defined.
	G3.2.3	The expected domain of the model should be understood.
	G3.2.4	There should be adequate validation error data density throughout the expected and application domains.
	G3.2.5	Sparse regions (i.e., regions of low data density) in the expected and application domains should be identified and justified.
	G3.2.6	The model should be restricted to its application domain.
G3.3		Any inconsistencies in the validation error have been accounted for appropriately.
	G3.3.1	The validation error should be investigated to ensure that it does not contain any subgroups that are obviously not from the same population (i.e., non-poolable).
	G3.3.2	The expected domain should be investigated to determine if it contains any non-conservative subregions that would impact the predictive capability of the model.
	G3.3.3	The model's predictions trend as expected in each of the various model parameters.

	G3.4		The model's uncertainty has been appropriately calculated from the validation error.
		G3.4.1	The validation error statistics should be calculated from an appropriate database.
		G3.4.2	The validation error statistics should be calculated using an appropriate method.
		G3.4.3	The model's uncertainty should be appropriately biased.
	G3.5		The model has been correctly implemented.
		G3.5.1	The model has been implemented in the same computer code that was used to generate the validation error.
		G3.5.2	The model's prediction of the CBT is being applied using the same evaluation methodology used to predict the validation data set for determining the validation error.
		G3.5.3	The model results in an accurate or conservative prediction when it is used to predict transient behavior.

NRC FORM 335 (12-2010) NRCMD 3.7

U.S. NUCLEAR REGULATORY COMMISSION

BIBLIOGRAPHIC DATA SHEET

(See instructions on the reverse)

1. REPORT NUMBER (Assigned by NRC, Add Vol., Supp., Rev., and Addendum Numbers, if any.)

NUREG/KM-0013

2. TITLE AND SUBTITLE

Credibility Assessment Framework for Critical Boiling Transition Models
A Generic Safety Case to Determine the Credibility of Critical Heat Flux and Critical Power Models
Draft Report for Comment

3. DATE REPORT PUBLISHED

MONTH	YEAR
March	**2019**

4. FIN OR GRANT NUMBER

5. AUTHOR(S)

J.S. Kaizer, R. Anzalone, E. Brown, M. Panicker, S. Haider, J. Gilmer, T. Drzewiecki, and A. Attard

6. TYPE OF REPORT

Technical

7. PERIOD COVERED (Inclusive Dates)

8. PERFORMING ORGANIZATION - NAME AND ADDRESS (If NRC, provide Division, Office or Region, U. S. Nuclear Regulatory Commission, and mailing address; if contractor, provide name and mailing address.)

Division of Safety Systems
Office of Nuclear Reactor Regulation
U.S. Nuclear Regulatory Commission
Washington, DC 20555-0001

9. SPONSORING ORGANIZATION - NAME AND ADDRESS (If NRC, type "Same as above", if contractor, provide NRC Division, Office or Region, U. S. Nuclear Regulatory Commission, and mailing address.)

Same as above

10. SUPPLEMENTARY NOTES

J.S. Kaizer

11. ABSTRACT (200 words or less)

Critical boiling transition (CBT) occurs when a flow regime that has a higher heat transfer rate transitions to a flow regime that has a significantly lower heat transfer rate. Models that predict a CBT are a necessary part of reactor safety analysis because they are used to determine plant safety limits. Therefore, the review of CBT models has been a focus of the U.S. Nuclear Regulatory Commission (NRC) since its inception in 1975.

This work presents a generic safety case in the form of a credibility assessment framework that combines aspects of goal structuring notation and maturity assessment. This framework is focused on the credibility assessment of CBT models with specific application to reactor safety analysis. The NRC has performed many such assessments and has generated this framework based on the experience of current and former NRC staff, as well as previous staff reviews as summarized in staff evaluations. This document includes a survey of the important technical and regulatory literature; a detailed technical discussion of CBT models and their application; and a suggested framework for CBT models. This NUREG/KM summarizes the knowledge the NRC staff has developed over the course of 40 years of CBT model and analysis reviews.

12. KEY WORDS/DESCRIPTORS (List words or phrases that will assist researchers in locating the report.)

Critical heat flux, critical power, departure from nucleate boiling, critical quality, boiling crisis, burnout, dryout, critical boiling transition

13. AVAILABILITY STATEMENT

unlimited

14. SECURITY CLASSIFICATION

(This Page)

unclassified

(This Report)

unclassified

15. NUMBER OF PAGES

16. PRICE

United States
Nuclear Regulatory Commission

NUREG/KM-0013
Draft
March 2019

NUREG/KM-0013
Draft

CREDIBILITY ASSESSMENT FRAMEWORK FOR CRITICAL BOILING TRANSITION MODELS

A Generic Safety Case to Determine the Credibility of Critical Heat Flux and Critical Power Models

March 2019