Office of Director for Developmental Test, Evaluation, and Assessments D(DTE&A)

Model Readiness Levels: A Mathematical Construct for Validation and Trust

December 6, 2021

Darryl Ahner, PhD, P.E. Former Director of STAT COE

AFIT.EDU/STAT

What Are Scientific Test and Analysis Techniques (STAT)?

Scientific Test and Analysis Techniques (STAT) are the **scientific and statistical methods and processes** used to enable the development of efficient, rigorous test strategies that will yield mission assurance during development. STAT encompasses such techniques as design of experiments, observational studies, reliability growth, software testing, and survey design used within a larger decision support framework. The suitability of each method is determined by the specific objective(s) of the test to assist the program manager to understand and quantify technical risk.

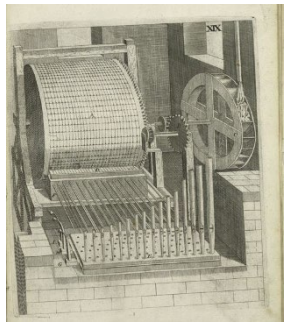
DoDI 5000.89 requires STAT for both developmental and operational testing

STAT: Innovation, Agility, Quality

AGILITY

Implements rapid & responsive solutions to program changes that occur during testing

Continually ensures alignment of desired mission capability with ***measurable*** objectives



INNOVATION

- Efficiently ***generates STAT-based information*** for technology development and testing
- ***Implementation costs are offset*** by test cost savings and post-production cost avoidance

QUALITY

- Test and evaluation ***enables the DoD & DHS to acquire systems that work*** while quantifying risk using STAT
- Produce the required data **to characterize system behavior and combat mission capability** in accordance with DoDI 5000.89 and other guidance
- STAT enables quantitative estimates of ***technical performance requirements*** and produces ***mission-oriented metrics***

We seek for Digital Engineering to enable Innovation, Agility, and Quality as well

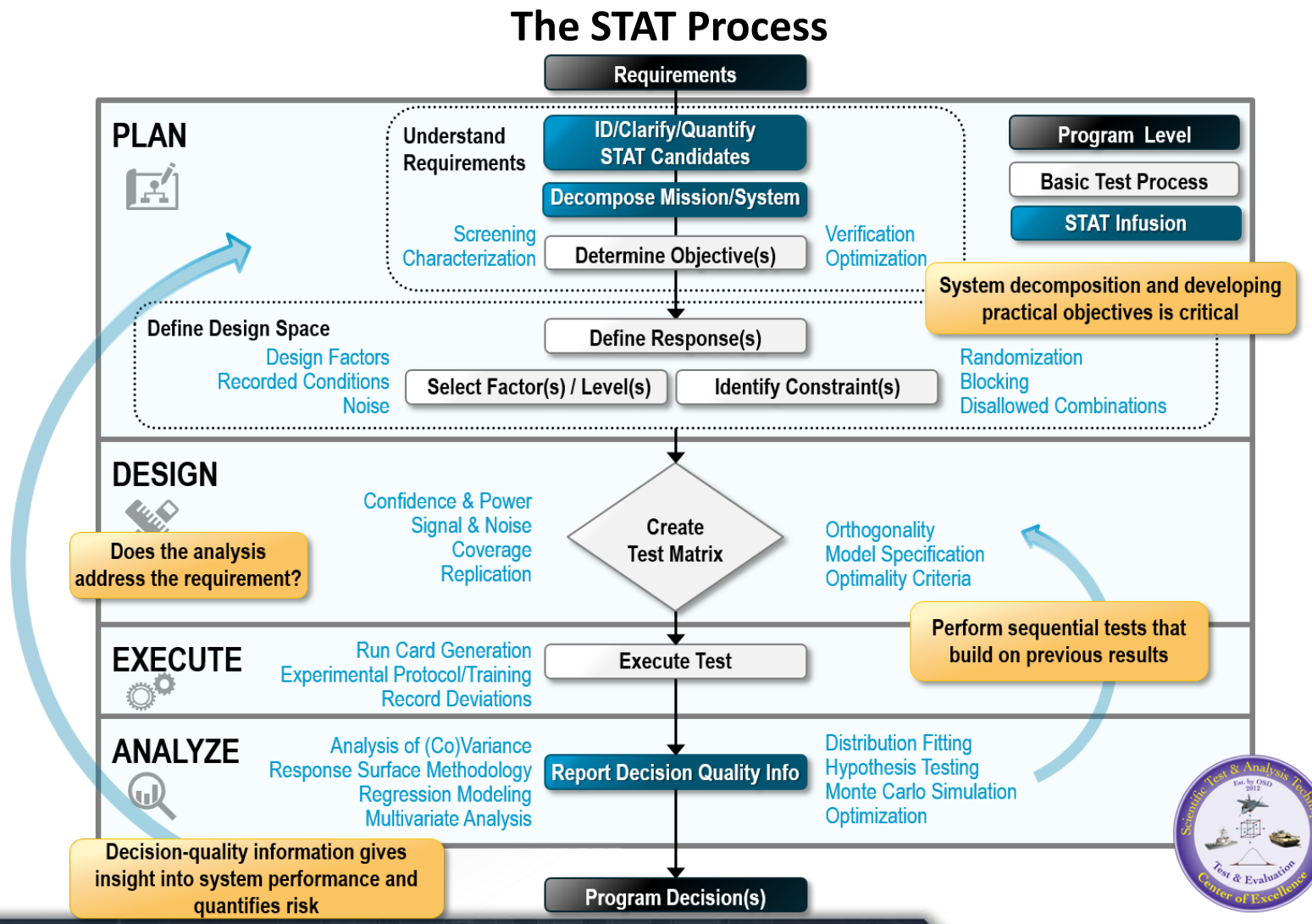
Digital Engineering and Testing in Virtual Environments

- Digital Engineering leverages models in a digital architecture to
  - **Accelerate** engineering design, test, evaluation, & analysis (TE&A)
  - **Virtualize** T&EA to reduce testing costs
  - **Continuously update** models & data to improve processes & procedures
  - **Reuse** models & data to prevent rework and carry forward lessons learned
- Digital Engineering builds on and augments other processes & technologies
  - MBSE, engineering design, Defense Acquisition
- Enterprise Digital Engineering feeds back into future engineering, requirements, logistics, and lifecycle management
- Model validity and readiness are critical to Digital Engineering and can be addressed via quantifiable, STAT-rigorous methods

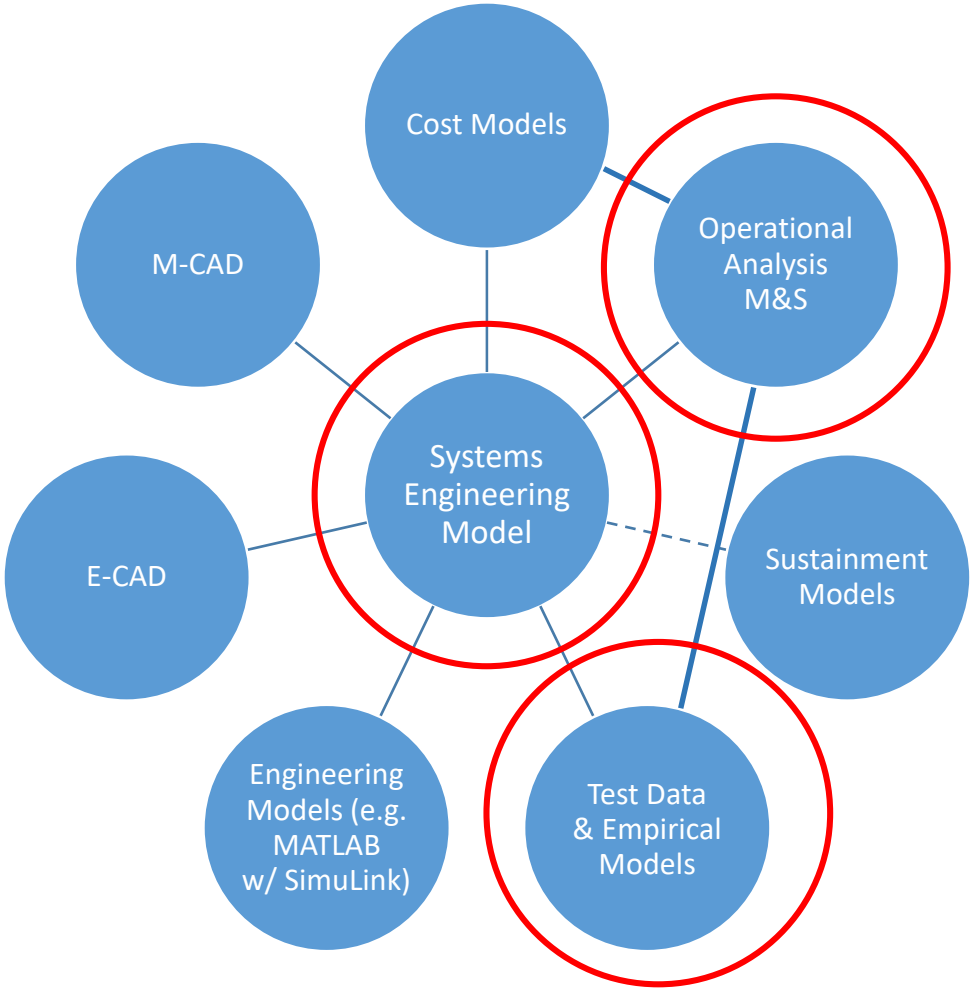
Digital Engineering and T&E support each other to accelerate development while reducing physical investment, STAT supports both

Test Planning in Digital Engineering

- Effective test planning is critical to successful testing and information for model Validation
- Test planning in Digital Engineering presents unique concerns:
  - Physical testing may not be feasible at all stages
  - M&S testing needs VV&A to be trustworthy
  - Difficult to know if models are ready for use when little physical reference data exists for VV&A



Digital Engineering Models

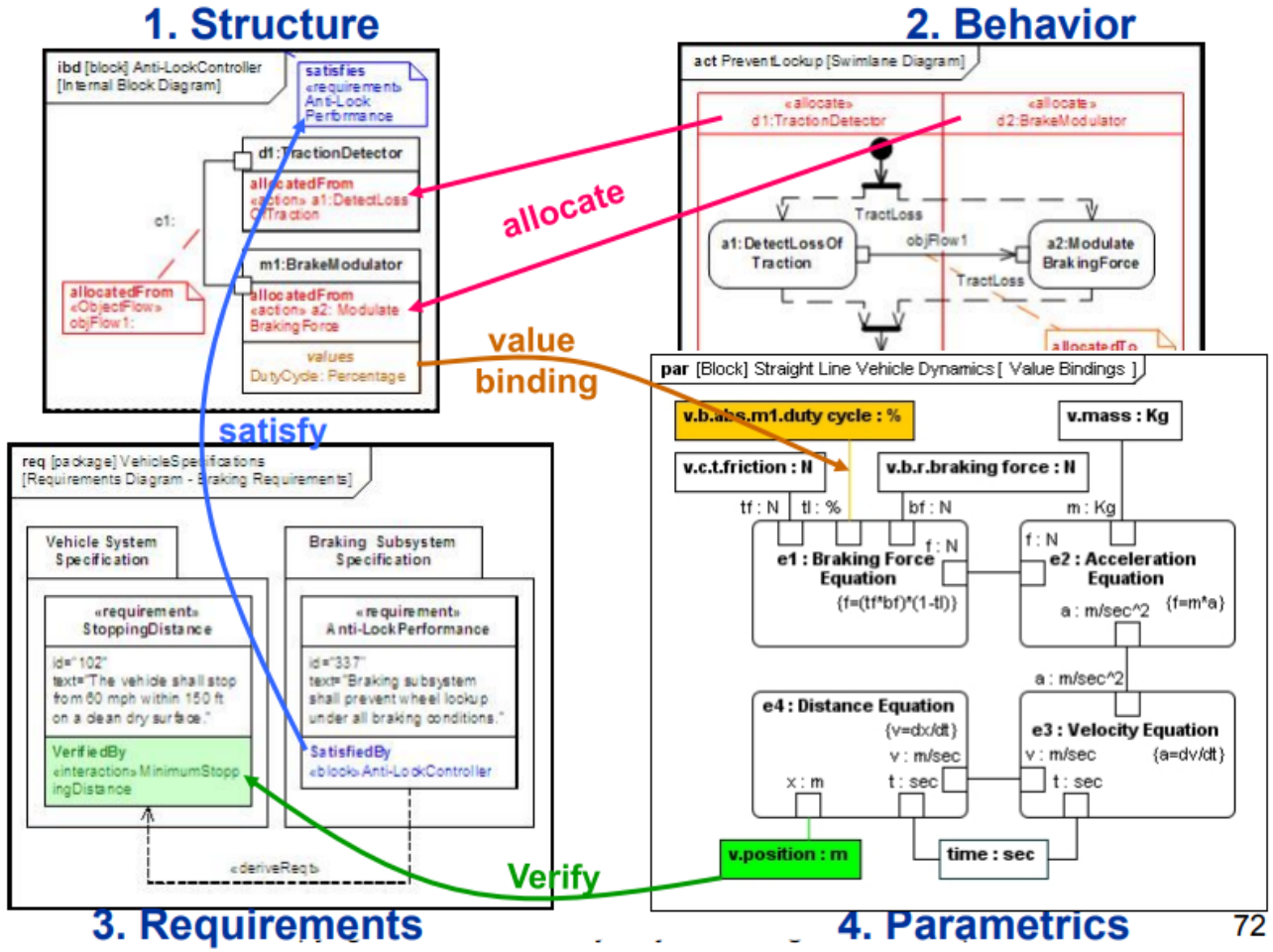


Focus of Today's Presentation

The MODELS are the authoritative source of truth

Model-Based Systems Engineering (MBSE)

- MBSE is a Systems Engineering Framework that tracks system attributes and requirements in a system architecture model
- Built on modeling languages like SysML
  - SysML uses Structure, Behavior, Requirements, and Parametric elements to capture and enforce relationships between system physical, functional, and required attributes
- MBSE provides numerous benefits
  - Provides Requirements traceability
  - Captures component relationships
  - Tracks and applies design constraints
  - Facilitates shared understanding and communication of design



Friendenthal, S, Moore, A., & Steiner, R. OMG Systems Modeling Language (OMG SysML™) Tutorial, 2009

All Digital Engineering models must be consistent

STAT in Physical and Simulated Test Environments for V&V Comparison

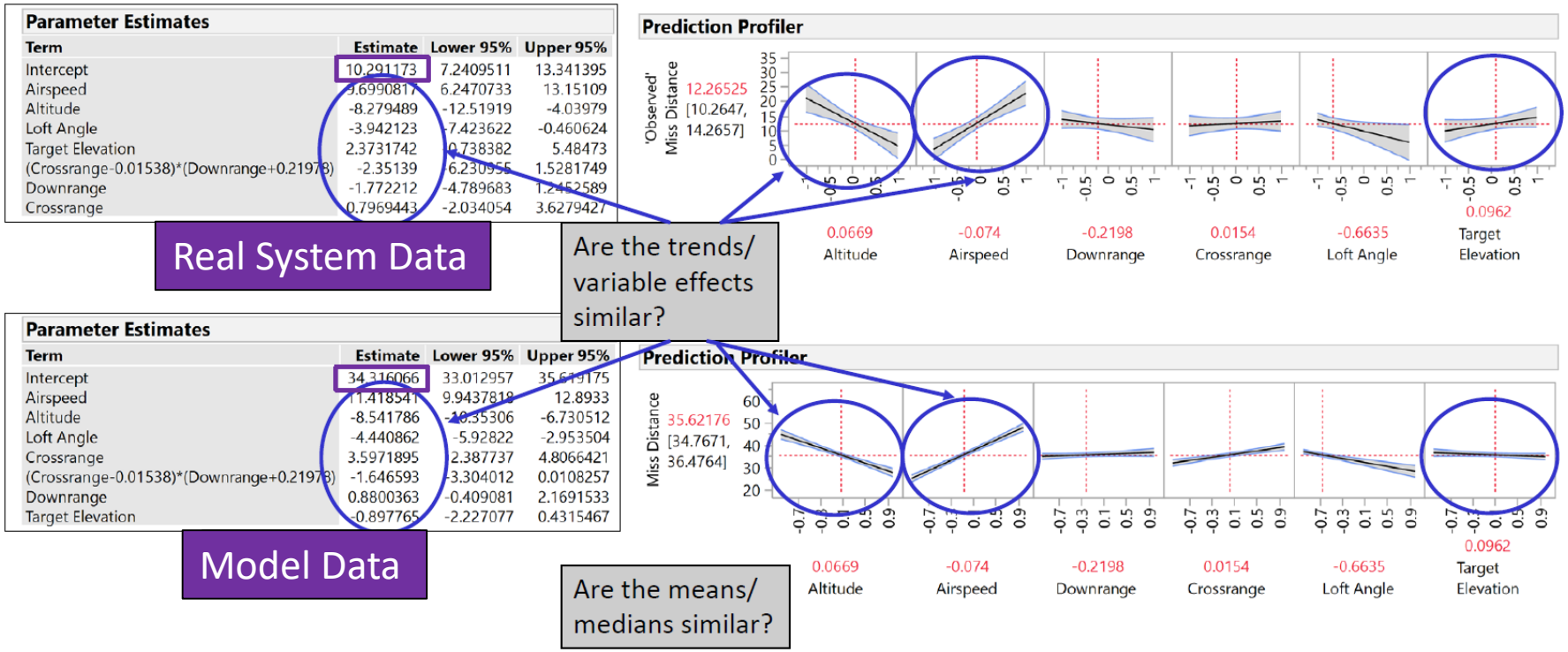
- STAT applies to both physical and simulated (M&S) environments
- STAT still strives for effective and efficient testing
- For successful physical and simulation testing, planning is critical

Physical Testing	Simulation Testing
Typically limited by schedule, test range, and test resources	May involve deterministic simulations with no variability
Typically limited by safety and manpower concerns	STAT methods seek to cover model space as fully as possible
STAT methods seek to understand system variability and coverage	Allow exploration of potentially dangerous operating regions
Meaningful results require rigorous test design	Larger numbers of test runs may be available
	Meaningful results require rigorous test design AND trusted models

Simulation testing seeks to understand the system being represented by a model, not the model itself

STAT for Validation of Models in Simulation

- STAT for validation differs from STAT for virtual testing because validation seeks to assess the quality of the model, not the properties of the modeled system



How does Digital Engineering Support T&E?

- Digital Engineering leverages models in a digital architecture to ensure program ownership of their technical baseline
- DE accomplishes this through:
  - **Reuse** models & data to prevent rework and carry forward lessons learned
    - Enabled by Digital Thread
  - **Virtualize** T&E to reduce testing costs
    - Enabled by Digital Twin and supported by Digital Thread
  - **Continuously update** models & data to improve processes, procedures, operations, & logistics
    - Enabled by Digital Telemetry and supported by Digital Thread
  - **Accelerate** engineering design, test, & evaluation
    - **Leverage the interplay of Digital Twin, Digital Telemetry, and Digital Thread**
- Use Cases (GBSD, T7, Space/ISR systems)

Digital Engineering leads to better systems engineering and acquisition decisions

Model Readiness Levels

- MRLs provide:
  - A measure for model developers to continuously improve their models
  - A means for decision makers to better understand the risk with making decisions based on models
- In order to accomplish this, MRLs need to:
  - Be usable
  - Be comprehensive
  - Have mathematical rigor and consistency
- Existing methods do not address all of these properties

An MRL needs to assess multiple aspects of the model

What Dimensions Matter in an MRL?

Fidelity

Image Credits: STAT COE

Model Referent

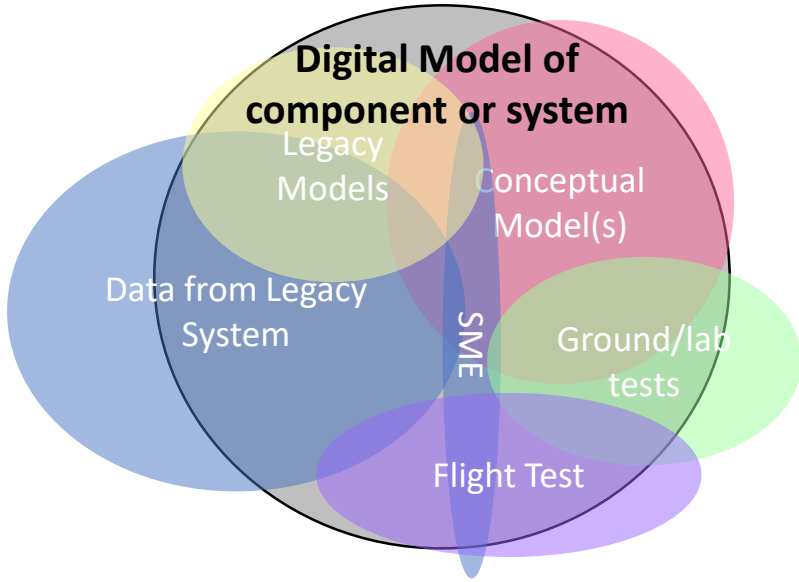
Fidelity quantifies model similarity to referent in 4 dimensions of **detail**

Authority

Authority Level	Relevant Referent
1	SME Judgement
2	First Principles/Physics Predictions
3	Subcomponent Lab Test Data
4	Component Lab Test Data
5	Lab-Scale System Test Data
6	Prototype Test Data
7	HWIL & SWIL Data
8	Live System Test Data
9	Operational Real-World Data

Referent Authority Ranking quantifies trust in **baseline** of comparison

Scope



Scope quantifies degree to which model and referent represent the **same system**

The MRL draws on the 3 Pillars of Validation to ensure the right things are compared, they behave the same way, and we trust the baseline

Dimensions of Fidelity

- **Fidelity**

  The level of detail in the model or simulation relative to the real world.

  - **Resolution**
    - Amount of uncertainty in observed value (M&S output) due to computational limitations
    - Rounding errors, precision of numerical calculations, precision of measurements
  - **Accuracy**
    - The difference between what the model should produce and what it actually produces
    - Are there tendencies to underestimate or overestimate results?
  - **Repeatability**
    - The variability in a *simulation* across runs with the same input conditions
    - Synonymous with statistical definition of precision
  - **Consistency**
    - Limiting behavior of M&S as resolution increases and/or repetitions increase trends towards expected result

Accuracy and Consistency require a valid referent.
Resolution and Repeatability do not (internal checks).

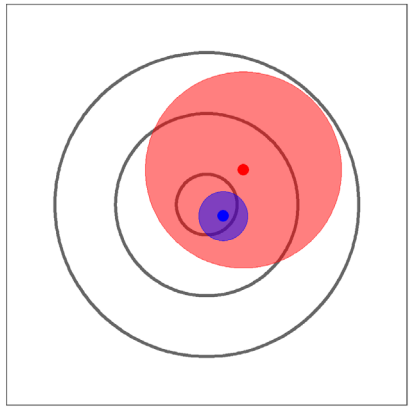
Fidelity

Model

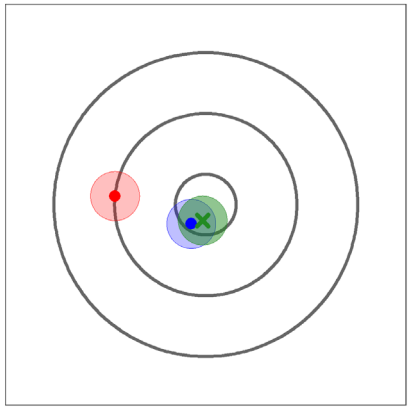
Referent

- Fidelity quantifies the level of detail in a model by its behavior in four dimensions
  - Accuracy:
  - Repeatability
  - Consistency
  - Resolution

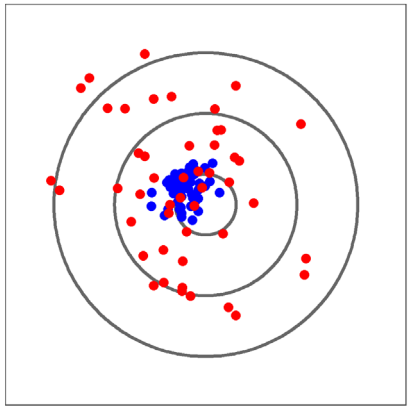
Resolution



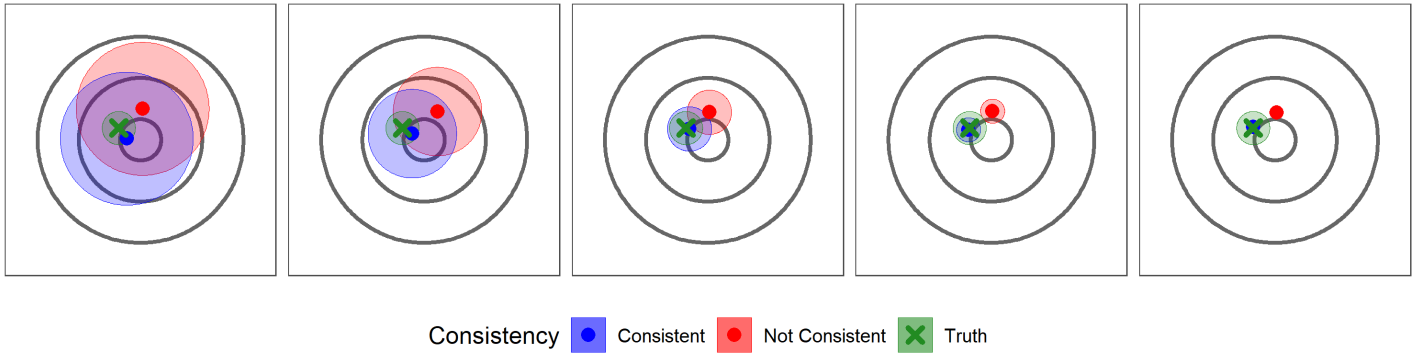
Accuracy



Repeatability



Consistency



Fidelity quantifies model similarity to referent in 4 dimensions of detail

Image credits: STAT COE

Referent Authority

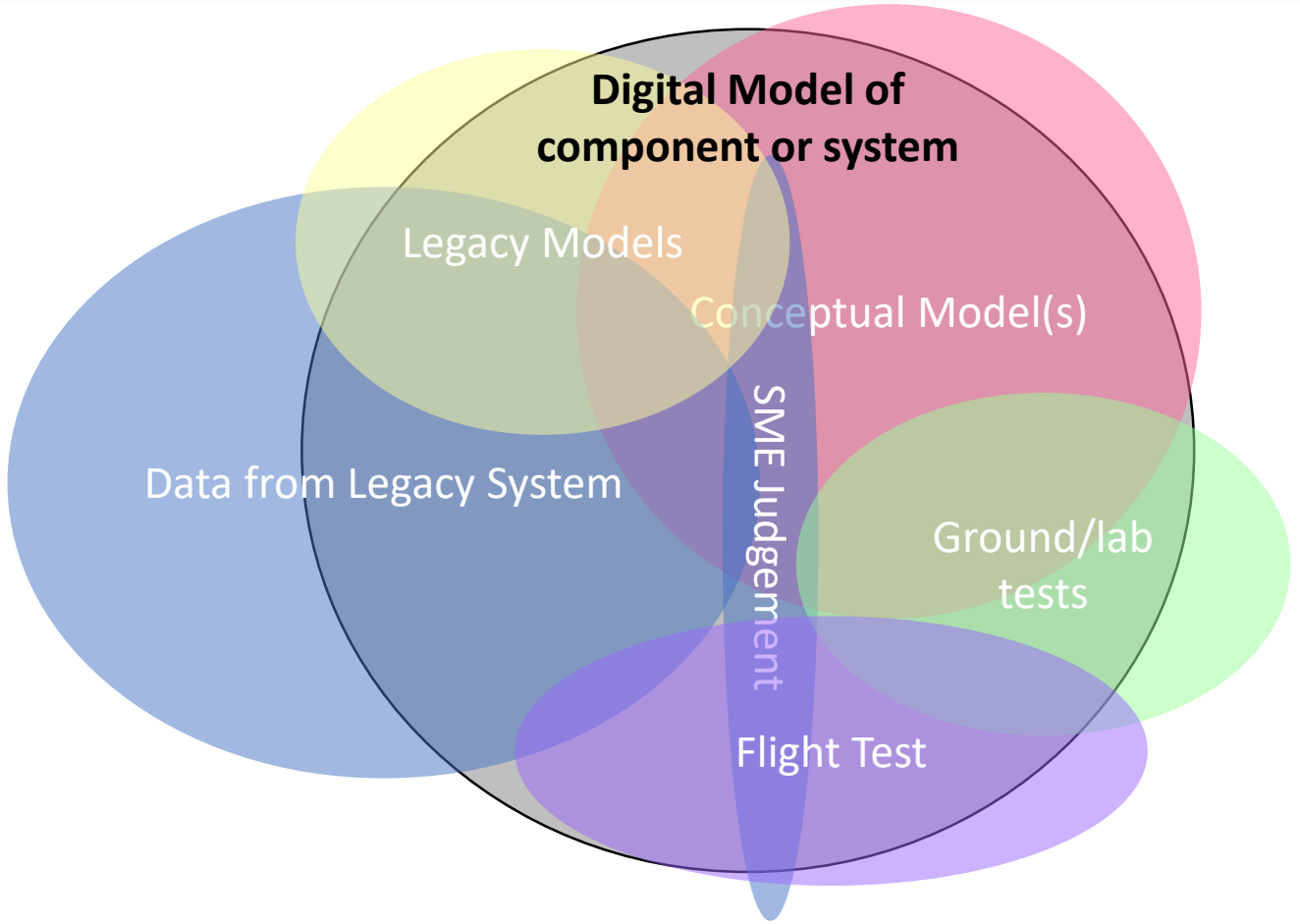
- Referent authority level is similar to TRLs:
  - To declare a technology at a TRL, some evidence must exist to show that the technology has been demonstrated at that TRL
  - That evidence is a referent trusted to that Readiness Level
  - A referent can transfer trust to a model up to its own level of authority
- Multiple referents of the same or different authority levels may be used
  - Referents can be weighted appropriately using a Rank Order Centroid (ROC) method applied at each authority level

Authority Level	Relevant Referent
1	SME Judgement
2	First Principles/Physics Predictions
3	Subcomponent Lab Test Data
4	Component Lab Test Data
5	Lab-Scale System Test Data
6	Prototype Test Data
7	HWIL & SWIL Data
8	Live System Test Data
9	Operational Real-World Data

Referent Authority quantifies trust in baseline of comparison

Scope

- Scope quantifies degree to which model and referent represent the same system
  - Scope compares the modeled variables, effects, and constraints to those present in the referent data
  - Referents which cover more of the model scope can validate more of the model and contribute more to model readiness
  - Several referents may be used to cover the full scope of a model
  - Coverage is multidimensional

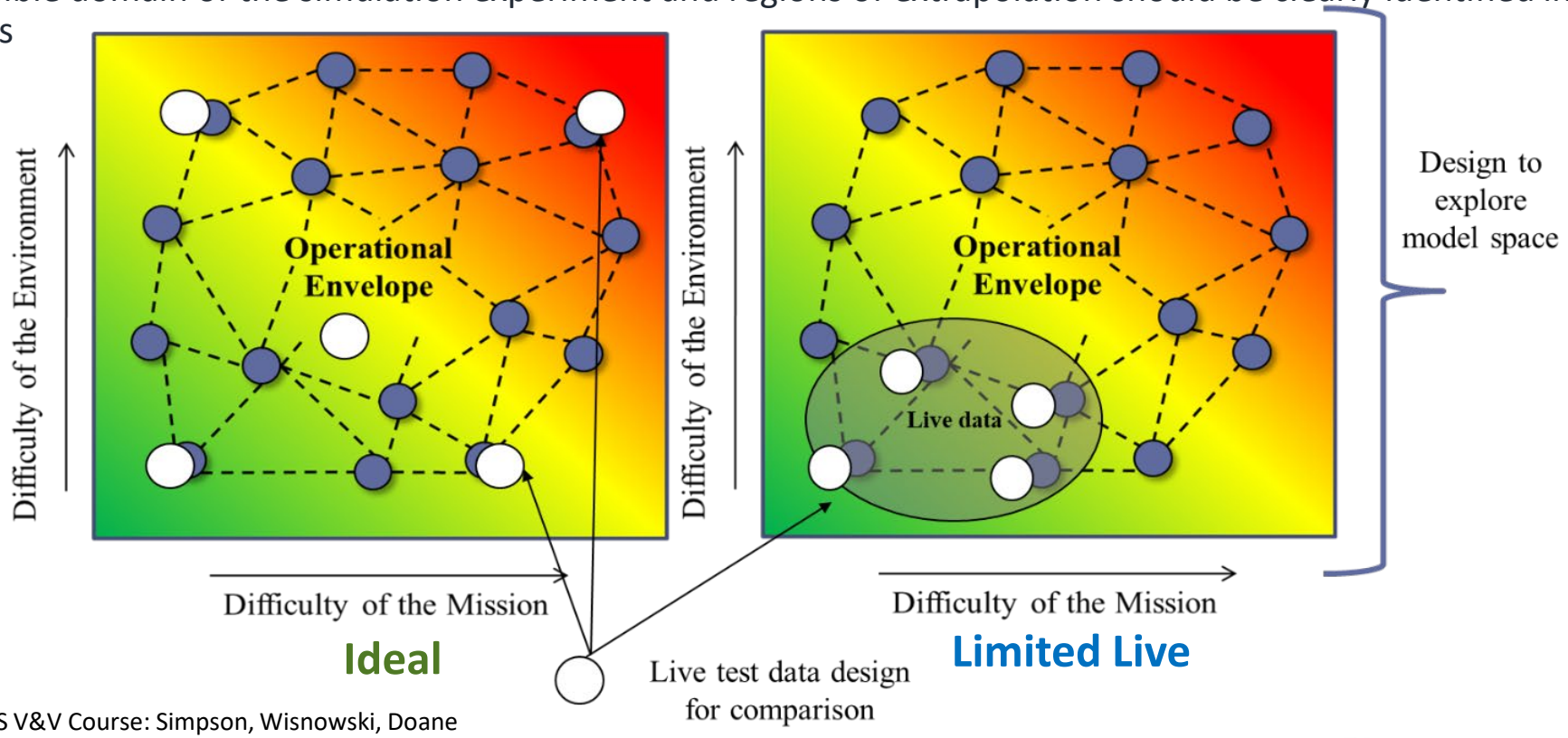


Scope quantifies degree to which model and referent represent the same system

Scope: Test Design Simulation vs. Live

Strategies for selecting the design space and points for live and simulation-based testing

- **Ideal**: the live design encompasses the simulation design space so comparisons between them are interpolations, not extrapolations
- **Limitation in Live Space**: often due to practical constraints that exist in live testing. Here the domain of the live testing should span the maximum possible domain of the simulation experiment and regions of extrapolation should be clearly identified in the validation limitations



Slides from STAT in M&S V&V Course: Simpson, Wisnowski, Doane

Fidelity: A Potential Measure

Fidelity Metric: Exponential Accuracy & Variability

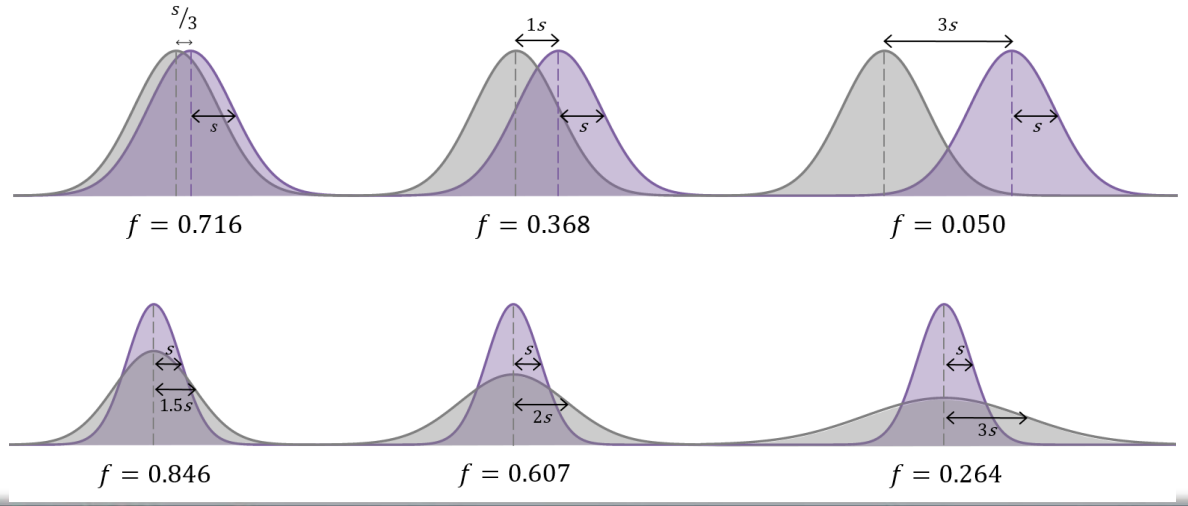
Model

Referent

- Metric assesses both accuracy, repeatability, and resolution
  - Define $s^* = s + \delta$ as variability, where
    - s is the standard deviation, used for representing repeatability
    - δ is the resolution

$$f = f_a f_v = e^{-\left|\frac{\bar{x}_m - \bar{x}_r}{s_r^*}\right|} e^{-\frac{(s_m^* - s_r^*)^2}{s_m^* s_r^*}}$$

  - Requires that all models (stochastic or not) have a measured resolution



- Difference of means equal to one standard deviation:

$$f_a = e^{-\frac{|\bar{x}_m - \bar{x}_r|}{s_r}} = e^{-1} = 0.368$$

- Difference of means equal to three standard deviations:

$$f_a = e^{-3} = 0.050$$

Takeaways

- Digital Engineering provides the potential for increased innovation, agility, and quality
- STAT complements Digital Engineering in two ways:
  - Rigorous and defensible model validation
  - Efficient and effective test design
- MRLs must be useable & comprehensive with mathematical rigor and consistency
- MRLs facilitate continuous model improvement and informed decisions from models
- Test and evaluation is about determining the test events required, ensuring rigorous test planning that feeds effective evaluation

Questions?

More information at www.afit.edu/STAT as it becomes releaseable