Principles of Verification, Validation, Quality Assurance, and Certification of M&S Applications

OSMAN BALCI
Professor

Department of Computer Science
Virginia Polytechnic Institute and State University (Virginia Tech)
Blacksburg, VA 24061, USA

https://manta.cs.vt.edu/balci

M&S Application Quality

- **Objective:** Develop an M&S application that possesses sufficient quality.
- **M&S Application Quality** is assessed by employing many quality characteristics (indicators) such as
  - **Accuracy** (the most important quality characteristic and is assessed by conducting **Verification and Validation**)
  - Dependability
  - Functionality
  - Interoperability
  - Performance
  - Portability
  - Supportability
  - Usability

Importance of the Principles

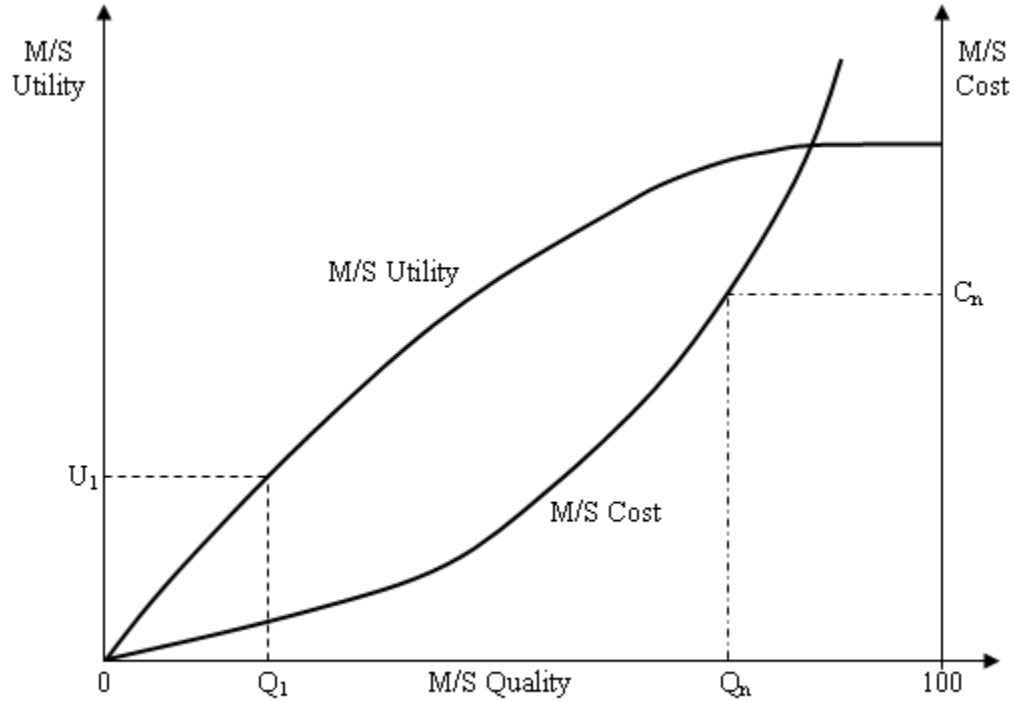
- According to the Webster's dictionary, a principle is defined as
  - "1. an accepted or professed rule of action or conduct.
    2. a fundamental, primary, or general law or truth from which others are derived. 3. a fundamental doctrine or tenet; a distinctive ruling opinion."
  - All three definitions above apply to the way the term "principle" is used herein.
- Principles are crucially important to understand the foundations of Verification and Validation (V&V), Quality Assurance (QA), and Certification.
- The principles presented herein help technical as well as non-technical people better comprehend what QA and V&V are all about.
- They serve to provide the underpinnings for over 100 V&V techniques that can be used throughout the M&S life cycle.
- Understanding and properly applying these principles is a requirement for successful certification of a M&S application.
- The principles are described next in no particular order.

Principle 1: M/S QA and V&V must be conducted hand in hand as integrated within the entire M&S application development life cycle

- Quality is not something that can be imposed after the fact.
- Quality has to be incorporated within the development processes.
- In other words, M/S QA is not a stage or step in the life cycle, but a continuous activity throughout the entire M&S life cycle.
- The QA activities throughout the entire life cycle are intended to reveal any quality deficiencies that might be present as the M&S development progresses from problem formulation to presentation of simulation results.
- This allows us to identify and rectify quality deficiencies at any point in the life cycle.

Principle 2: M/S quality/accuracy assessment outcome should not be considered as a binary variable where M/S quality/accuracy is either perfect or totally imperfect

- Since a model, by definition, is an abstraction, perfect representation is never expected.
- Therefore, M/S quality is not assessed to conclude with a binary decision, where 1 implies "perfect quality" and 0 implies "totally imperfect quality".
- M/S quality must be judged as a degree on a scale from 0 to 100 as shown in the Figure.



Notional Relationships among M/S Quality, M/S Utility, and M/S Cost

Principle 2: M/S quality/accuracy assessment outcome should not be considered as a binary variable where M/S quality/accuracy is either perfect or totally imperfect

- Although M/S quality/accuracy must be judged as a degree on a scale from 0 to 100, the degree cannot be quantified.
- Therefore, we express nominal judgments in assessing M/S quality/accuracy.
- The Table below shows an example set of nominal judgments.

Nominal Score	Numerical Score	Description
Excellent	[90 .. 100]	M/S meets all of the requirements for a particular quality indicator (e.g., accuracy) under a given set of intended uses.
Very Good	[80 .. 89.99]	M/S satisfies most of the requirements for a particular quality indicator under a given set of intended uses.
Satisfactory	[70 .. 79.99]	M/S meets many of the requirements for a particular quality indicator under a given set of intended uses.
Marginal	[60 .. 69.99]	M/S fails to meet most of the requirements for a particular quality indicator under a given set of intended uses.
Deficient	[40 .. 59.99]	M/S is deficient in meeting the requirements for a particular quality indicator under a given set of intended uses.
Unsatisfactory	[0 .. 39.99]	M/S is unacceptable with respect to a particular quality indicator under a given set of intended uses.

Principle 3: A M/S is built for a prescribed set of intended uses and its quality/accuracy is judged with respect to those intended uses

- A model, by definition, is a representation and as such the representation can be created in many different ways depending on the objectives for which the model is intended for use.
- The intended uses of a M/S dictate how representative the M/S should be.
- Sometimes, the nominal score “Very Good” M/S accuracy may be sufficient; sometimes, “Excellent” accuracy may be required depending on the criticality of the decisions to be made based on the M/S results.
- Therefore, M/S quality/accuracy must be judged with respect to a predefined set of intended uses, for which the M/S is created.

Principle 3: A M/S is built for a prescribed set of intended uses and its quality/accuracy is judged with respect to those intended uses

- The adjective "sufficient" must be used in front of terms such as quality, accuracy, verity, validity, and credibility, to indicate that the judgment is made with respect to the prescribed set of intended uses.
- It is more appropriate to say
  - "the model is sufficiently valid"
  - than saying "the model is valid."
  - Here "sufficiently valid" implies that the validity is judged with respect to the prescribed set of intended uses and found to be sufficient.

Principle 4: M/S QA and V&V require independence to prevent developer's bias

- **M/S QA and V&V are meaningful when conducted in an independent manner by an unbiased person, group, or agent who is independent to the M&S application developer.**
- **The M/S developers may be biased when it comes to QA and V&V, because they may fear that negative QA and V&V results may be used for their performance appraisal & may damage their reputation.**
- **M/S QA and V&V should be conducted under technical, managerial, and financial independence.**
  - **Technical Independence implies that the M/S QA and V&V group or agent determines, prioritizes, and schedules its own tasks and efforts.**
  - **Managerial Independence implies that the M/S QA and V&V group or agent reports to the M&S application sponsor independently of the developer organization or group.**
  - **Financial Independence implies that the M/S QA and V&V group or agent is allocated its own budget for the M/S QA and V&V and does not rely on the M&S application development budget.**

Principle 5: M/S QA and V&V are difficult and require creativity and insight

- M/S QA and V&V are difficult due to many reasons including
  - lack of data,
  - lack of sufficient problem domain-specific knowledge,
  - lack of qualified subject matter experts,
  - many qualitative elements to assess, and
  - inability to effectively employ M/S developers due to their conflicts of interest.
- Designing an effective test, identifying test cases, and developing a test procedure to follow require creativity and insight.
- V&V experience is required to be able to determine which of the more than 100 V&V techniques are most effective for a given V&V task.

Principle 6: M/S QA and V&V are situation dependent

- QA and V&V are applied depending on
  - the particular QA and V&V task,
  - M/S type,
  - M/S size,
  - M/S complexity, and
  - the nature of the artifact subjected to QA and V&V.
- The top ten most effective V&V techniques for one QA and V&V situation may not be so for another.
- The QA and V&V approach, techniques, and tools must be selected depending on the QA and V&V task at hand.

Principle 7: M/S accuracy can be claimed only for the intended uses for which the M/S is tested

- M/S accuracy is judged and certified for a particular intended use, which defines the M/S input conditions.
- The M/S that works for one set of input conditions under a given intended use may produce absurd output when conducted under another set of input conditions.
- For example:
  - Assume that a simulation model is developed for the intended use of finding the best light timing for a traffic intersection during the evening rush hour (input conditions).
  - Assume that the simulation model is certified for such use under the stated input conditions.
  - Sufficient accuracy of that simulation model cannot be assumed for other input conditions such as morning rush hour or noon rush hour.

Principle 8: Complete testing is not possible for large and complex models and/or simulations

**"The only exhaustive testing there is,
is so much testing that the tester is exhausted!"**

- Exhaustive (complete) testing requires that the model is tested under all possible input values.
- Combinations of feasible values of model input variables can generate millions of logical paths in the execution of a large and complex model.
- Due to time and budgetary constraints, it is impossible to test the accuracy of millions of logical paths.
- When using test data, it must be noted that the "law of large numbers" simply does not apply.
- The question is not how much test data are used, but what percentage of the potential model input domain is covered by the test data. The higher the percentage of coverage the higher the confidence we can gain in model quality/accuracy.

Principle 9: M/S QA and V&V activities should be considered as confidence building activities

- We are unable to claim sufficient accuracy of a reasonably large and complex M/S with 100% confidence due to M/S complexity, lack of data, reliance on qualitative human judgment, and lack of complete testing.
- The QA and V&V activities are conducted until sufficient confidence is obtained for a particular quality indicator such as accuracy.
- Therefore, M/S QA and V&V activities should be viewed as "confidence building" activities.
- Accuracy is certainly the most important quality indicator and V&V is conducted to assess it.
- However, for a large and complex M/S, we are unable to substantiate sufficient accuracy with 100% confidence. In this case, assessment of other quality indicators helps us increase our confidence in sufficient accuracy of the M/S.

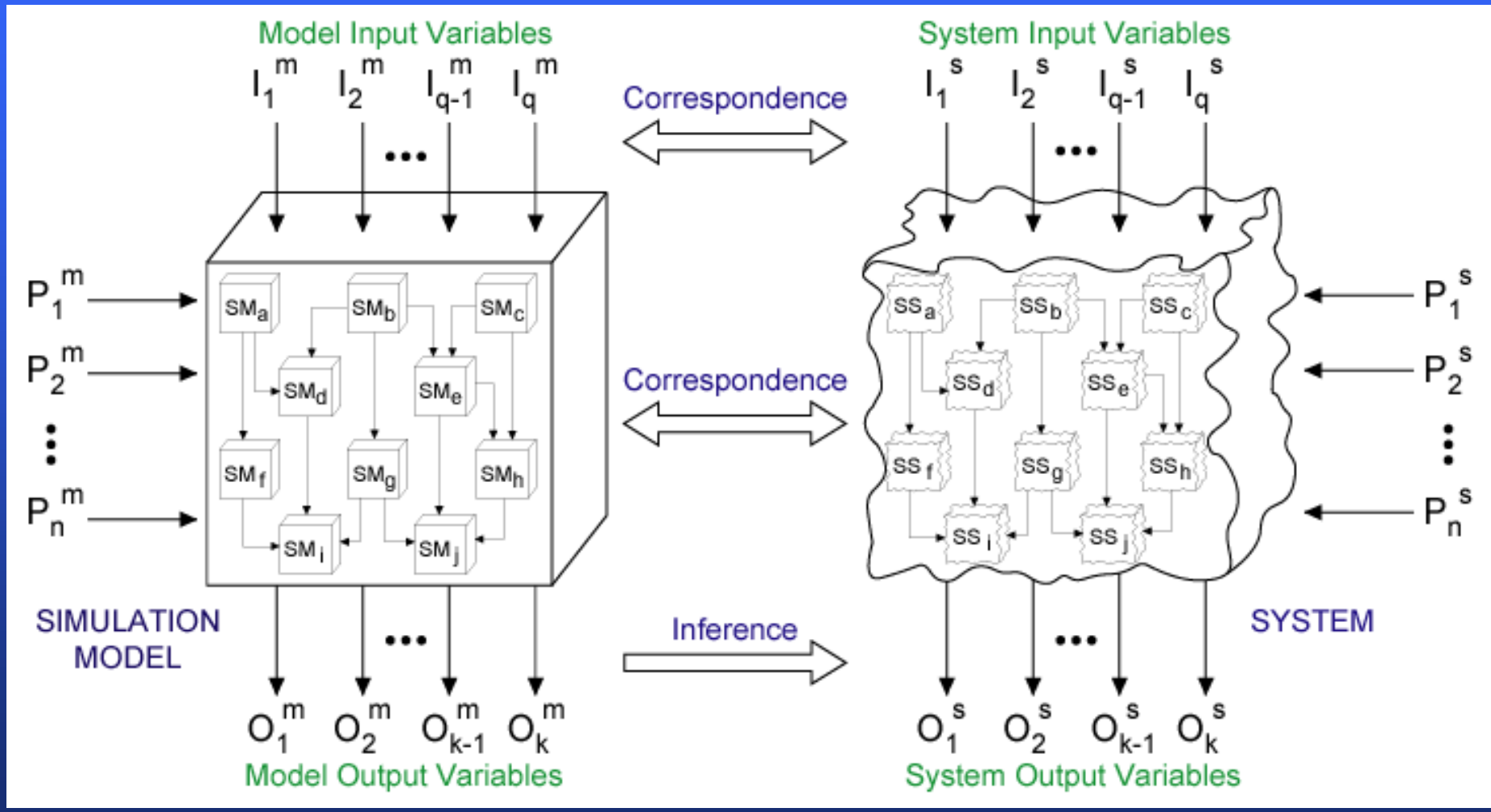
Principle 10: M/S QA and V&V activities must be planned and documented throughout the entire life cycle

- The M/S QA and V&V activities must not be conducted in an *ad hoc* fashion.
- Planning such activities are required for
  a) identifying methodologies and techniques to use,
  b) identifying software tools to acquire,
  c) identifying participants,
  d) assigning roles and responsibilities,
  e) allocating resources such as personnel, facilities, tools, and finances, and
  f) scheduling of QA and V&V tasks throughout the entire M&S life cycle.
- All QA and V&V activities must be documented for certification, regression testing, re-testing, and re-certification.
- All of the test designs, test data, test cases, and test procedures must be documented and preserved for re-use during the maintenance stage of the M&S life cycle.

Principle 11: Errors should be detected as early as possible in the M&S application development life cycle

- M&S application development must start with problem formulation and must be carried out process by process in an orderly fashion in accordance with a comprehensive blueprint of development (i.e., M&S life cycle).
- Skipping the early stages of development and jumping into programming results in an approach called “build-and-fix” and must be avoided.
- Detection and correction of errors as early as possible in the M&S life cycle results in reduced development time and assures better quality.
- Some vital errors may not be detectable in later stages of the M&S life cycle due to increased complexity.
- It is relatively easier to detect, localize and correct errors in an incremental manner as the development progresses.

Principle 12: Double validation problem must be recognized and resolved properly



Principle 12: Double validation problem must be recognized and resolved properly

- A typical validation test is conducted by running the simulation model with the "same" input data that drive the system, and then comparing the model and system outputs to determine how similar they are.
- The amount of correspondence between the model and system outputs is examined to judge the validity of the model.
- However, in conducting this validation test, another validation test should be recognized and performed before this test.
- That validation test deals with substantiating that the model and system inputs match each other with sufficient accuracy.
- This test is also referred to as input data model validation, which must be successfully performed before the model validation test.

Principle 13: Successfully testing each submodel (module) does not imply overall model validity

- Submodels representing subsystems can be tested individually.
- Since a model or submodel, by definition, is an abstraction, perfect representation is never expected and some representation error is allowed.
- Each submodel can be found to be acceptable with respect to the intended uses with some tolerable error in its representation.
- However, the allowable errors for the submodels may accumulate to be unacceptable for the whole model.
- Therefore, the whole model must be tested even if each submodel is individually found to be acceptable.

Principle 14: Formulated problem accuracy greatly affects the acceptability and credibility of simulation results

- It has been said that a problem correctly formulated is half solved or proper formulation of a problem is 50% of its solution.
- The M&S life cycle starts with problem formulation. Based on the formulated problem, the system or domain containing the problem is defined and its characteristics are identified.
- Based on the defined problem domain, M&S requirements are engineered and the requirements become the point of reference for the M/S development throughout the rest of the life cycle.
- An incorrectly defined problem results in simulation results that are irrelevant.
- Formulated problem accuracy greatly affects the credibility and acceptability of simulation results.
- Sufficient time and effort must be expended to properly define the problem.

Principle 15: Type I, II and III errors should be prevented

- Three types of errors may be committed in developing a simulation model:

1. Type I Error is committed when the simulation results are rejected when in fact they are sufficiently credible.
2. Type II Error is committed when the simulation results are accepted when in fact they are not sufficiently credible.
3. Type III Error occurs when the wrong problem is solved and is committed when the formulated problem does not completely contain the actual problem.

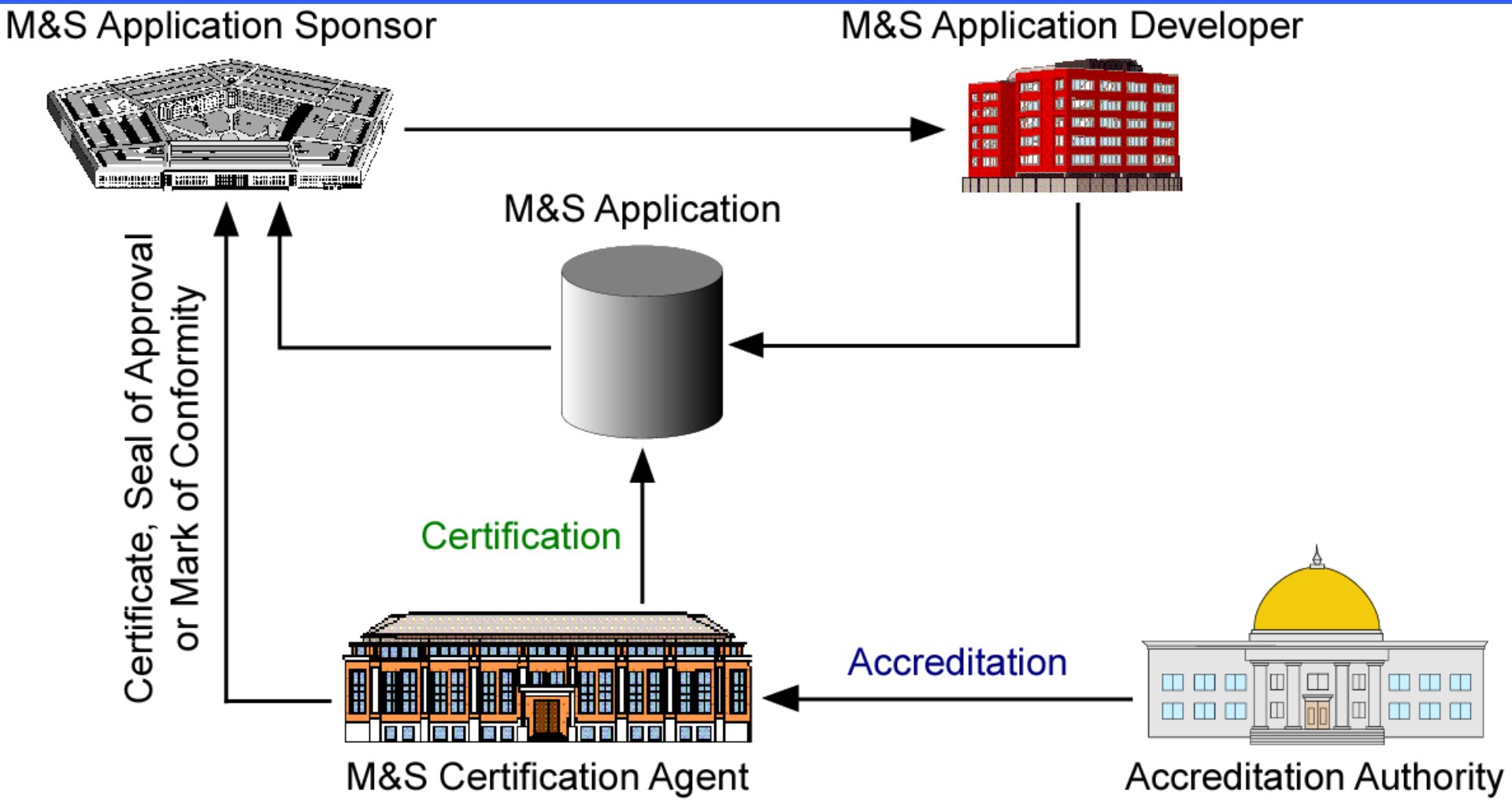
Principle 15: Type I, II and III errors should be prevented

- **Committing Type I Error unnecessarily increases the simulation model development cost.**
- **The consequences of Type II and Type III Errors can be catastrophic especially when critical decisions are made on the basis of simulation results.**
- **Type III Error implies that the problem solution and the simulation results will be irrelevant when it is committed.**
- **Two risks are defined: †**
  - **Model Builder's Risk = probability of committing Type I Error**
  - **Model User's Risk = probability of committing Type II Error**
- **The QA and V&V activities must focus on minimizing these risks as much as possible.**

---

† Osman Balci and Robert G. Sargent (1981), "A Methodology for Cost-Risk Analysis in the Statistical Validation of Simulation Models," *Communications of the ACM 24*, 4 (Apr.), 190-197.

Principle 16: Certification should be conducted by an independent third party



Principle 16: Certification should be conducted by an independent third party

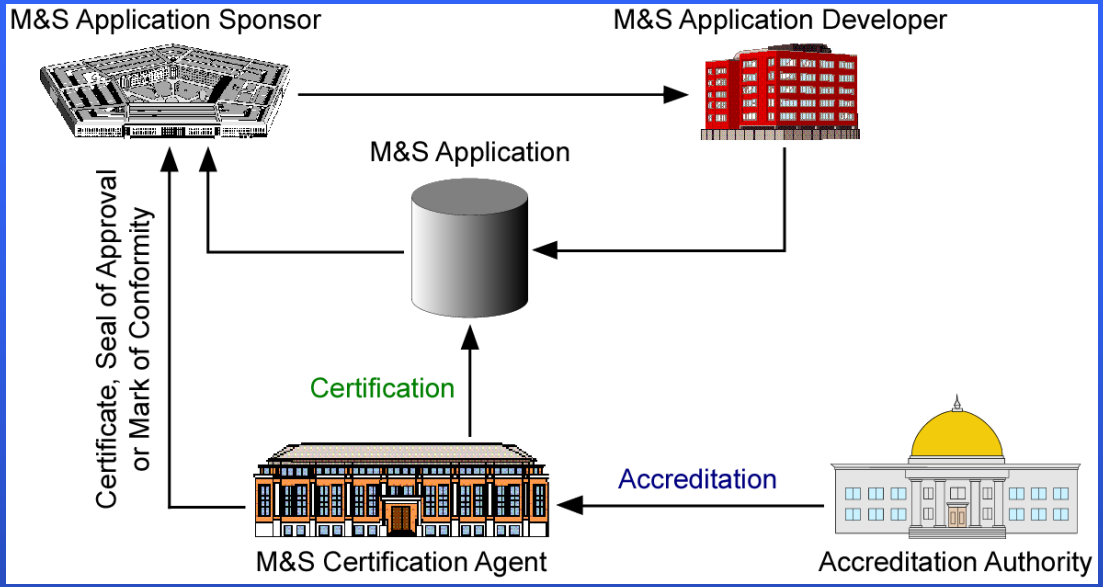
Certification is meaningful if and only if it is conducted in a truly independent manner. True independence requires technical, managerial, and financial independence [IEEE 1998].

- **Technical Independence implies that the certification agent (third party) determines, prioritizes, and schedules its own tasks and efforts.**
- **Managerial Independence implies that the certification agent reports to the M&S application sponsor independently of the M&S developer organization.**
- **Financial Independence implies that the certification agent is allocated its own budget for the M&S certification and does not rely on the M&S development budget.**

Principle 17: Certification should be conducted concurrently throughout the entire development life cycle of a new M&S application

- M&S application sponsor must appoint an independent M&S Certification Agent at the start of the M&S development project.
- For new M&S application development, certification should be conducted concurrently hand in hand with the development activities.
- Concurrent certification enables the discovery of problems throughout the development life cycle before it is too late or too costly to correct.
- For certification of an already developed M&S application with or without modifications, effective and detailed documentation, test cases, test data, and test procedures used during development should be provided to facilitate the certification.

Principle 18: A certification agent should be accredited



- A company or organization interested in serving as M&S certification agent applies to an accreditation authority, which examines the acceptability and maturity of the applicant's certification processes and the qualifications of the key people who execute the certification processes.
- Based on the examination results, the accreditation authority gives formal recognition that the applicant agent is competent to carry out the certification processes and provide certification which is unbiased, fair, cost effective, and consistent.

Principle 19: M&S application sponsor should clearly dictate the rules of conduct between the M&S application developer and M&S application certification agent

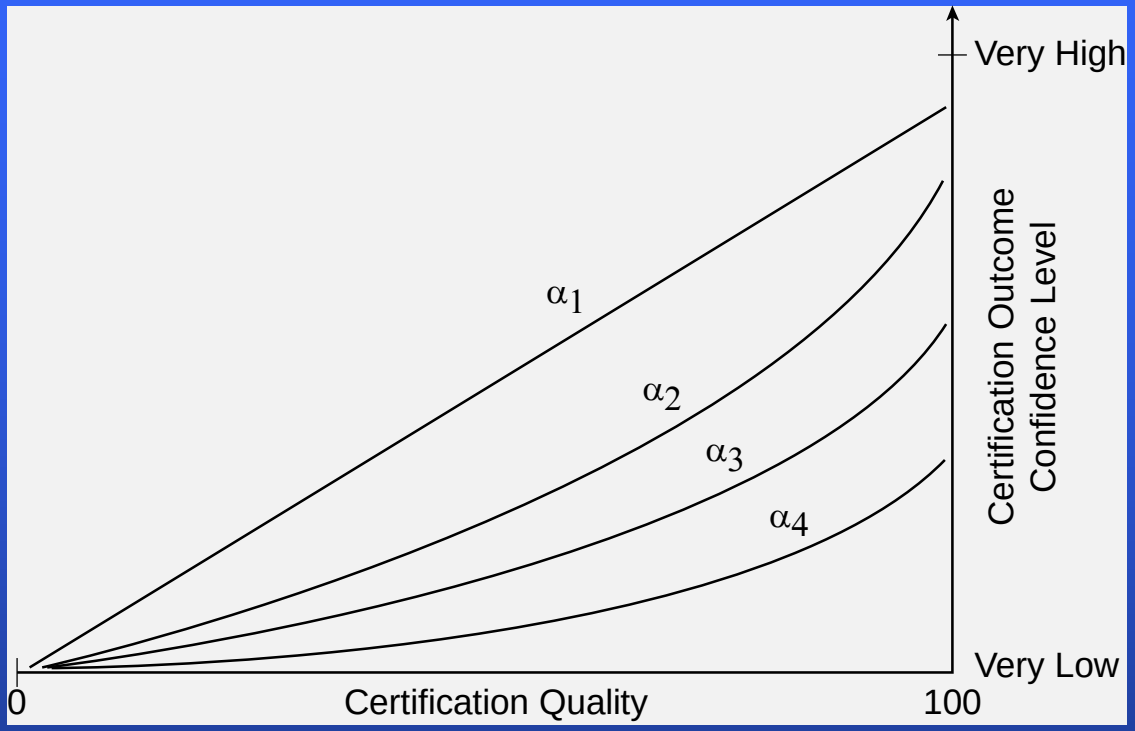
- Successful certification requires the certification agent to have full access to the M&S application with its associated documentation and data.
- However, the M&S developer has full control of the M&S application and might not fully cooperate in providing the required material and information to the certification agent.
- Sometimes, developers view certification as a performance appraisal activity, and they fear that their reputation and potential future funding are at stake if the certification agent identifies problems.
- Therefore, they sometimes show no desire to cooperate and behave in an adversarial manner against the independent certification agent personnel.
- The M&S application sponsor has a critical role in resolving this problem by dictating “rules of conduct”.

Principle 20: Certification outcome should be presented with a level of confidence

- Certification outcome should not be a binary decision, where 1 implies “certified” and 0 implies “not certified.”
- Certification outcome should be reached with a confidence level expressed as a nominal value such as very low, low, average, high, and very high.
- Certification may not be carried out at a desired level of quality due to many factors including lack of data, schedule delays, loss of resources, changing requirements, and development refocus.

Principle 20: Certification outcome should be presented with a level of confidence

- The level of quality at which certification is conducted influences the level of confidence with which a certification outcome is reached as depicted in the Figure.



- As the certification quality increases so does our confidence in reaching the certification outcome.
- The relationship between certification quality and certification outcome confidence level is situation dependent and is notionally illustrated by curves with different α shape parameter values as shown in the Figure.