DETERMINING M&S CREDIBILITY: What The Accreditor Needs To Know

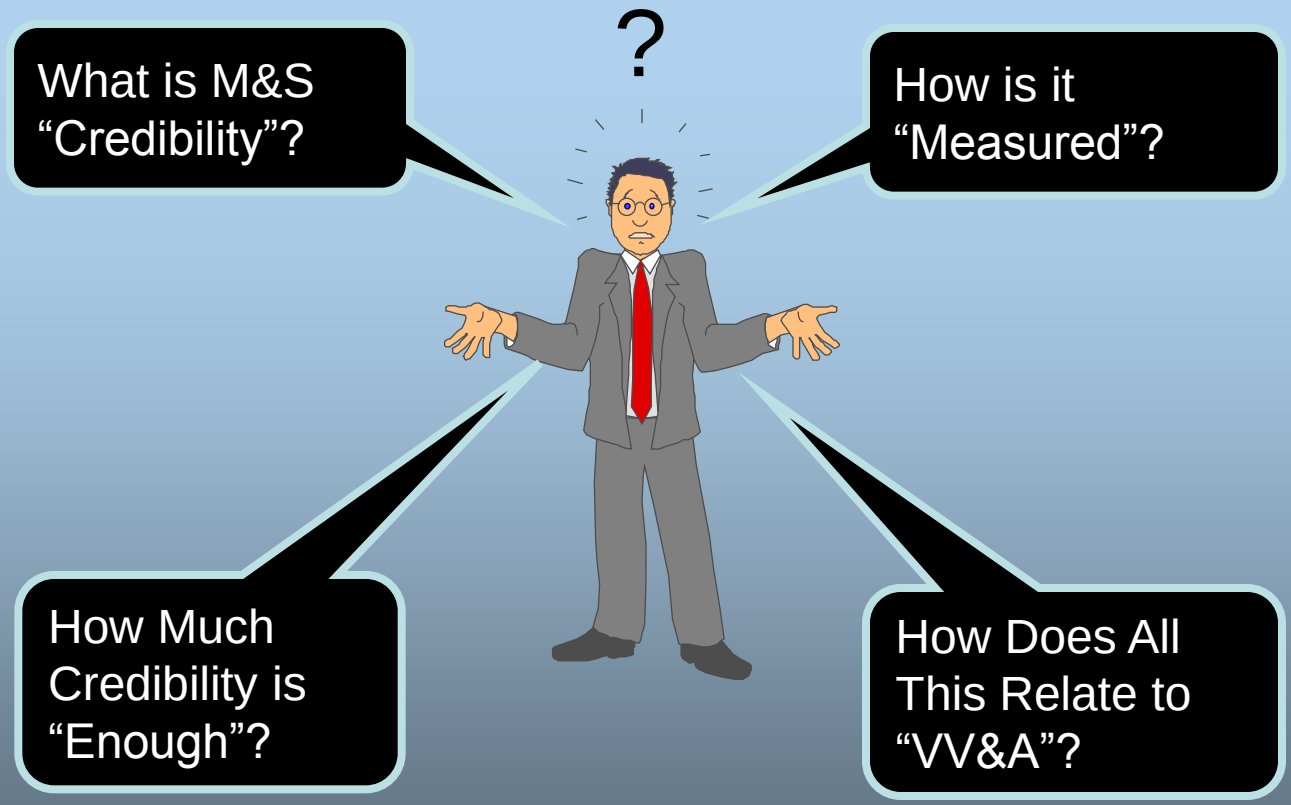
Presented To

The National Defense Industrial Association Modeling and Simulation Committee, 19 April 2016

James N. Elele, PhD
Department of The Navy M&S Verification Validation & Accreditation Lead Trusted Agent (VV&A) For Navy's Commander Operational Test Force
James.Elele@navy.mil
Ph: 301-342-4154

THE ISSUE:

M&S CREDIBILITY AND HOW TO SHOW IT



WHERE CAN I GET HELP?

FIRST THINGS FIRST

VV&A Definition (DoDI 5000.61)

- **VERIFICATION:** The process of determining that a model implementation and its associated data accurately represents the developer's conceptual description and specifications.
  - Does the model do what the originator intended?
  - Is it relatively error free?
  - Did we build it right?

- **VALIDATION:** The process of determining the degree to which a model and its associated data are an accurate representation of the real world from the perspective of the intended uses of the model.
  - Do model results match real world data well enough for your needs?
  - Note that M&S validation is not the same as software validation
  - Did we build the right thing?

- **ACCREDITATION:** The official certification [determination] that a model, simulation, or federation of models and simulations and its associated data are acceptable for use for a specific purpose.
  - Does the accreditation authority have adequate evidence to be confident that a model (and its input data) are credible and suitable for a particular use

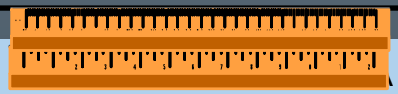
The goal of a VV&A program is to generate, gather, maintain, and apply M&S credibility artifacts to support the decision to use M&S.

KEY QUESTION:

What Makes A Simulation "Credible"?

- **Most people think "validation" is the hallmark of simulation credibility. But is it?**
- ***Validation has well-known limitations:***
  - **Limited scope of validation tests**
    - **You *cannot validate*** *over the whole domain* **of the M&S**
  - **Validation data are *difficult and costly* to obtain**
    - **Moreover "Real World" *always contains factors*** *not accounted for* **in the simulation.**
    - **Obtaining *sufficient validation test data* can be costly**
  - **Some simulations simply *cannot be validated* in the conventional sense of the word.**
    - **Wars are not experiments designed for data collection**
      - **e.g., how can we validate mission and campaign level warfare M&S?**

Fortunately, there are other measures of simulation credibility



CREDIBILITY

A Definition:

A model or simulation, its data, and its results have credibility if the decision-maker and other key project personnel accept them as "correct." (RPG)

Note:

- **A credible simulation is *not necessarily* valid, and vice versa.**
- **A model or simulation that is *both valid and credible* is more likely to be formally accredited for use in a particular application.**
- **The following factors help establish credibility for a model or simulation:**
  - **Decision-maker's understanding and agreement with the simulation's assumptions**
  - **Demonstration that the simulation has been *validated and verified***
  - **Decision-maker's ownership of and involvement with the project**
  - **Reputation of the simulation developers and analysts**
  - **History of previous use by other organizations or agencies.**

Important:

- **This Tutorial will discuss an approach for making a simulation valid and credible based on the M&S capability, accuracy and usability (to be defined)**

Why Is M&S Credibility So Important ?

Simulations Are Being Used To Support Very Many High Value Decisions:

- Warfare (Scenarios; Equipments' Performances & Effectiveness) Simulations
- Acquisition/Technology Development
- Risk Determination, Mitigation/Reduction
- Playing "What if" investigations for extremely hazardous conditions.
- Medical, Scientific, Training & Testing
- **HOW MUCH RISK CAN I ACCEPT IF M&S RESULTS TURN OUT TO BE WRONG?**

Because

POLICY MANDATES M&S USE: DODI 5000.02 SECTION 5.a(5) ; Jan 2015

- The Program Manager will integrate modeling and simulation activities into program planning and engineering efforts. These activities will support consistent analyses and decisions throughout the program's life cycle.
- "Ensure that all test infrastructure and/or tools (e.g., models, simulations, automated tools, synthetic environments) to support acquisition decisions ***will be verified, validated, and accredited*** (VV&A) by the intended user or appropriate agency"
- The PM *shall* use verified, validated, and accredited models and simulations, *and ensure credible applicability for each proposed use.* (DoD 5000.2-R; Section C2.6.7.2; 2002)

M&S Have Become Ubiquitous:

- Demand For M&S Applications Keeps Increasing (For Scientific Studies, For Acquisition, For Testing &Training, etc.)
- Budget: Resources Keep Shrinking
- M&S Is Considered More Economical Than Other Tools and Methods
- Requirements To Represent Very Complex Phenomena In Realistic Battlefield Environment Keep Increasing

VV&A AND M&S CREDIBILITY

WHAT ARE THE ISSUES?

M&S Credibility And The Evidence Needed To Show It

- **VV&A are used to establish the M&S Credibility through testing (verifying & validating), demonstrating and documenting the core M&S characteristics:**

THE THREE PILLARS OF M&S CREDIBILITY

  - **CAPABILITY**
    - **What the M&S can do**
  - **ACCURACY**
    - **How well it does it**
  - **USABILITY**
    - **How much facilities are available to ensure the M&S is used correctly**

AND DEMONSTRATE: FIT FOR PURPOSE

  - Assessing the Risk associated with using erroneous M&S results

- **VV&A Provides Confidence to decision makers that M&S is credible; and**
- **VV&A are Risk reduction processes**
- ***GOAL OF VV&A: To define, determine, generate and document information needed to assess the M&S credibility and justify using the M&S for the specified application***

How Is Credibility Established?

- M&S Credibility is established through testing, demonstrating and the documentation of:
  - **Capability:** What the M&S can do
    - Functions modeled & their levels of detail (Fidelity)
      - Assumptions & Limitations
  - **Accuracy:** How well it does it
    - *Software* (design & implementation verification, quality measures, code verification)
    - *Data* (pedigree, precision, integrity)
    - *Output* (validation, Configuration Management, M&S output comparison to "reality")
  - **Usability:** Is there enough information to allow error-free use?
    - Documentation, Technical Support

⇔ Did we meet the Requirements & Acceptability Criteria for these? How well?

- Credible simulations are less likely to provide incorrect results
  - They reduce the risk that an inappropriate or unsuitable simulation will be used to solve your problem
- Credible simulations provide confidence in the M&S output/results

Greater Credibility - Low Risk; Less Credibility - High Risk

CREDIBILITY AND M&S VV&A

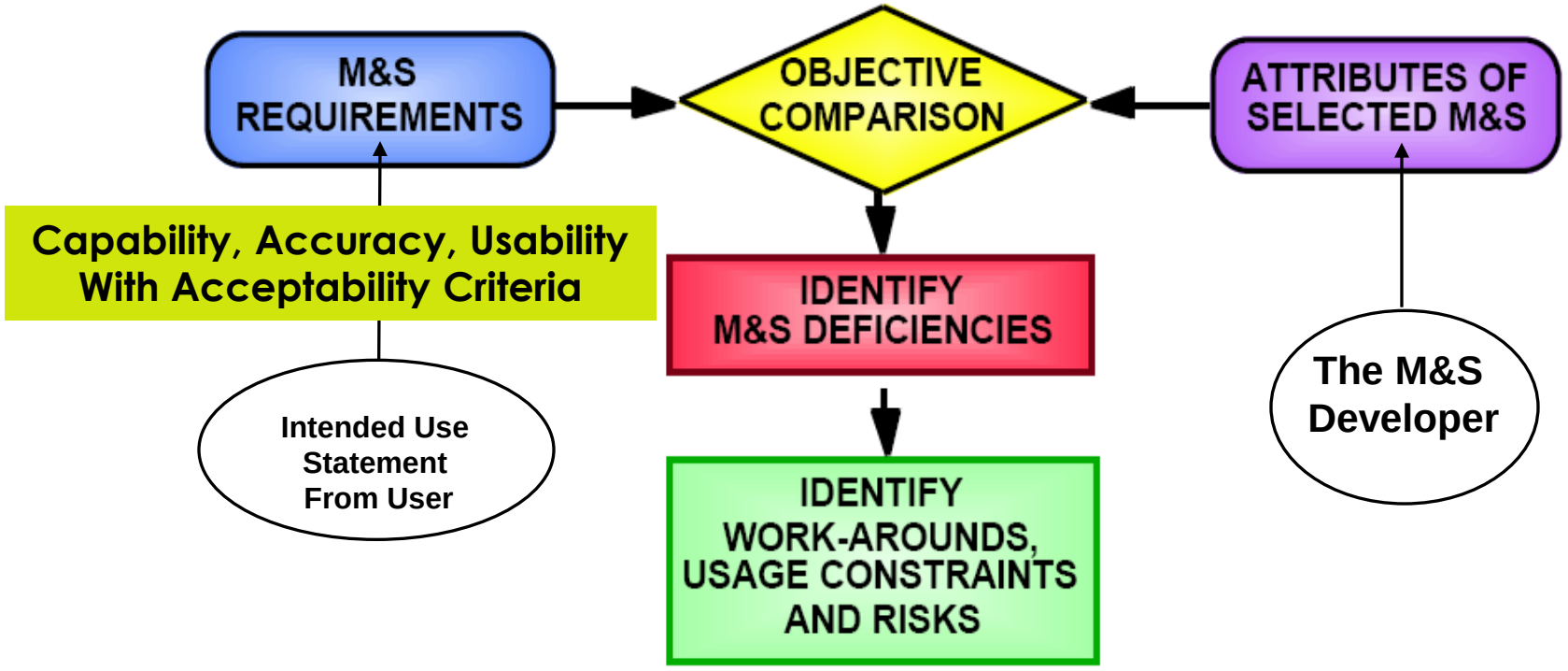
"The What And The Why"

- **ISSUE: The Risk Associated With Decisions Supported With Erroneous M&S Results.**
- VV&A is a Risk-reduction/mitigation process
- Sources Of Risk
  - M&S may solve the wrong problem/ Poor problem and Intended Use definitions.
  - Invalid/Not-credible M&S may be accepted.
  - Erroneous M&S output/results may be accepted and used.
- VV&A reduces/mitigates risks by ensuring that M&S:
  - address the correct problem.
  - produce, valid, accurate, consistent, and reproducible results.
  - provide confidence that decisions supported with M&S are credible and reliable
- VV&A provide documented evidence/proof for supporting acquisition decisions
- VV&A provide evidence needed to establish that M&S is credible!

THE PROBLEM: Knowing "How To Accredit"

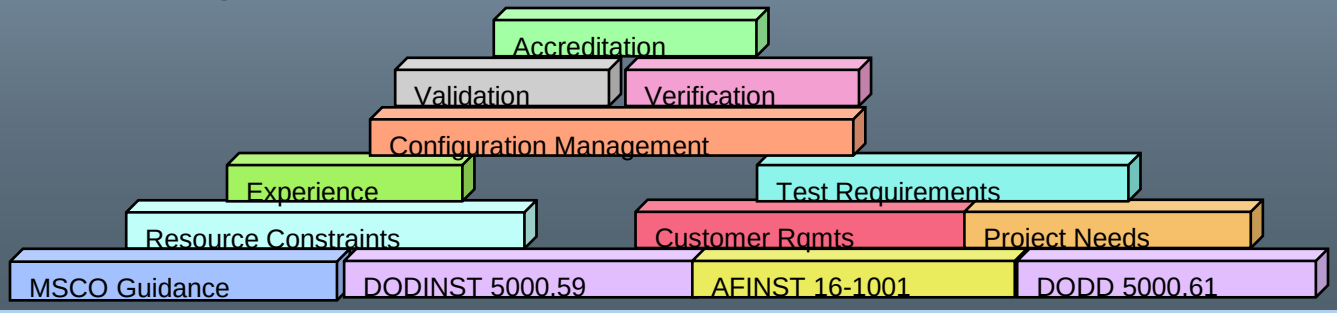
A PRACTICAL DEFINITION OF ACCREDITATION

AN OBJECTIVE COMPARISON OF SIMULATION CAPABILITIES WITH M&S REQUIREMENTS DERIVED FROM THE PROBLEM TO ESTABLISH CREDIBILITY OF M&S RESULTS



WHAT IS NEEDED

- A practical VV&A approach that is cost-effective for establishing M&S Credibility .
  - ***Is systematic and repeatable***
  - ***Is logical for demonstrating that M&S is credible***
  - ***Is not too complex to implement!***
- A method that can be used for fulfilling the M&S User's Intended application needs (and Service policies) while meeting the specific M&S credibility requirements for that Intended Use (IU):
  - Identifying M&S Credibility Requirements That Make Sense
  - Planning and Executing Cost-Effective VV&A Programs
  - Conducting Accreditation Assessments
  - Making VV&A Decisions Based On Risks Associated With M&S Use
    - Most efficient and effective implementation approach
  - Documenting Results of VV&A Activities AND Lessons Learned



SEEKING FOR ANSWERS

OUR AIM:
To provide the M&S Community & Industry with a COST-EFFECTIVE pathway to defining and meeting simulation credibility requirements.

HOW CAN I DETERMINE THE CREDIBILITY OF MY M&S AND THE ASSOCIATED DATA?

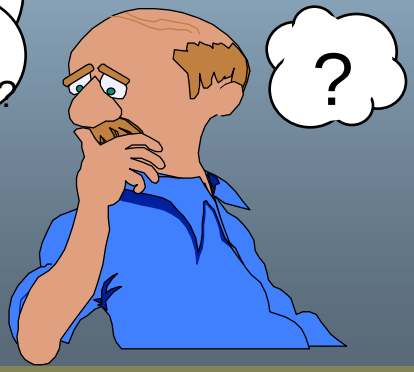
WHAT MAKES M&S CREDIBLE?

HOW MUCH VV&A IS ENOUGH?

HOW DOES VV&A FIT IN ALL OF THESE?

HOW DO I BALANCE COST AND CREDIBILITY?

ARE THERE SOME COST OPTIMUM APPROACHES?



PM & THE M&S USER

WHAT ARE MY RISKS WHEN M&S OUTPUTS ARE WRONG & USING THEM LEADS TO WRONG DECISION OUTCOMES?

- **Battlefield Operational Outcomes?**
- **Personnel Losses?**
- **Financial Loses?**
- **Other Resources (Schedule, Facilities etc.)?**

TOO MANY QUESTIONS! SOMEBODY DO SOMETHING!!

THE PROCESS

OUR APPROACH:

- **It all starts with a "*well-articulated* *Intended Use Statement*"**
  - *This is **absolutely critical**; for valid requirements definition*
  - *Without a clearly stated intended use, you **CANNOT** "prove" your M&S is credible and can meet the defined need*
  - *This Is What VV&A Needs To Prove (That This M&S Can Do)!*
- **What Makes *A Good* Intended Use Statement?**
- ***A Good Intended Use Statement (IUS) shall define:***
  - *What you are trying to model (General intended application description).*
  - *The Key Questions you are trying to find answers to (Critical Problem Statements/Descriptions).*
  - *How you are planning to use the results (e.g. T&E, Training, Usage Environs)*
  - *The Key Outputs (results) you are looking for (MOPs, MOEs, KPPs, etc.)*
- **NOTE: Why is this so important?**
  - ***This is what the M&S is going to be accredited for!***
  - ***The Intended Use is critical for defining the M&S (VV&A) Requirements!***

Example of A "Good" Intended Use Statement

The General Intended Use Statement For Traffic Flow Model (TFM)*

- The Traffic Flow Model (TFM) will simulate a standard 4-way intersection in order to provide analysis of traffic flow control and provide support for implementing improvements in congestion redirect.
- The TFM will be applied for the following uses:
- To estimate the performance level of current traffic control systems to efficiently dissipate high density traffic congestion.
- To simulate variations of traffic flow situations in a base environment.
- To determine the effect of altering traffic control systems.
- TFM will be further developed in the future with increased complexity and fidelity to represent the entire traffic system of PAX NAS.

*Created by NAVAIR 5.4 VV&A Team Employees and Interns as part of a study project on "How to Build a Credible Simulation"

EXAMPLE OF AN IUS FOR A TRAFFIC FLOW M&S

Intended Use for Traffic Flow Model (TFM)

The Traffic Flow Model (TFM) will simulate a standard 4-way intersection in order to provide analysis of traffic flow control and provide support for implementing improvements in congestion redirect.
The TFM will be applied for the following uses:

- To estimate the performance level of current traffic control systems to efficiently dissipate high density traffic congestion.
- To simulate variations of traffic flow situations in a base environment.
- To determine the effect of altering traffic control systems.

TFM will be further developed in the future with increased complexity and fidelity to represent the entire traffic system of PAX NAS.

Intended Use	Key Questions to Be Addressed by M&S	Application	Model Outputs/Data
The Model & Simulation (M&S) will be used to simulate traffic flow at a standard 4-way intersection on PAX NAS.	What is the efficiency of the current traffic control system at a single intersection of PAX NAS?	Determine the necessity of further management of high traffic density at this single intersection.	▪ Traffic queue lengths ▪ Queue wait times ▪ Traffic flow rates (Based on time of day and present traffic light settings)
	What is the current representative traffic density at the designated intersection?	Determine the baseline for traffic improvement.	▪ Traffic queue lengths ▪ Queue wait times ▪ Traffic flow rates (Based on time of day) and present traffic conditions)
	What is the effectiveness of the current traffic light cycle and timing at representative intersection?	Determine effectiveness of traffic light function when compared to present traffic congestion.	▪ Present traffic light timing, and cycle patterns
The M&S will be used to support improvement of traffic flow at single intersection of PAX NAS.	How can optimal traffic flow be maintained through control of traffic light functions at a given intersection?	Determine optimal light cycle timing to efficiently dissipate high density traffic.	▪ *Independent/ User Controlled Variable* o Adjustable light cycle time intervals ▪ *Dependent Variables* o Traffic queue lengths o Queue wait times o Traffic flow rates.
The M&S will be used to support future modeling and evaluation of the entirety of PAX NAS traffic congestion.	How can the modeling of a single intersection of PAX NAS be expanded to model all traffic lights on base?	Determine how traffic lights on base intersections can be managed to work both independently and as a system to optimize traffic flow.	▪ Multiple representative traffic intersections ▪ Traffic light cycle timing capable of acting as independent elements or as a part of a system.

*Created By NAVAIR 5.4.2.3 (BMVVB) Employees and Interns as part of a study project on "How to Build a Credible Simulation"

THE PROCESS: From The IUS Determine The VV&A Requirements

Requirements Are Determined/Categorized According To The Three Pillars Of M&S/Tool Credibility

Intended Use

What the M&S shall do…

- Requirement #1
- Requirement #2
- …
- …
- Requirement # (*n*- 1)
- Requirement #*n*

ACCEPTABILITY CRITERIA

CAPABILITY REQUIREMENTS

What can the M&S/Tool do (the functions) and to what level of detail (Fidelity Level)?*

Detailed Functional Decomposition

ACCURACY REQUIREMENTS

How well does the facility/M&S/Tool do what it is designed to do and is it relatively error-free?

- ***Software Accuracy***
- ***Data Accuracy***
- ***Output Accuracy***

ACCEPTABILITY CRITERIA

USABILITY REQUIREMENTS

What do users need to operate the Facility/M&S/Tool correctly and without introducing errors?

***RISK ASSESSMENT**: How fit is the Facility/M&S/Tool to satisfy the intended use with respect to the specified Capability and Accuracy?*

VVV&A Will Test, Demonstrate And Document How Well These Requirements, Their Acceptability Criteria And The Metrics/Measures Have Been Satisfied

THE PROCESS

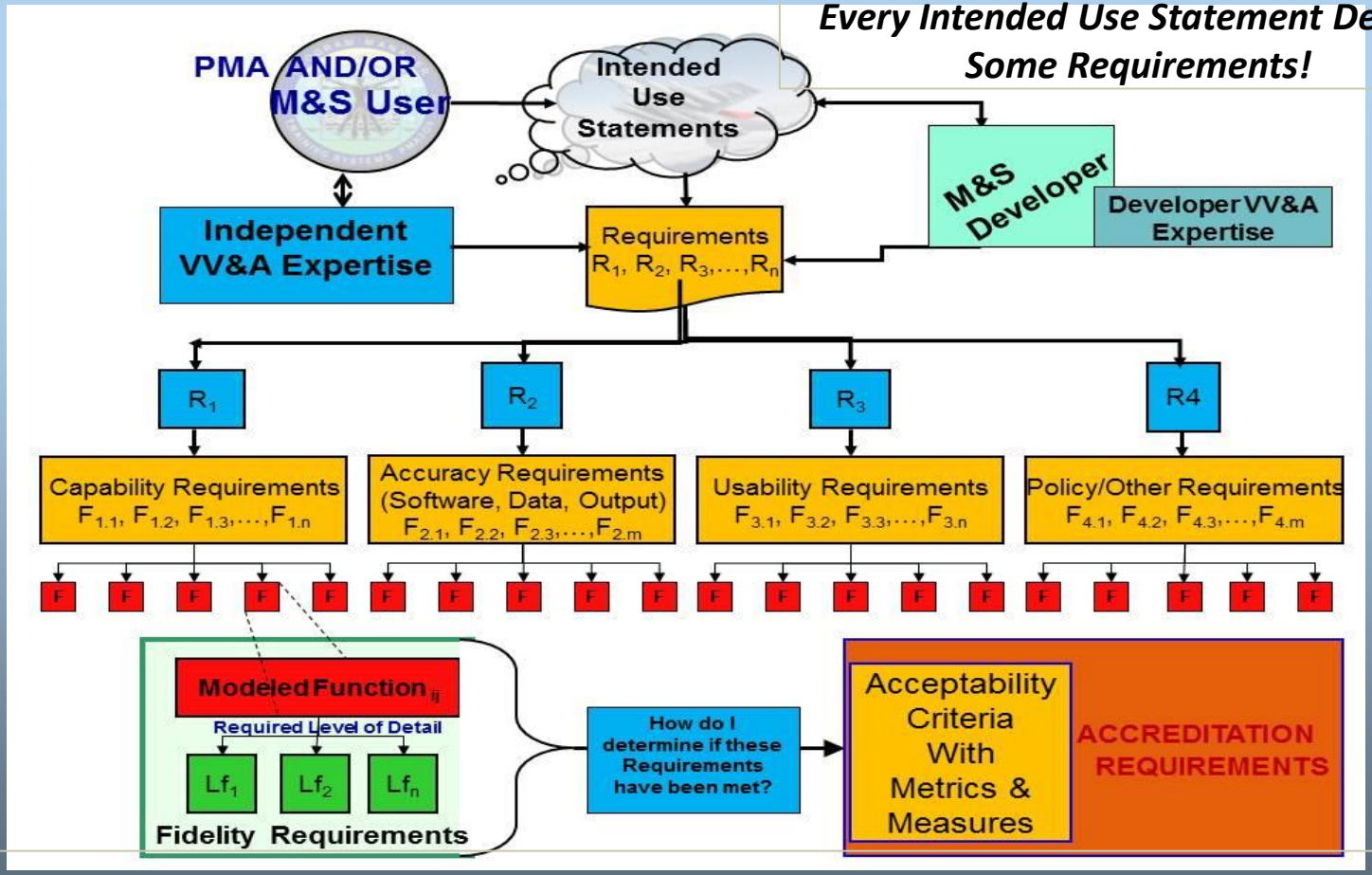
Using The Requirements Determine The Acceptability Criteria With Associated Metrics & Measures

Defining Requirements and Acceptability Criteria is as important as having an good intended use statement; otherwise, V&V cannot be accomplished.

- **Requirements:** Features and ***characteristics*** a model and/or simulation must have to be able ***to meet its intended use***
- **Acceptability Criteria:** Quantitative or qualitative properties a model or simulation must ***demonstrate to meet requirements*** for the intended use
- **Metrics/Measure:** ***How to determine*** whether or not the ***acceptability criteria has been met***

Determining Acceptability Criteria

Every Intended Use Statement Defines Some Requirements!



Note: The best way to obtain Acceptability Criteria is to decompose the Capability, Accuracy and Usability Requirements to their lowest level so their implementation and satisfiability can be easily verified. Tailoring Based On Level Of Risk and On Available Resources Is Usually Done Using Acceptability Criteria

MORE DETAILS ON

CAPABILITY
ACCURACY
USABILITY
M&S Use Risk Assessment

What Functions Are Modeled And To What Level Of Detail (Fidelity)?

The Conceptual Model Description

- Descriptions of simulation capability should include:
  - *Purpose:* Clear description of the specific purpose for which the simulation was developed and the phenomena being simulated
  - *Modeled elements:* Listing of the physical entities represented in the simulation, the **functions modeled** or **interactions** they have with other entities, and the level of detail (degree of fidelity) to which each entity is modeled. **Functional decomposition diagrams** that call out the level of details modeled (Fidelity of Representations)
  - *Environment:* Description of the physical environment in which the entities interact within the simulation ; data exchanged and their formats
  - *Relationships:* Description of I/O relationships between elements and the rules governing interactions among elements and the environment
  - *Assumptions and Limitations:* Description of any aspect of the design or implementation that limits the scope of potential uses
    - Summary of Assumptions, Limitations, and Errors (SALE) allows a model User to quickly identify all known model characteristics that might limit its applicability to a particular problem.

Sample High Level Requirement Table for Capability

The Comprehensive Air Performance Toolset (CAPT) is a simulation designed to support a generalized Unmanned flight system-under-test (SUT).

CAPT Requirement	Acceptability Criteria	Metrics/Measures
CAPABILITY: CAPT shall simulate turbulent environmental condition using a physics-based mathematical formula	The M&S shall incorporate either von Karman or Dryden form of turbulence model (modeled to specified parametric level of detail)	SME review and compare between the output from the selected turbulence model and the expected results (obtained from field test, lab measurements, etc.)
"What (function) CAPT shall have to meet the Intended Use…"	***"What CAPT shall demonstrate to meet the Requirement…"***	***"How to determine whether or not CAPT has met the Acceptability Criteria…"***

Sample Requirement Table for Capability (Cont.)

The Comprehensive Air Performance Tracking (CAPT) Simulation is a tool that supports a flight system-under-test.

CAPT Requirement	Acceptability Criteria	Metrics/Measures
CAPT shall simulate structural bending modes	XYZ values for structural bending mode from CAPT agree with flight test data values with no more than 0.1 difference	SME Review and comparison of output values and flight test data values
"What (function) CAPT shall do to meet the Intended Use…"	***"What CAPT shall demonstrate to meet the Requirement…"***	***"How to determine whether or not CAPT meets the Acceptability Criteria…"***

SIMULATION ACCURACY

M&S ACCURACY*: The Error-freeness of the simulation*

- **Most People think of Simulation Accuracy in terms of Output Accuracy. *But simulations are very complex systems!***
- **Simulation accuracy is *influenced by many factors*.**
- **Assuming "correct" conceptual definitions of functional Capability (both scientific and heuristic):**
  - **The most important factors influencing accuracy are:**
    - **Software Accuracy**
    - **Data Accuracy**
    - **Output Accuracy.**

The Quality and "Error-freeness" of the software

- **"Good" SW Accuracy Verification Is Characterized by:**
  - Results of both logical and code verification
  - Description of techniques, tools, and test conditions used
    - The variety of possible techniques precludes a detailed listing here
  - Documented acceptance criteria for software tests
    - For subjective criteria, qualifications of evaluators should be documented
  - Documented software test results, including:
    - Technical (dynamic and static test results), requirement coverage, etc
    - Any limitations or errors identified through the verification efforts
    - Implications for simulation use
  - Confidence in software accuracy is a function of:
    - Development environment: Verification evidence (activities and results), & Any S/W quality assessments done on the software
    - User must be able to answer the question: "How much confidence do I have that the simulation is well-constructed and good enough for my use?"

ACCURACY: SOFTWARE QUALITY ASSESSMENT

Application Architecture Standards
- Multilayer design compliance (UI vs App Domain vs Infrastructure/Data)
- Data access performance
- Coupling Ratios
- Component (or pattern) reuse ratios

Coding Practices
- Error/exception handling (all layers UI/Logic/data)
- If applicable - compliance with OO and structured programming practices
- Secure controls (access to system functions, access controls to programs)

Complexity
- Transaction
- Algorithms
- Programming practices (eg use of polymorphism, dynamic instantation)
- Dirty programming (dead code, empty code…)

Documentation
- Code readability and structuredness
- Architecture -, program, - and code-level documentation ratios
- Source code file organization

Portability: Hardware, OS and Software component and DB dependency levels

Technical and Functional Volumes
- # LOC per technology, # of artifacts, files
- Function points - Adherence to specifications (IFPUG, Cosmic references..)

SOFTWARE ACCURACY VERIFICATION

- **Good Software Verification Evidence Is Characterized by:**
  - Logical and code verification, regression testing, integration testing
  - Description of techniques, tools, and test conditions used are too many to list here:
    - MSCO VV&A Recommended Practices Guide is a good source document (Reference Documents - V&V Techniques)**
  - Documented acceptance tests and test criteria for software (must be articulated in V&V documents)
    - Where criteria are subjective, qualifications of evaluators should be documented
  - Documented acceptance tests results, including
    - Technical results
    - Any limitations or errors identified through verification
    - Implications of limitations for M&S intended Use
  - M&S development and implementation contracts must specify what is needed for "Good software V&V Evidence"
  - Evidence of well-implemented Configuration Management is being followed in practice

** http://msco.mil/VVA_RPG.html

SOFTWARE ACCURACY REQUIREMENTS

Sample High Level Requirement for Software Accuracy

The Comprehensive Air Performance Toolset (CAPT) is a simulation designed for supporting an unmanned flight SUT. All simulation software used to support CAPT must satisfy the following: requirements:

CAPT Requirement	Acceptability Criteria	Metrics/Measures
CAPT software shall be tested adequately to demonstrate its proper operation against the requirements identified	Software test results and verification activities have been conducted	Software test and verification results are available and documented according to a user-specified format
	The software development environment has been well-structured and documented (CMMI Level III Certification or equivalent compliance is required)	Artifacts of the software development process are available and follow user-specified format
"What (function) CAPT shall do to meet the Intended Use…"	***"What CAPT shall demonstrate to meet the Requirement…"***	***"How to determine whether or not CAPT meets the Acceptability Criteria…"***

DATA ACCURACY

Conducting Data Quality Assurance

Everyone Says: Garbage In Garbage Out, but…

- **Here Is What To Do:**

➢ **Data Verification:**

- **Data Pedigree:** Ensure that input data sources and appropriate classifications are identified, documented and maintained
- **Data Collection:** Ensure that data collection conditions and their limitations are identified and documented
- **Embedded Data:** Verify internal embedded data and data transformations via computations are consistent and correct (*desk audits, simplified table top computations/simulation etc*.)
- Verify that input /output data handling and usage in the model are defined

➢ **Data Validation:** Ensure that input and embedded data/constants are consistent with the best or accepted values/estimates by *comparing and confirming* with known values

➢ **Data Certification:** Formal approval of the validity and pedigree of a data set for use for a specific purpose including appropriateness of classification by appropriate authority

- **Why Do It?**
  - Builds confidence that input data sets are acceptable for use
  - Gives the user confidence that key input parameters used by the model are as accurate as best estimates permit
  - Certifies that input data are acceptable for use in the given application

DATA ACCURACY (Cont.)

Conducting Data Quality Assurance And Documenting Data Precision Requirements

- **Various Techniques Used:**
  - Data Analysis, semantic analysis, data interface testing, and data flow testing
  - Check and verify input data values precision and bounds
- **Example: The value of the data T *shall* satisfy: $T \neq t$; $T \leq t$; $T = \bar{T} \pm t$; $T \geq t$**

- **Data identification and selection processes should be initiated as early in the process as possible (e.g., during conceptual model development)**
  - **They can continue throughout this phase.**
- **Focus on how to prepare the selected data for use.**
  - **For hard-wired/embedded data, check for appropriate formats (e.g., floating or integer) and degree of accuracy (e.g., should (pi) be represented by 3.14; 3.1416 or 3.141592?) to meet the needs of the simulation.**
- **Always document the rationale for and data modifications.**
- **Input data determining what form the data are available in, what form the model needs the data to be in, and the best way to make the transformations.**
- **In most situations, instance data will have to be transformed from their original state.**
  - **Determine and select appropriate transformation techniques (may have to develop and validate the chosen transformation).**
- **Data Transformation Examples:**
  - **Converting all rates of speed to kilometers per hour, all ranges to kilometers**
  - **Converting the probabilities of acquisition, shot, hit, and kill into a single-shot kill probability**
  - **Aggregating kills of individual classes of targets by individual classes of weapon systems over time into a overall kill probability**
- **For output instance data, a major concern is ensuring the design (i.e., algorithms, code, interfaces) can produce, collect, and store the desired output data. In many situations, output data have to be transformed (e.g., aggregated, combined) to produce usable results. Whether data collection occurs external to the model or not, appropriate transformation techniques should be selected (or developed) and validated during design to make sure that proper collection is possible.**

OUTPUT ACCURACY

THIS IS WHAT IS COMMONLY KNOWN AS "VALIDATION"

RECALL DEFINITION OF VALIDATION: The Degree to which Simulation Outputs Match the "*Real World*".

- **Validation always comes down to a *comparison* between simulation predictions and some representation of the *"Real World"***
- **For M&S, there are *three* ways to define the "*Real World*" :**
  1. **Another "Valid" simulation (Benchmarking; some call it Registration)**
  2. **Subject Matter Expert (SME) Review (Face Validation)**
  3. **Test Data (Range/Lab/Field measured data)**
     - Sensitivity Analysis (i.e. output stability analysis) supports all three.

❖ **PROBLEM: *Heavy reliance on M&S "Validation/Output Accuracy" at the expense of Verification has become a very common practitioner issue!***

❑ **NOTE: *Validation Is very Important, But…..***

– ***Comparing M&S output to "Real/Measured Data," no matter how well matched, cannot compensate for wrong conceptual representation; algorithm, logic and coding errors which can only be detected through careful verification.***

– ***A good programmer can always match the output of erroneous simulation to measured data to "prove" the simulation validation requirements. But this cannot make the simulation valid!***

– ***For Critical Simulations, more resources should be spent on ensuring the correctness of conceptual representations, algorithm, logic and coding correctness with validation serving as the confirmation.***

– ***For Validation, "good enough" is determined by how the simulation will be used and the available resources. Subject Matter Expert reviews and confirmation of the correctness of conceptual representation, algorithm, logic and coding correctness provide extra needed confidence.***

– ***With good enough verification*, validation becomes the logical way to guarantee when to Quit "tweaking" the simulation and accept the simulation for its Intended Application.**

OUTPUT ACCURACY ISSUES

THE NEED FOR

"FIELD MEASURED M&S VALIDATION DATA"

Simulation Credibility

Output Accuracy Issues: Is There An Overemphasis On Validation?

- **Validation Short of Verification is a serious problem because:**
  - A good programmer can tweak a simulation to match any given "Real Measured Validation Data"
  - Natural variability and susceptibility to experimental errors in data
  - It is probably as important (if not more) to verify concepts, logic and error-free software implementation
- **Passive acceptance of benchmark simulation suitability can become an issue**
  - Wide use or Service "blessings" ***are indicators, not proofs,*** of simulation suitability
- **Face validation (SME Review) is a subjective evaluation, BUT there are objective criteria for a "good" face validation:**
  - The right people (SMEs, with)
  - The right experience (looking at)
  - The right data (in)
  - The right context (is necessary for good Face validation)

IMPORTANT: Simulation Management Boards DO NOT necessarily serve well as Face Validation SMEs

- **The main value of "Real Measured Validation Data"** is to **provide confidence** about when to stop "**Tweaking**" the Simulation. This provides confidence that the simulation agrees **with "This One Real Measured Instance Of Data!"** The number of **Instances Of agreements may or may not** be critical :Actual system operations may never encounter these scenarios.

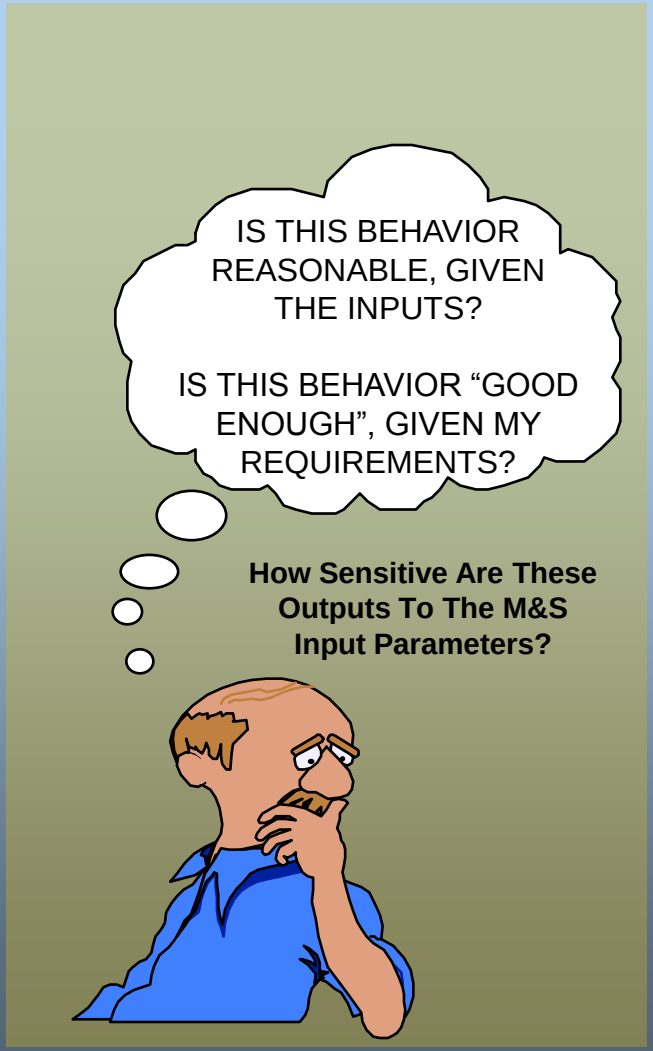
SOMETHING TO CONSIDER:

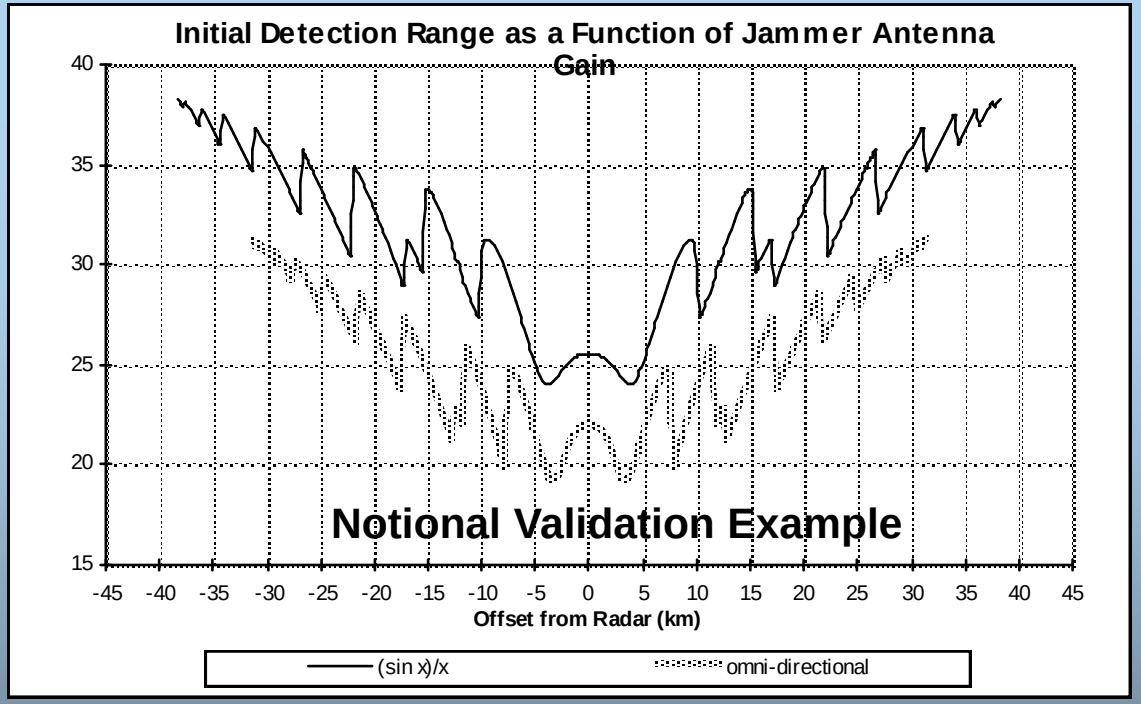
How much measured data does NASA require to validate simulations used for the design of satellites that land on the surface of an orbiting COMET in space?

OUTPUT ACCURACY ISSUES

M&S Credibility And Output Data Analysis

What are available techniques and when are they applicable (Statistical Data Comparison Tests & Interval Estimations)?
Sensitivity Analysis And Design of Experiments (DOE): Very valuable but must be done right!





Antenna Pattern	Mean (m)	s (m)	Normalized Mean Difference	% Change
Omni-directional (baseline)	25.26	3.38	-	-
(sin X)/X	30.67	4.00	0.05	9.67

SIMULATION USABILITY: Often Neglected

USABILITY: Is defined in terms of reduced likelihood of "error in operation" and "misuse of the M&S" in application

- Usability is not ease of simulation use, though this is a very desirable characteristic.
- Usability refers to availability of simulation user support features and guidance that:
  - Facilitate credible use of the simulation
  - Use compatible hardware and software
  - Reduces the probability that it will be employed inappropriately
- Simulations are credible
  - Only within a well-defined usage context
  - And only when properly used within that context
- Credible and valid simulations will produce wrong answers if not operated correctly

Any simulation attribute that reduces the probability of simulation likelihood of "error in operation" and "misuse" enhances its credibility

USABILITY REQUIREMENTS

Documentation Requirements for Usability

- **Usability Documentation shall be in terms of reduced likelihood of error in operation and M&S misuse**
- **VV&A Documentation shall include:**
  - ✓ User support documents (such as, but not limited to user manuals, web access, helpdesk, phone, etc.)
  - ✓ Usage history and demonstration of user community acceptability
  - ✓ Configuration management and control with documented demonstration of actual implementation and use in operation
  - ✓ Any other available support resources
- **Document shall articulate Implications of any Limitations in Usability Requirements for the Model's Intended Application**

REQUIREMENTS FOR USABILITY

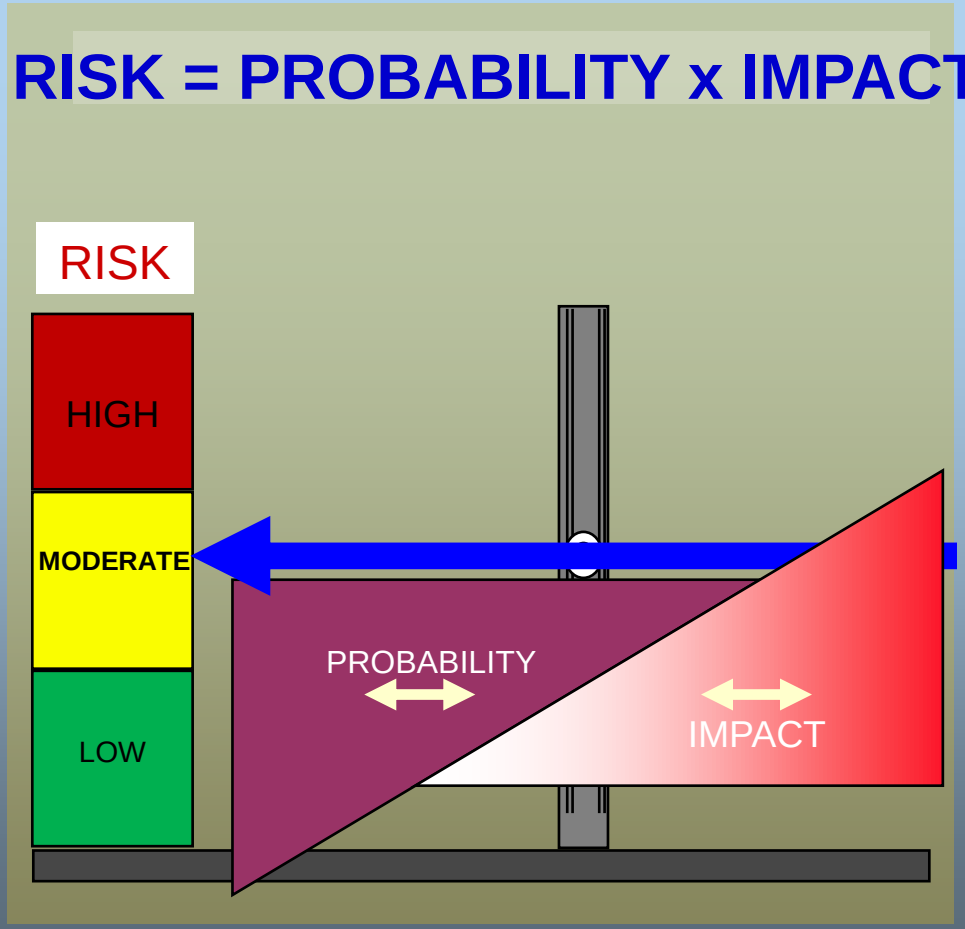
Sample High Level Requirement Table for Simulation Usability

CAPT Requirement	Acceptability Criteria	Metrics/Measures
Processes and documentation shall be in place to ensure proper operation and appropriate interpretation and use of outputs.	CM processes are sufficient and adequately documented and being followed.	CM plans and artifacts are available; review by potential users and other SMEs.
	Users are appropriately skilled and have the necessary training.	Documentation of user training and experience, and any credentials are available for SME review.
	User manuals and training are adequate to enable the user to properly execute the simulation.	SME review of user manuals and training materials.
"What (function) CAPT shall do to meet the Intended Use…"	***"What CAPT shall demonstrate to meet the Requirement…"***	***"How to determine whether or not CAPT meets the Acceptability Criteria…"***

Simulation Output Use Risk Assessment

How Much VV&A Is Enough?

- ***It Depends On Risk***
  - Risk means something bad happens because you believed an incorrect simulation result
    - **Decisions based on M&S results can incur risks if M&S results are in error**
  - The NAVAIR Process is a way to plan and implement VV&A activities based on risk
    - It has been used to support many ACAT I programs for the Navy; the Air Force (High Energy Laser) and the Army (Underbody Blast M&S)



VV&A: M&S OUTPUT USE RISK

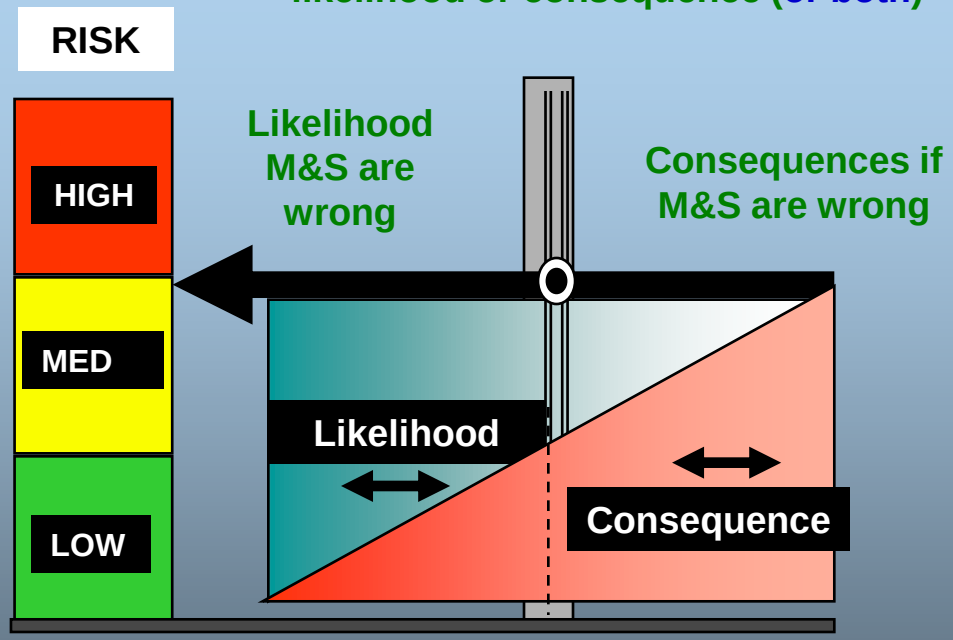
DOD DEFINITION: RISK = LIKELIHOOD X CONSEQUENCE

Risk Of Concern

- ***Here, the risk of interest is the risk associated with using M&S results***
  - M&S includes the models and simulations as well as the necessary input data
- ***Likelihood is the odds that the M&S and/or their input data are incorrect or inappropriate for the intended use***
- ***Consequence is the impact if the M&S output is wrong but you believe it and act on it***

Note: The risk associated with model development – will it be done on time and within budget—is an important but separate issue. Here we focus on operational risk.

To reduce risk, reduce either likelihood or consequence (or both)

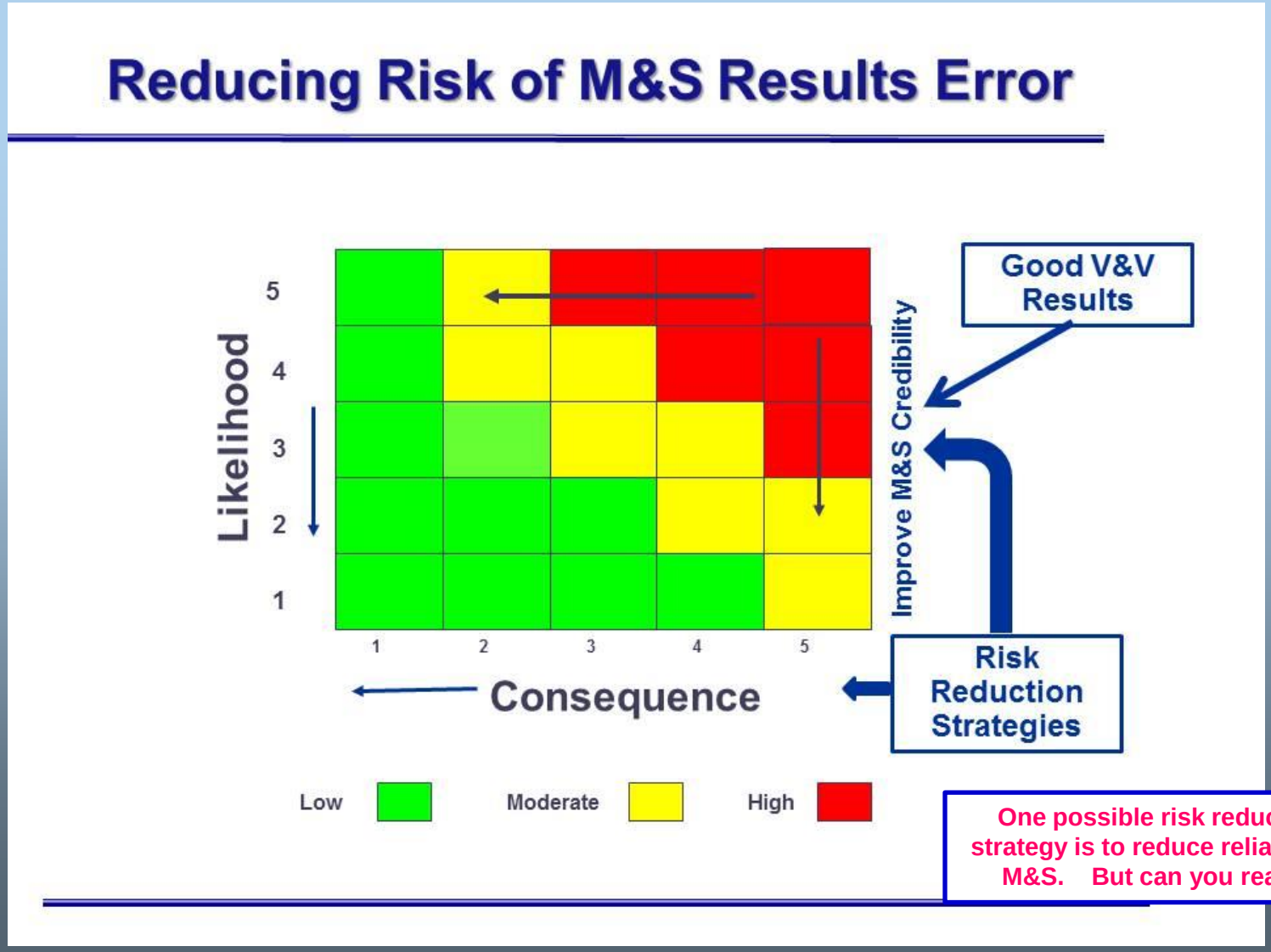


***PROBLEM:** How do you multiply two things (Likelihood and Consequences) you may not be able to define quantitatively?*
We Start by defining standardized scales for each element.

- NAVAIR has developed standardized scales for these elements with rules for associating them. These scales and association rules have been used successfully in many programs.

Risk-based VV&A

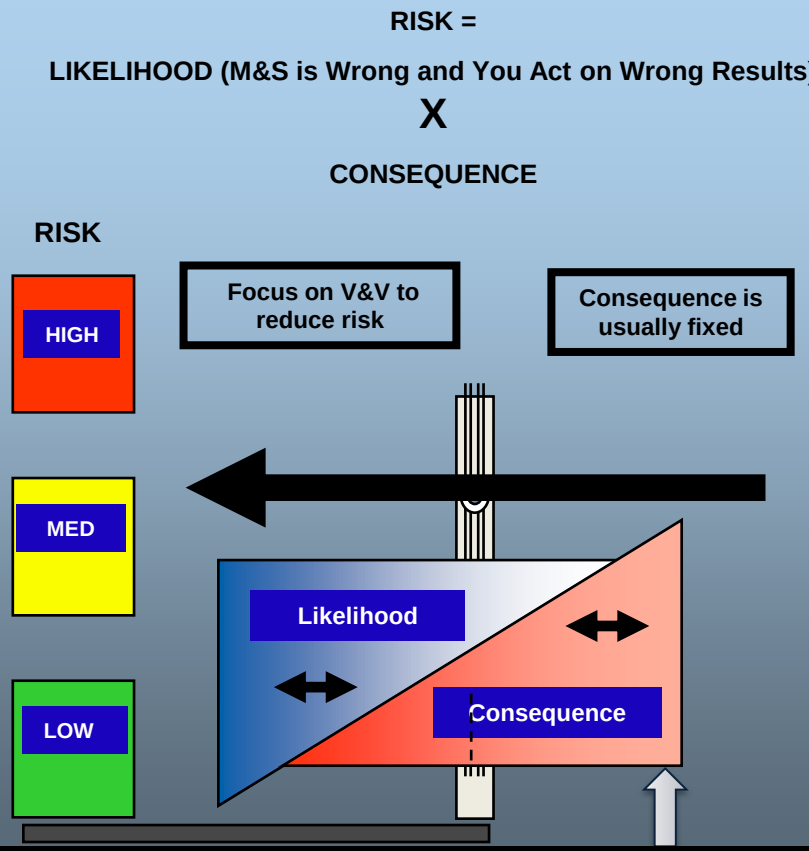
The Higher The Risk The More Rigorous The VV&A Required



VV&A AS RISK A REDUCTION PROCESS

Reduce Likelihood of Error, Reduce Risk

- *Verification*
  - Reduces the likelihood of undetected errors that are fatal to your intended use
- *Validation*
  - Reduces the likelihood that simulation outputs wont match the “real world” well enough for your problem
- *Accreditation*
  - Reduces the likelihood that an inappropriate or unsuitable simulation is selected for use in solving your problem



The goal of a VV&A program is to generate, gather, maintain, and apply M&S credibility artifacts to support the decision to use M&S.

VV&A: M&S OUTPUT USE RISK

DETERMINING M&S RISK

- ***Establishing M&S Credibility for acceptance by the Accreditation Authority are generally based on assessment of risk***
- **The Nature and extent of information required to support accreditation decision is at the discretion of the accreditation authority.**
- **Risk associated with using M&S is determined by the following:**
  - The **Role of M&S** results in decision making process
  - The **Importance** of decision that M&S is supporting
  - The Severity of the **Consequences** of making incorrect decisions
  - The Probability that the M&S results upon which the decisions are based will be incorrect (**Likelihood of Error)**
- **By evaluating how well the Capability, Accuracy and Usability Requirements (and their associated Acceptability Criteria) are able to satisfy the M&S Intended Use, the likelihood of the M&S output being in error can be determined**

VV&A: M&S OUTPUT USE RISK

DETERMINING
Role/Reliance on M&S for Decision Making

- **Here's an example scheme**

Role Level	Definition
4	**M&S will be the only method employed to make a decision**
3	**M&S will be the primary method, employed with other non-M&S methods**
2	**M&S will be a secondary method, employed with other non-M&S methods, and will provide significant data unavailable through other means**
1	**M&S will be a supplemental method, employed with other non-M&S methods, and will provide supplemental data already available through other means**

VV&A: M&S OUTPUT USE RISK

DETERMINING
Levels Of Importance of Decisions

Importance Level	Description
4	Intended use addresses multiple areas of significant program risk, key program reviews and test events, key system performance analysis, primary test objectives and test article design, system requirements definition, and/or high software criticality, used to make a technical or managerial decision.
3	Intended use addresses an area of significant program risk
2	Intended use addresses medium or low program risk, other program reviews and test events, secondary test objectives and test article design, other system requirements and system performance analysis, and medium or low S/W criticality used to make technical or managerial decisions.
1	Intended use addresses program objectives or analysis that is not a significant factor in the technical or managerial decision making process.

DETERMINING LIKELIHOOD OF ERROR

Risk Assessment For Specific M&S Characteristics

Characteristic	Criterion	Rating
Capability		
Intended Use and Acceptability Criteria	The General and Specific Intended Use(s) of the M&S is/are clearly stated; the Acceptability Criteria and their Metrics are clearly articulated	**RED**
Conceptual Model Validation	The conceptual model (framework, algorithms, data sources, and assumptions) is documented and correctly and adequately describes the needs and requirements of the intended use.	**YELLOW**
Model Fidelity (Function and Entity Level Decompositions)	The model's Functions, Entities and Data (framework, algorithms, data values, and assumptions) representation levels are documented and appropriate for the intended use.	**GREEN**
Accuracy		
Design Validation	The algorithms and/or mathematical formulations are correct and valid. The premises for the application of the algorithms and/or mathematical formulations are correct with no assumptions violated.	**YELLOW**
Input and Embedded Data	The simulation input and embedded data are credible, and subject to review and revision.	**GREEN**
System Verification	The M&S has been formally tested or reviewed and has been demonstrated to accurately represent the specific intended use(s) and requirements.	**RED**
Output Validation	The M&S responses have been compared with known or expected behavior from the subject it represents and has been demonstrated to be sufficiently accurate for the specific intended use(s).	**YELLOW**
Configuration Management	The M&S and its components are supported by a sound written Configuration Management (CM) Plan.	**GREEN**
Usability		
Documentation	The M&S is well documented as to capabilities, limitations and assumptions; documentation is readily available, up-to-date, and complete.	**RED**
User Community	The M&S is designed and developed for the level of competency of the intended users. The users have access to documents such as user's manual, training manuals, and/or reference guides. User support is available from the M&S developer or proponent.	**YELLOW**

VV&A: M&S OUTPUT USE RISK

Steps For M&S Risk Determination Process

Determine the "Importance of Technical Decision Supported by the Model/Simulation". Enter Importance rating.

Determine the "Level of Reliance on the Model/Simulation". Enter Reliance rating.

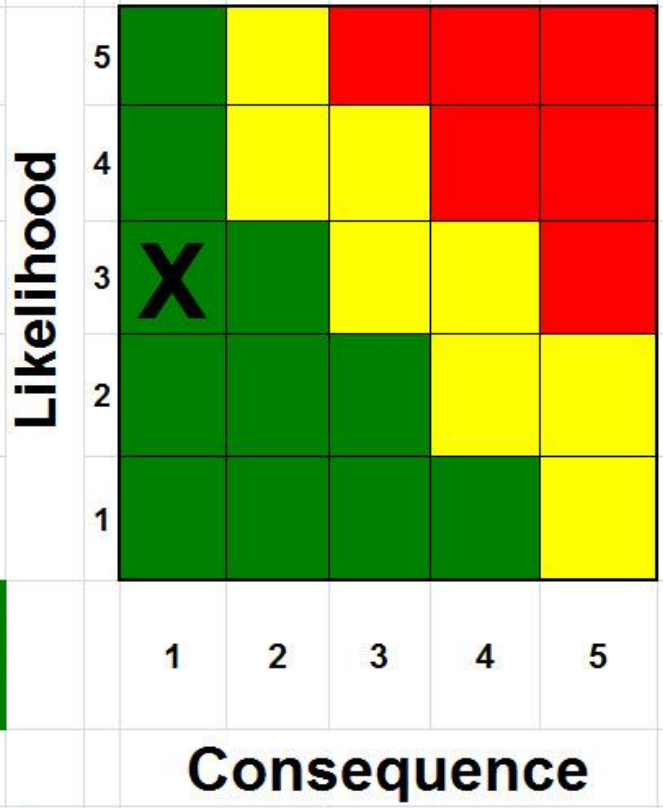
Consequence rating is automatically calculated from the Importance and Reliance ratings using the Combining Matrix.

Combining Matrix is shown in Table 2 for reference.

Risk is automatically calculated from the Likelihood and Consequence ratings.

RATINGS:

Likelihood (1 to 5)	3
Importance (1 to 4)	1
Reliance (1 to 4)	1
Consequence (1 to 5)	1
Overall Risk Level	Low



- NAVAIR Has Developed Standard Tables For Assessing Likelihood, Importance, Reliance, And Consequence

M&S CREDIBILITY:Process Summary

- **Start With "Well-defined" Intended Use Statement(s)**
- **Define Simulation Credibility (Requirements) :**
  - **Capability:** (Functions modeled and their level of detail (Fidelity))
    - Detailed function decomposition required
  - **Accuracy:** Software, Data, and Outputs (Validation)
    - Well defined/testable and documented requirements
  - **Usability:** User Facilities that facilitates operating the Simulation Correctly (User manuals, Help Desk, User websites, documentation)
    - Implemented Configuration management process that is followed
- **Define Acceptability Criteria And Their Measures**
  - Carefully determined processes with simulation outcome parameters
- **Conduct VV&A To Establish Simulation Credibility**
  - Verify, Validate And Accredit (determine, test, and document results of) Capability, Accuracy and Usability, then assess to complete the simulation credibility picture
- **Conduct Risk Assessment**
  - Determine the extent of V&V (and other information) needed to establish Simulation Credibility for the defined Intended Use
- **Summarize Assumptions, Limitations and Known Errors (SALE)**

SOMETHING TO REMEMBER

THE M&S CREDIBILITY EQUATION

START WITH A Well-articulated Intended Use Statement

CREDIBILITY = f(Capability, Accuracy, Usability)

VERIFY:

Well-defined Requirements,* Acceptability Criteria *with Metrics & Measures

CAPABILITY: What Can This Simulation Do?

Functions, Fidelity, Assumptions, Limitations, and Operational Characteristics needed to meet the IUS

ACCURACY: How well does it do it?

Software, Data, (Tested, documented and results under configuration management)

VALIDATE: **Simulation Outputs** (Compare: With test data; benchmark, and Face-validate via SME review)

USABILITY: Do I have what I need to use/operate it correctly?

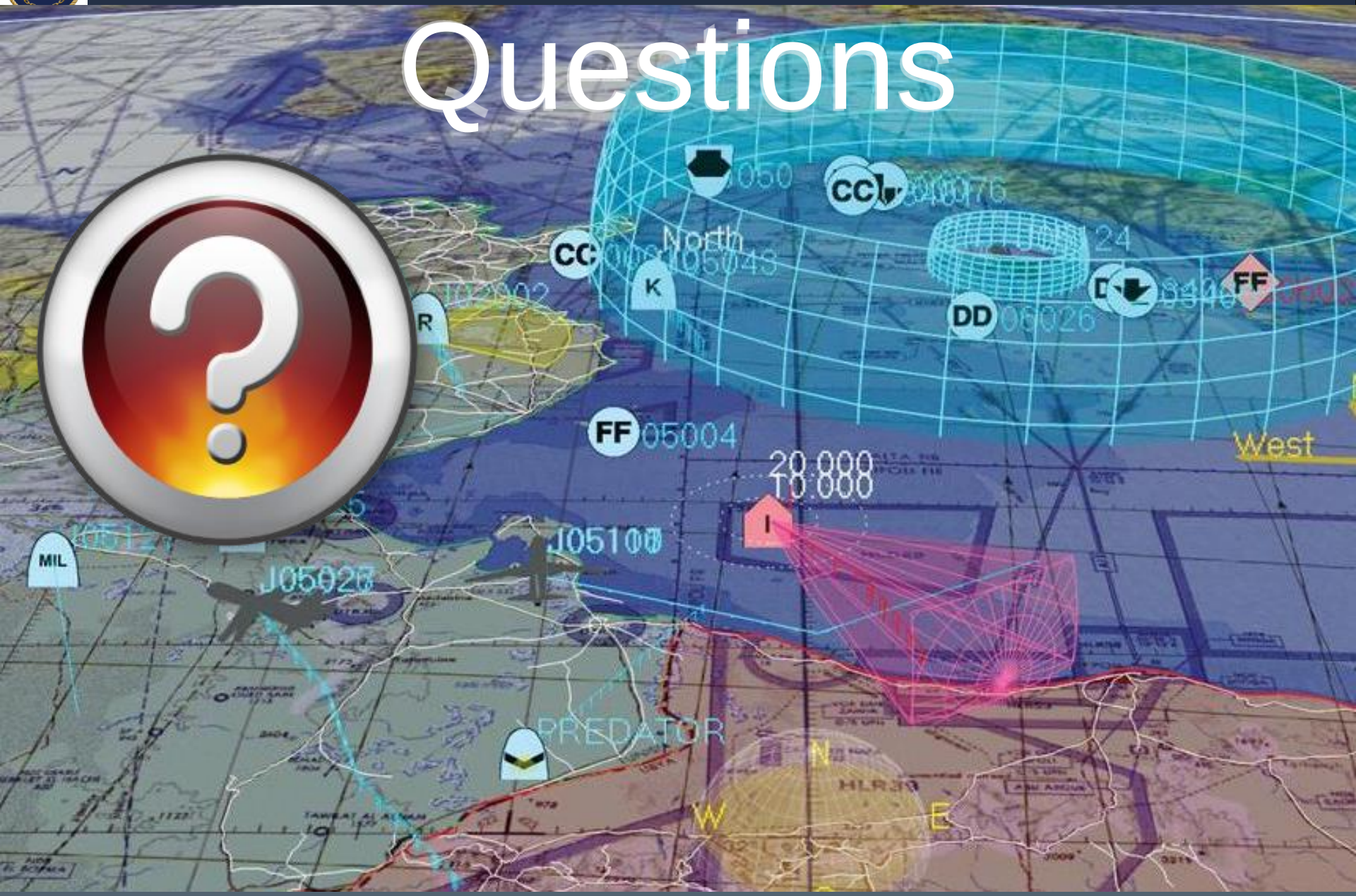
Training, Documentation, User Support Appropriate Hardware & Software

- **Is This "Tool" Fit For This Purpose?**

What Are The Risks If The Wrong Solution To The Credibility Equation Is Used?

These Are What The Accreditor Needs To Know!!

Questions


CC
North
K
DD
FF
FF 05004
20 000
10 000
West
MIL
J05100
J05025
PREDATOR
N
W
E

Biography

Dr. Jim Elele is the Department of the Navy's Lead for VV&A. Dr. Elele was elected into the NAVAIR's Science and Engineering Fellow program in 2014 and he currently serves as a Trusted Agent for VV&A for the Navy's Commander Operational Test Force. He started the NAVAIR's Battlespace Verification and Validation Branch (NAVAIR 5.4.3.7) and served as its Branch Head (from 2007) up until he was selected to lead the Navy's VV&A Program. He has extensive experience in Modeling, Simulation and VV&A, and had worked as Electromagnetic Environmental Effects (E3) engineer for over 10 years. He currently provides VV&A support to various Navy System Commands and acquisition programs, and was the M&S lead for the Marine Corps Heavy Lift Helicopter (CH-53K) Replacement acquisition program during which he invented the NAVAIR Risk-based VV&A Process. He started work as a US Army civilian employee in 1988, and was a member of the technical experts who created the Army's Mobile Subscriber Equipment Performance Model (MSEPAM). He served as data specialist for both the DIS Communications Protocol Committee and the US Army Extended Air Defense Test Bed (EADTB). He was instrumental in defining and creating the Army's Virtual Proving Ground (VPG) under Dr. David Brown of the Army's Test & Evaluation Command. He had also worked as a process engineer for IBM and General Electric (1980 to 1983) prior to joining the US government.

He earned a BS in Chemical Engineering (1980); MS (1985) and PhD (1988) in Applied Mathematics (all from the University of Arizona). Dr. Elele is an adjunct professor for Strayer University and (formerly for) Florida Institute of Technology. He has published over 30 technical papers in Journals & Conferences, and is the owner of 2 patents.

PROFESSIONAL ACCOMPLISHMENTS: <u>MAJOR AWARDS AND PATENTS</u>

- NAVAIR Associate Fellow (2010), NAVAIR Fellow (2014)
- US Army's Civilian Medal of Achievement (1995) for leading the development of the Virtual Proving Ground with Dr. David Brown.
- US Army Test & Evaluations Command's Federal Engineer of The Year; became a Runner up in the US Federal Engineer of the Year (1993).
- US Patent, # 5,720,555 (Feb. 1998) Temperature Indicating Container and Lid Apparatus.
- Invented the NAVAIR Risk-based VV&A process currently submitted for patent application (2012).