"What Does Test & Evaluation Mean in a Digital Engineering Enabled World?"
August 18, 2021 | 1:00 PM ET

Dr. Darryl K. Ahner, P.E.,
Director, Scientific Test and Analysis Techniques Center of Excellence (STAT COE),
Air Force Institute of Technology

TEST & EVALUATION

- ❑ Today's session will be recorded.
- ❑ An archive of today's talk will be available at: www.sercuarc.org/serc-talks/ as well as on the SERC YouTube channel.
- ❑ Use the Q&A box to queue questions, reserving the chat box for comments, and questions will be answered during the last 5-10 minutes of the session.
- ❑ If you are connected via the dial-in information only, please email questions or comments to SERCtalks@stevens.edu.
- ❑ Any issues? Use the chat feature for any technical difficulties or other comments, or email SERCtalks@stevens.edu.

SYSTEMS ENGINEERING RESEARCH CENTER

"What Does Test & Evaluation Mean in a Digital Engineering Enabled World?"

SERC TALKS

Dr. Darryl K. Ahner, PE
Director, Scientific Test and Analysis Techniques Center of Excellence (STAT COE),
Air Force Institute of Technology

TEST & EVALUATION

Dr. Laura Freeman, SERC Research Council Member, Director of the Hume Center's Intelligent Systems Lab and Assistant Dean for Research in the College of Science, and Research Associate Professor in the Department of Statistics at Virginia Tech

The Systems Engineering Research Center (SERC) is a federally funded University Affiliated Research Center managed by Stevens Institute of Technology.

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

Test & Evaluation in a Digital Engineering Enabled World

August 18, 2021

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

AFIT.EDU/STAT

Scientific Test and Analysis Techniques (STAT) Center of Excellence (COE)

- Established and funded by DASD(DT&E) in 2012 to provide ***Independent*** *graduate degree level technical STAT skills*
  - Plan to ***increase T&E efficiency & effectiveness***
  - COE directly supports ***DOD major acquisition programs***
- Mission: Provide independent advice and assistance to designated acquisition programs in the application of scientific test and analysis techniques in the development of test & evaluation strategies and plans
- Functions
  - Work directly with Program Managers supporting their efforts for more rigorous T&E
  - Provide technical assistance to the DD(DTEP) and DAU as requested
  - Capture STAT best practices for wider dissemination across acquisition community
  - Develop case studies that exemplify appropriate use of STAT in achieving more rigorous T&E
  - Identify STAT research needs, conduct research when appropriate, and engage the academic community
  - Provide training at the point of need to ensure program led rigor in testing
- DHS STAT COE since 2017

STAT enables more scientific but practical test design .

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.

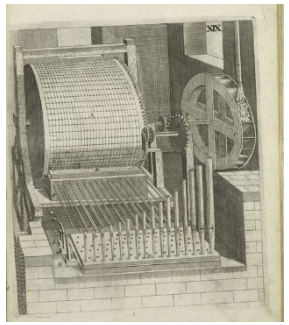
DoDI 5000.89 requires STAT for both developmental and operational testing

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***

The Role of Developmental Test & Evaluation

Testing justifies its cost if:

Enabling Resource:

Scientific Test and Analysis Techniques

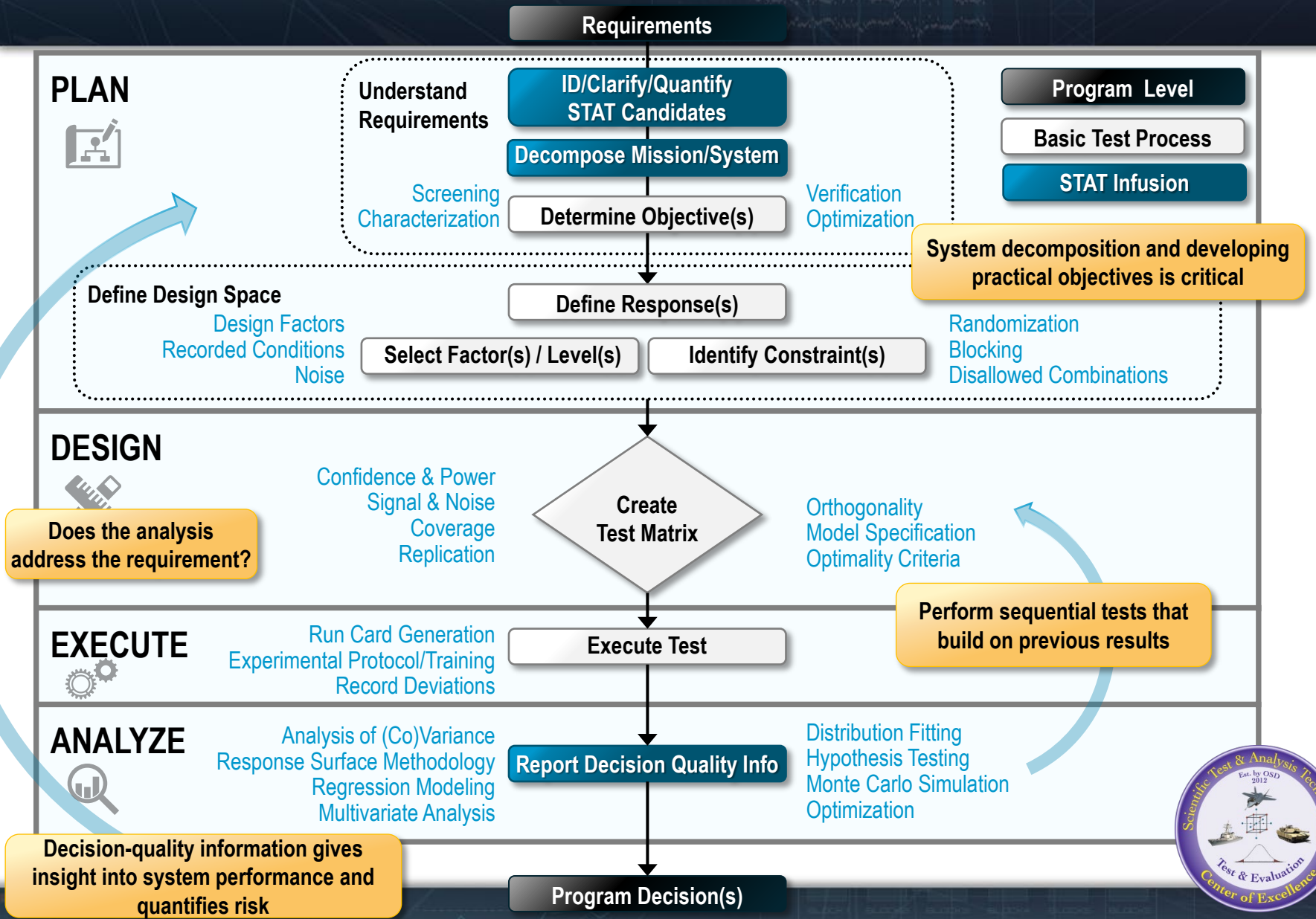
- it yields quality information (effective)
- it doesn't waste resources or time (efficient)
- it informs decisions
- it results in programs better understanding their technical baseline
- it identifies cause when desired performance is not achieved

Developmental Evaluation Framework

- it is conducted where system performance is uncertain
- it is conducted where lack of system performance creates risk to the mission

Well thought-out sequential test strategies have the potential during technology and system development to enable agility, inform innovation, and ensure quality
"Build a little, test a little, learn a lot" -- Adm. Wayne E. Meyer

STAT Process



STAT COE Complex System Challenge List

1. Efficient and effective test of cyber-physical systems to inform technical baseline knowledge and reengineering (often using design of experiments)
2. Reliability testing
3. Software testing (DevSecOps)
4. Validation of modeling and simulation
5. Test information integration within digital engineering
6. Model-based systems engineering informing test planning
7. Qualified system-specific test personnel
8. Test range complexity and realism
9. Interoperability testing
10. Cyber testing

Test planning for complex systems requires appropriate resources and a methodical approach (STAT process) to achieve information to inform development risk and possible reengineering

Digital Engineering and Testing in Virtual Environments

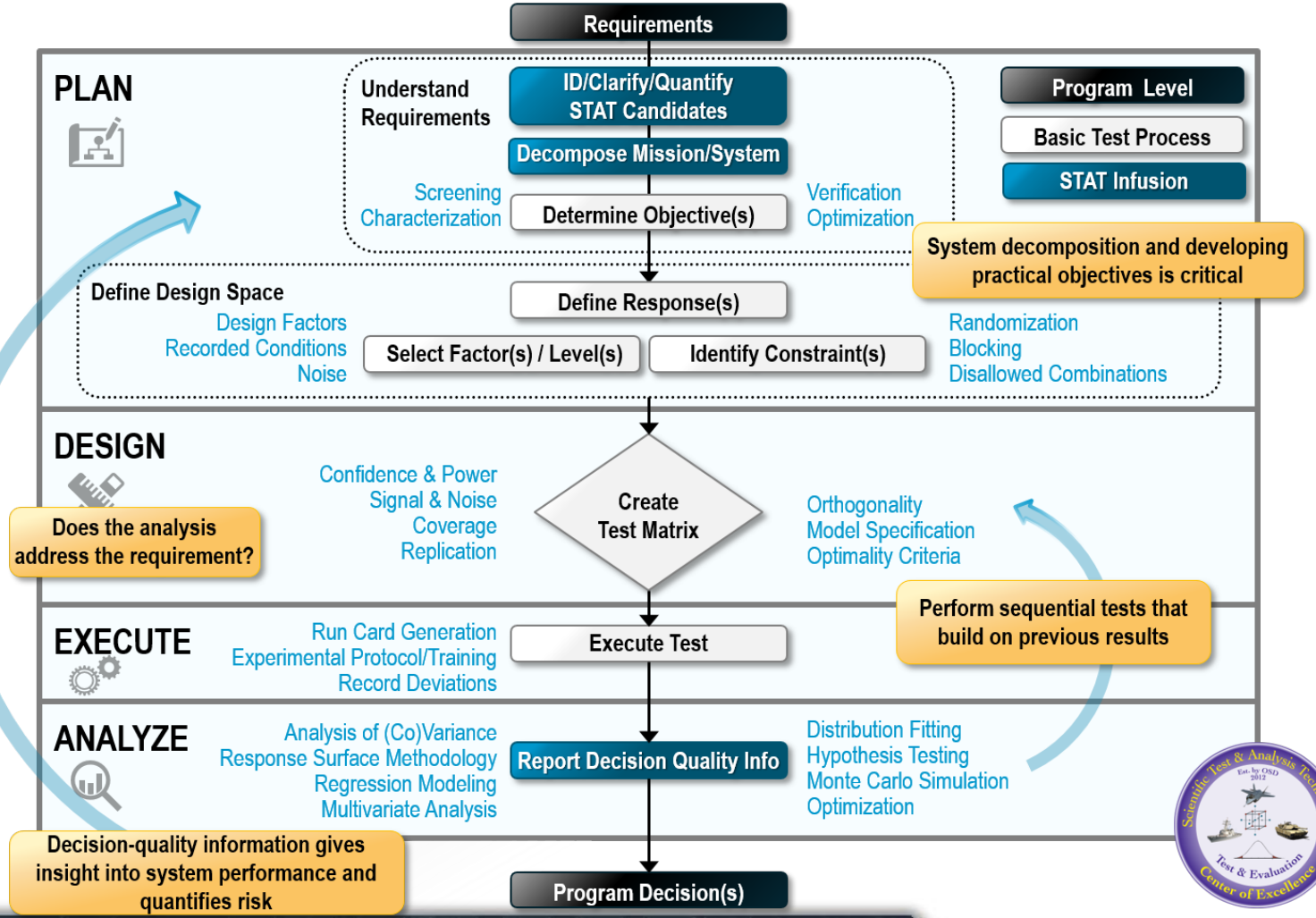
- Digital Engineering leverages models in a digital architecture to
  - **Accelerate** engineering design, test, evaluation, & analysis (TE&A)
  - **Virtualize** T&EA to reduce testing costs
  - **Continuously update** models & data to improve processes & procedures
  - **Reuse** models & data to prevent 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

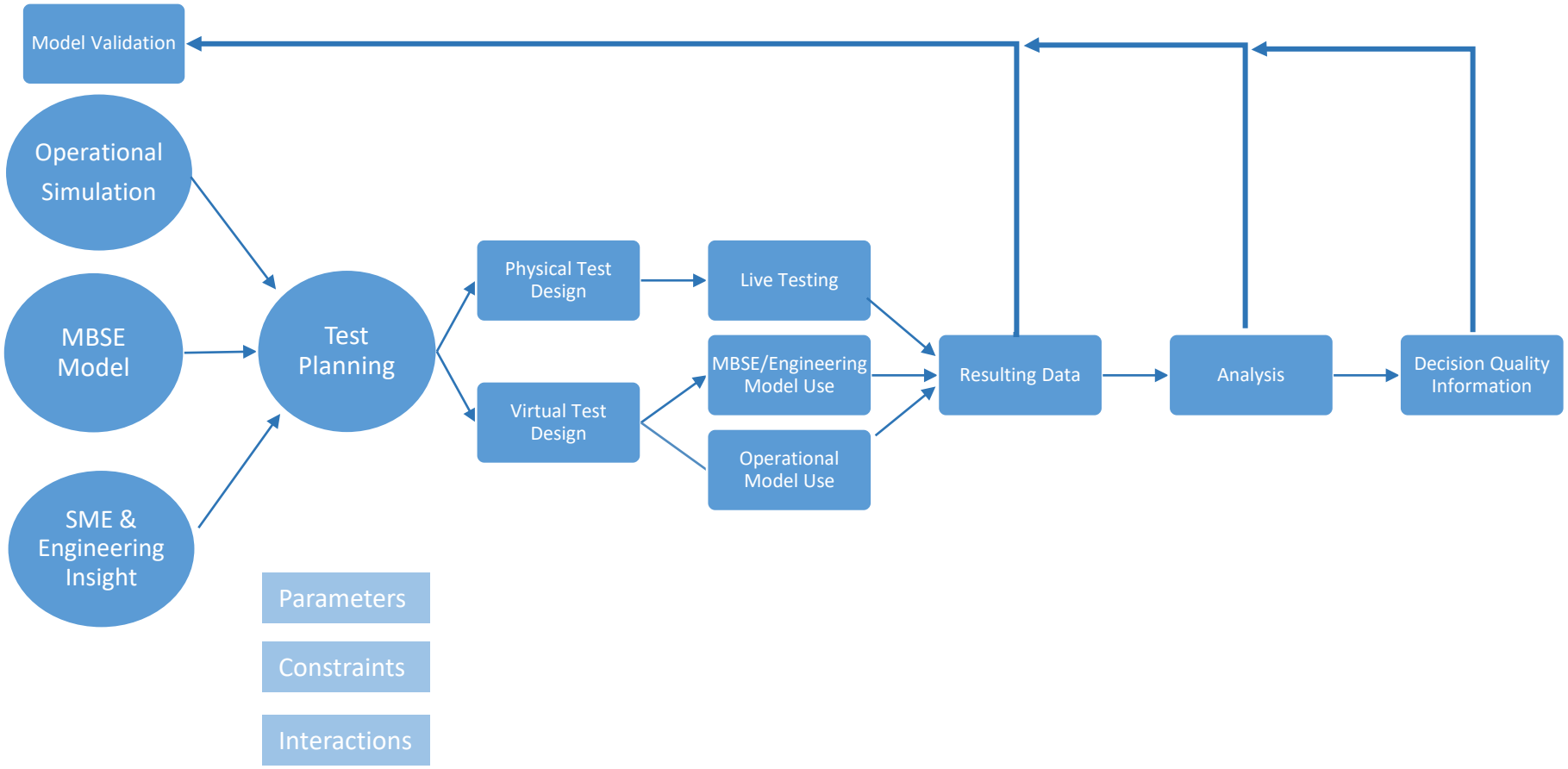
Test Planning in Digital Engineering

- Effective test planning is critical to successful testing
- 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

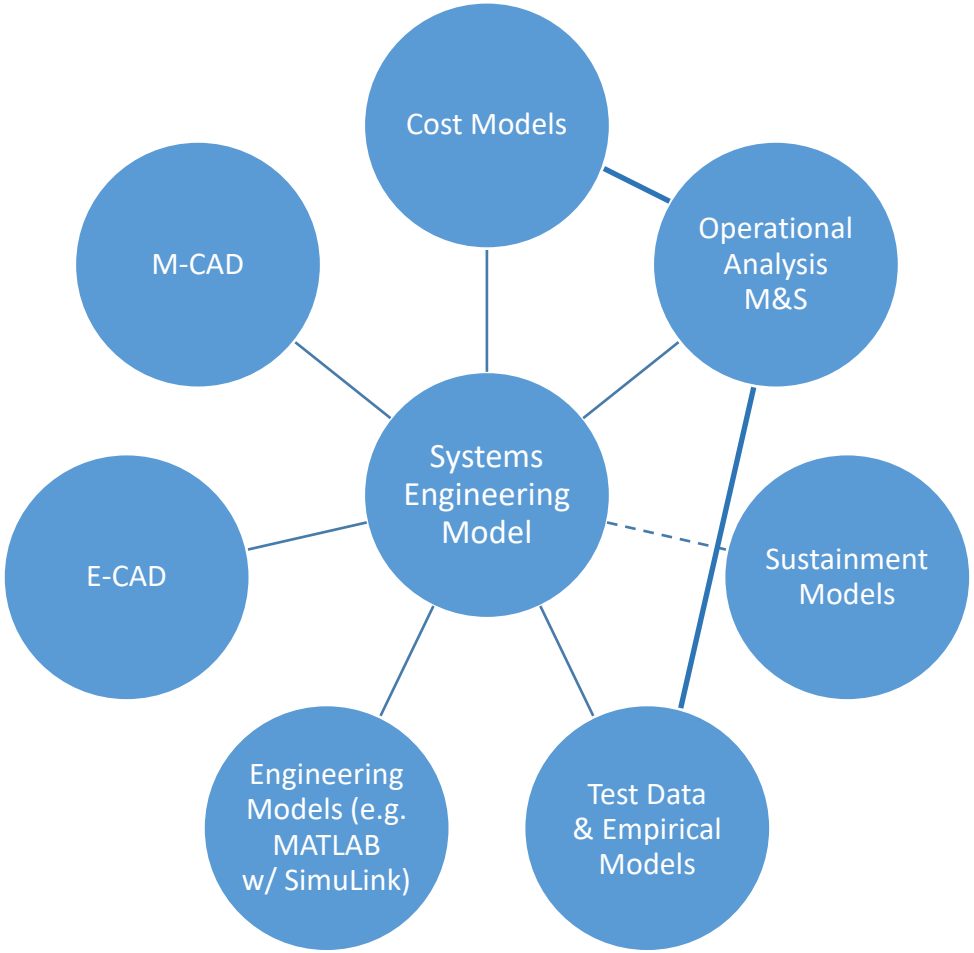
The STAT Process



Testing within Digital Thread



Digital Engineering Models



STAT in Physical and Simulated Test Environments

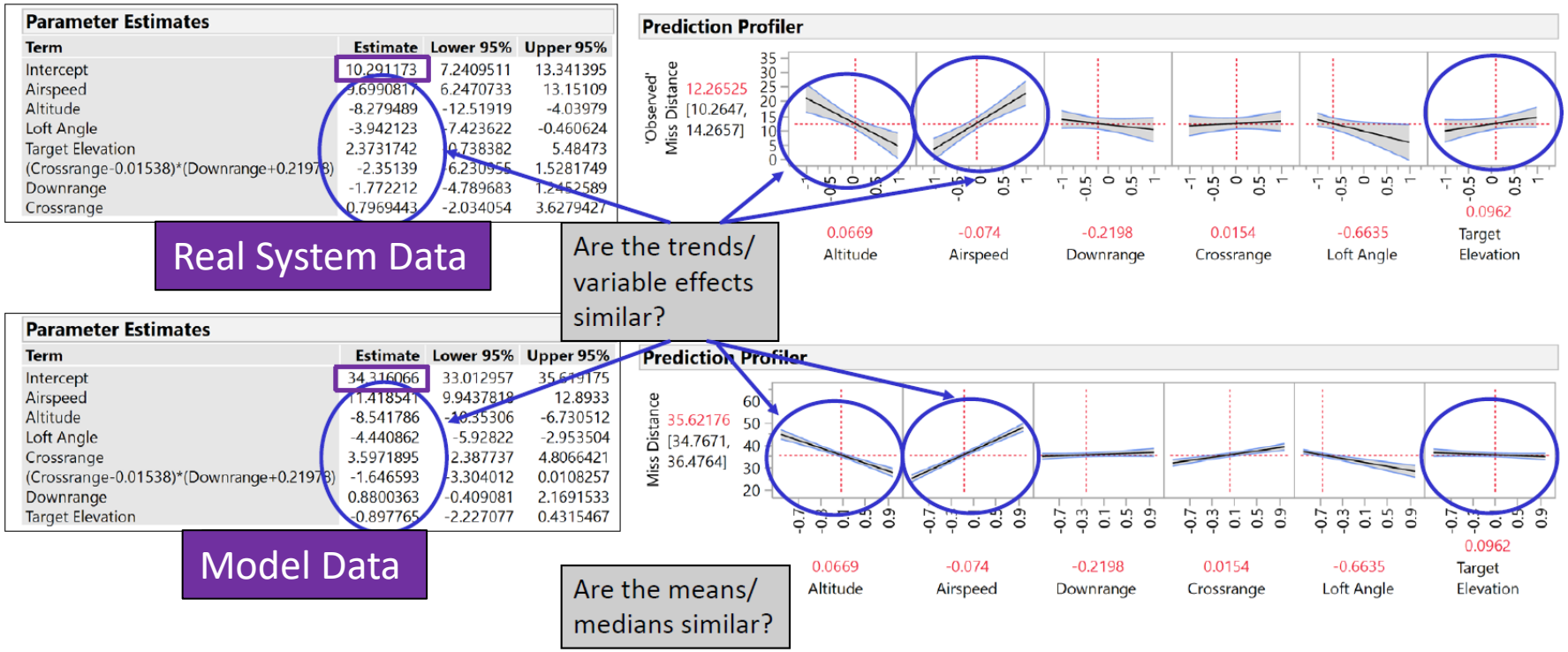
- STAT applies to both physical and simulated environments
- STAT still strives for effective and efficient testing
- For successful physical and simulation testing, planning is critical

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

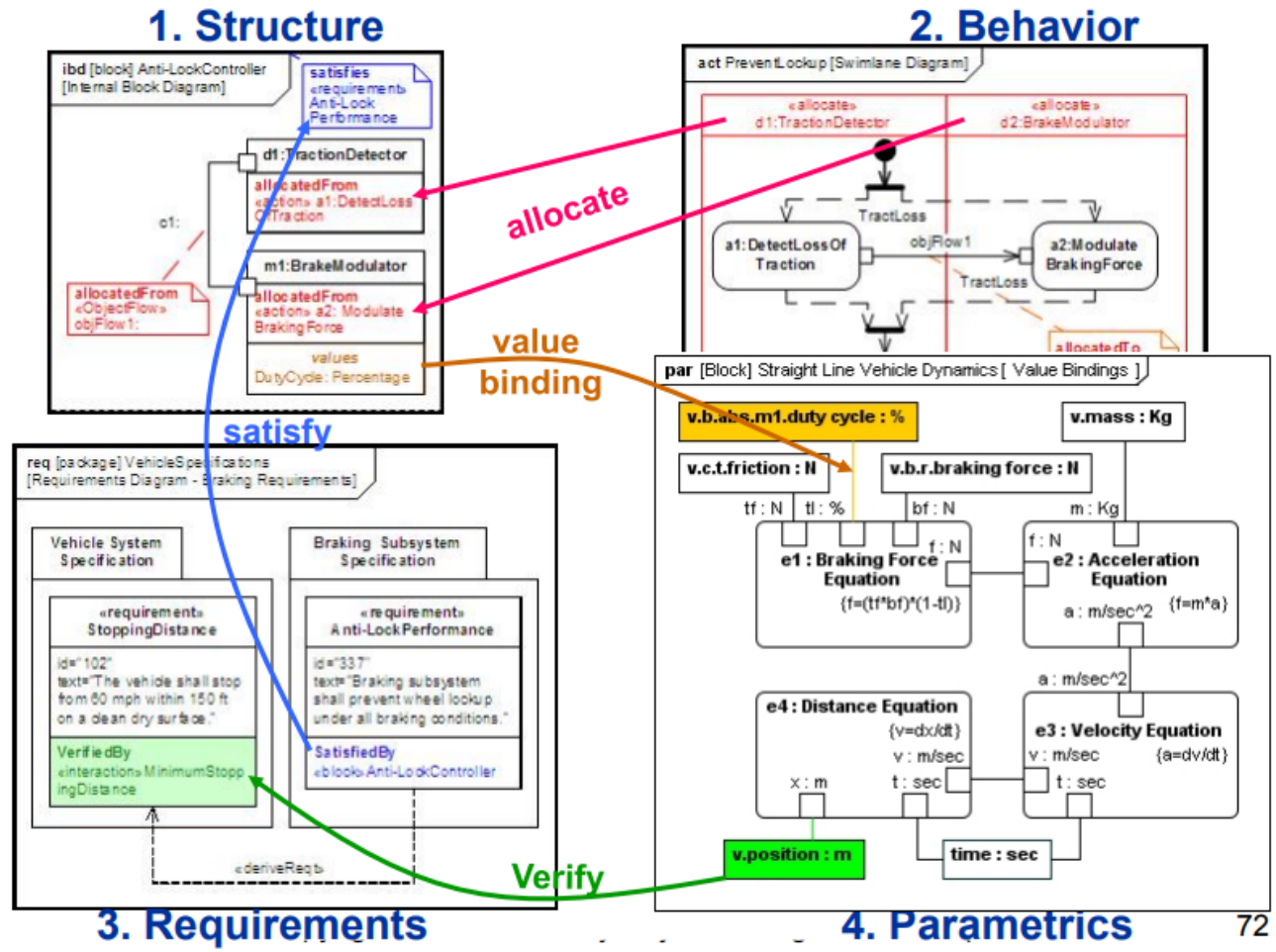
STAT for Validation of Models in Simulation

- 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



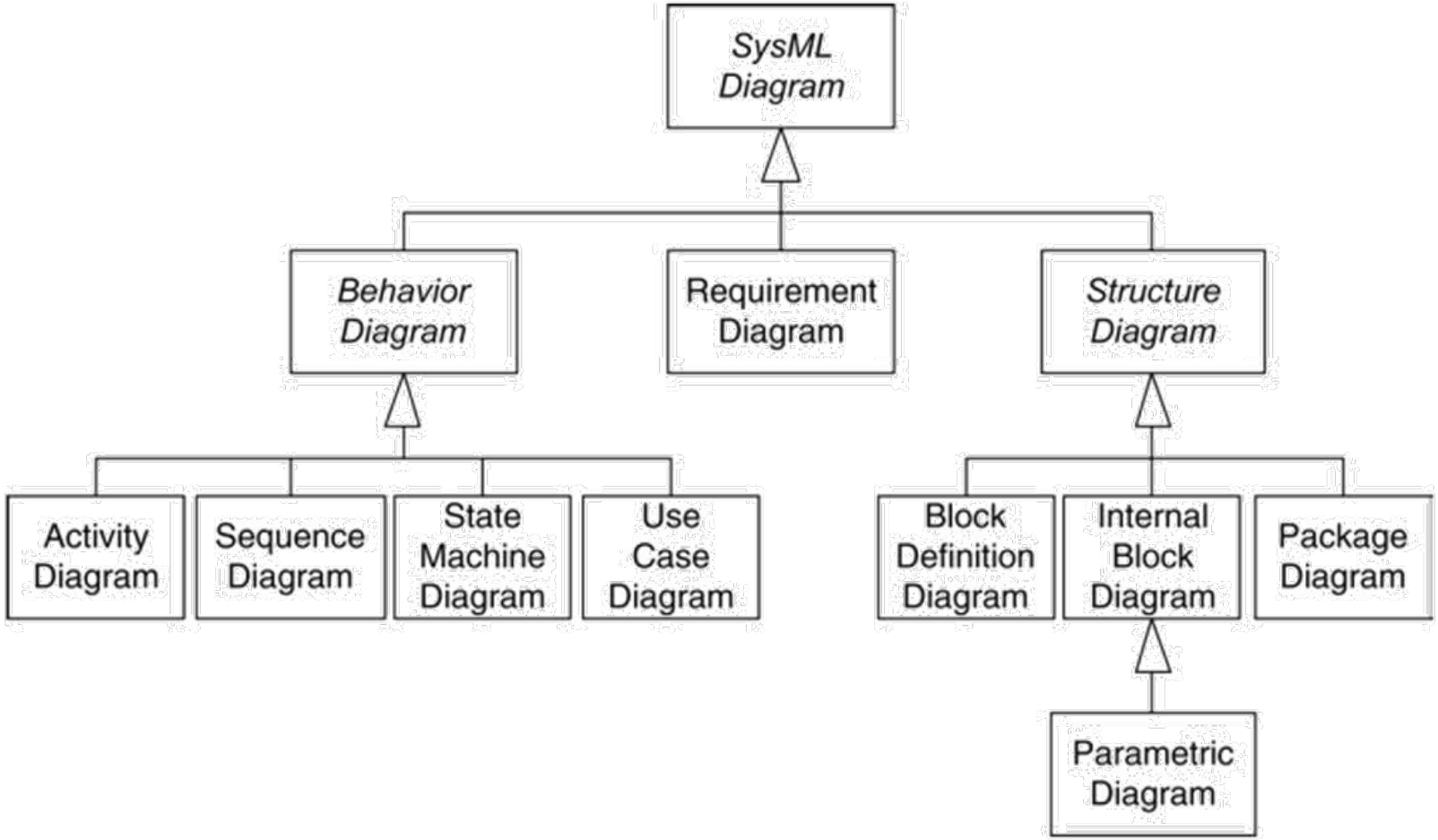
Model-Based Systems Engineering (MBSE)

- 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

SysML Diagram Taxonomy



Open arrow head indicates the item at the head is a generalization of the unadorned end

Delligatti, Lenny (2013). SysML Distilled: A Brief Guide to the Systems Modeling Language

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

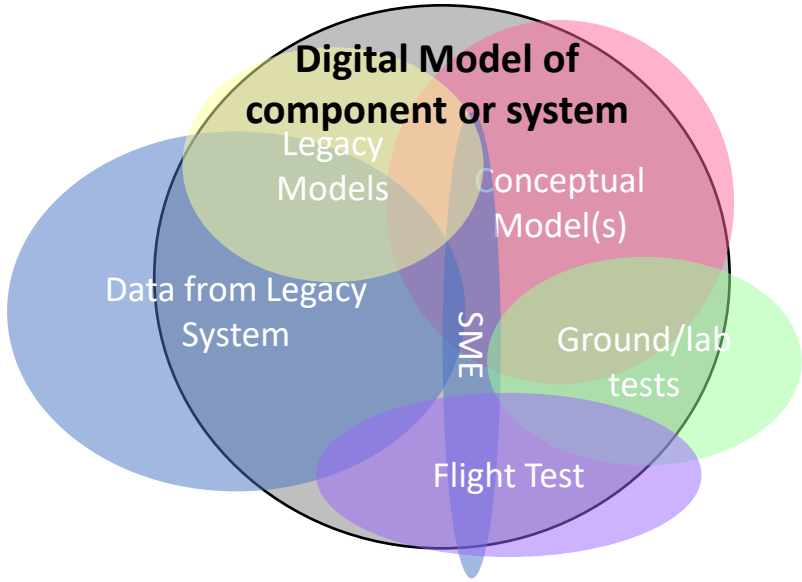
Fidelity quantifies model similarity to referent in 4 dimensions of **detail**

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

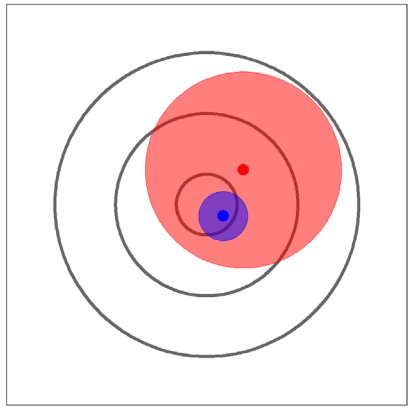
Fidelity

Model

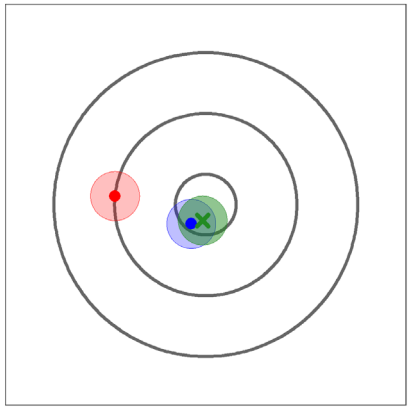
Referent

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

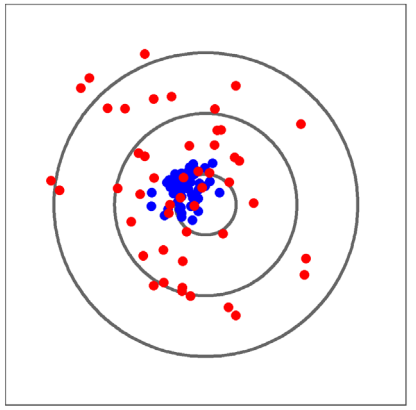
Resolution



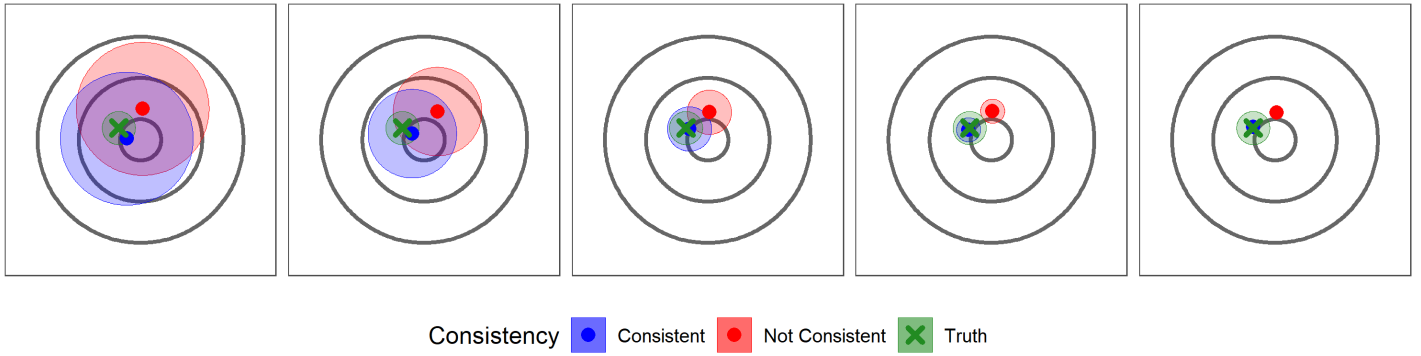
Accuracy



Repeatability



Consistency



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

Image credits: STAT COE

Referent Authority

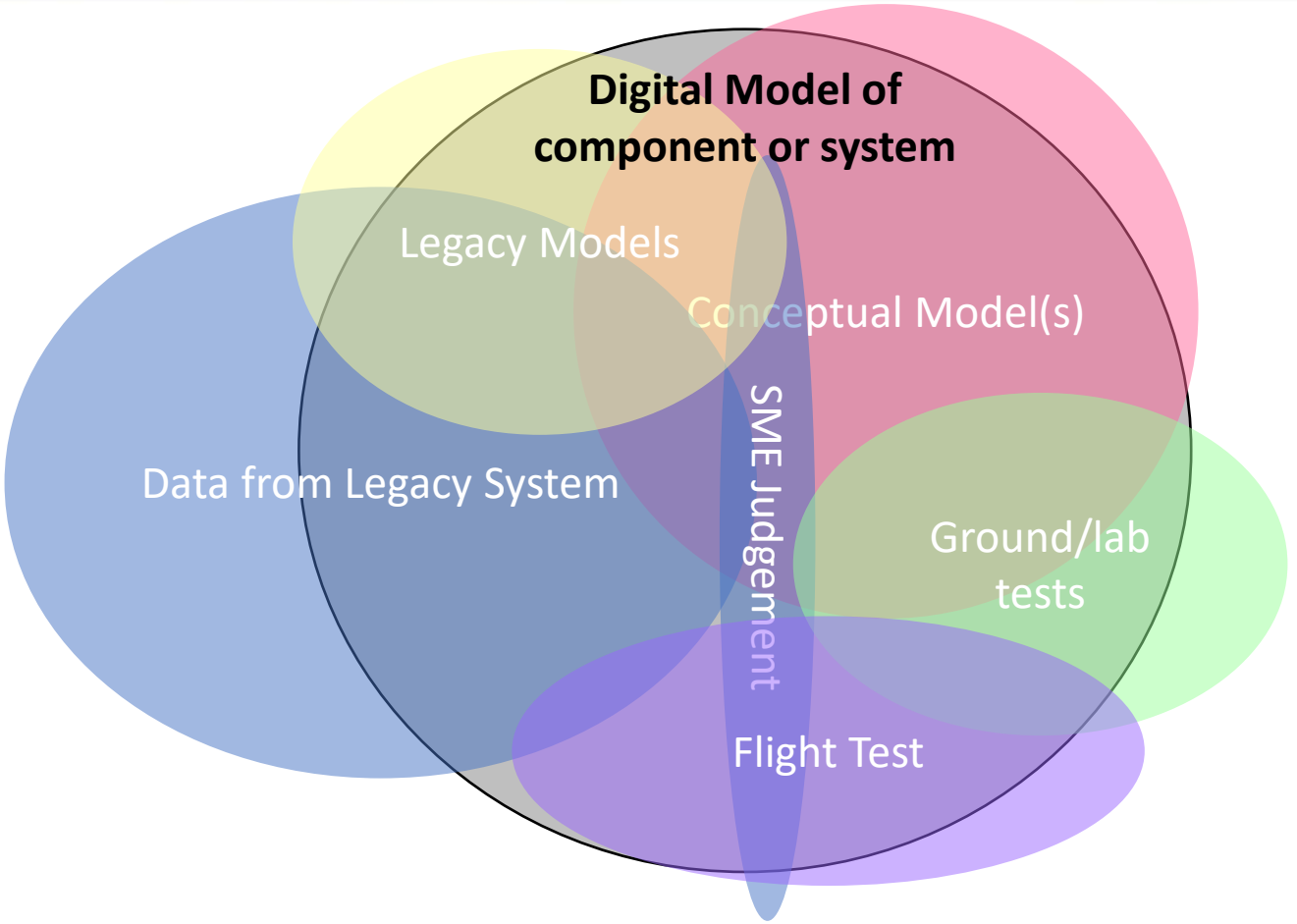
- Referent authority level is similar to TRLs:
  - To declare a technology at a TRL, some evidence must exist to show that the technology has been demonstrated at that TRL
  - That evidence is a referent trusted to that Readiness Level
  - A referent can transfer trust to a model up to its own level 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

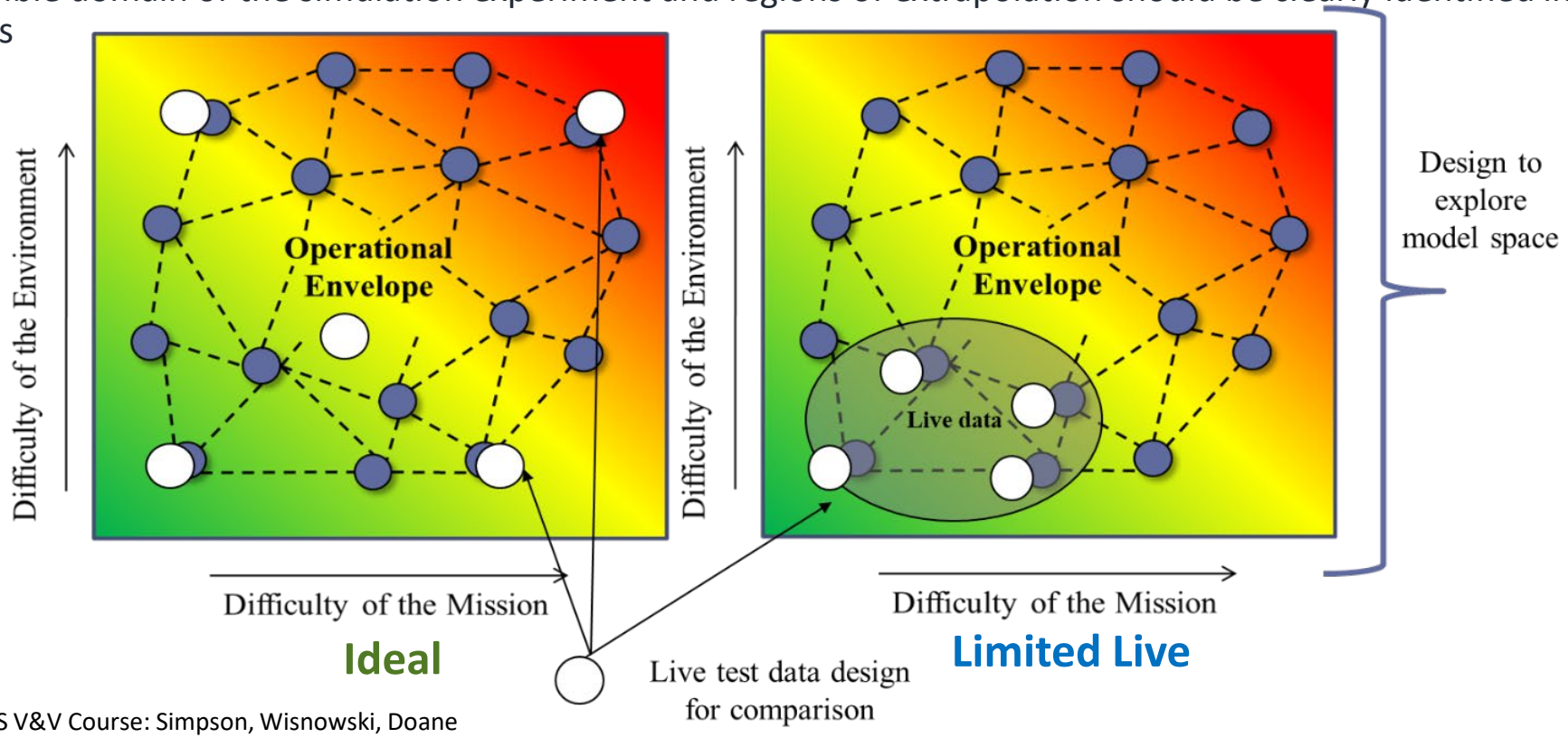


Scope quantifies degree to which model and referent represent the **same system**

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

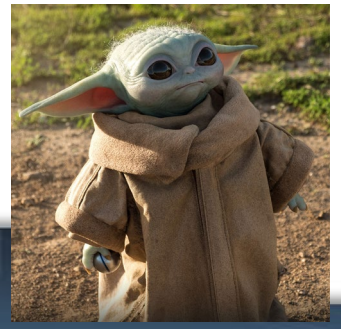


Slides from STAT in M&S V&V Course: Simpson, Wisnowski, Doane

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

This is the way



Questions?

QUESTIONS AND DISCUSSION

SAVE THE DATE: NOV 2-4, 2021

SERC SPONSOR RESEARCH REVIEW

SERC SPONSOR RESEARCH REVIEW
Day 1
IN-PERSON | VIRTUAL

NATIONAL PRESS CLUB WASHINGTON DC

INTRODUCING THE AIRC

LIVE DEMONSTRATIONS

NETWORKING

SERC SPONSOR RESEARCH REVIEW
Day 2
VIRTUAL

UPDATES ON SERC RESEARCH

DIGITAL ENGINEERING

TRUSTED SYSTEMS

ARTIFICAL INTELLIGENCE

HUMAN CAPITAL DEVELOPMENT

SERC DOCTORAL STUDENT FORUM
Day 3
VIRTUAL

STUDENT RESEARCH

FREE TUTORIALS

VIRTUAL NETWORKING

SERC TALKS

"Test and Evaluation" Series

UPCOMING TALKS

Series Moderator: Dr. Laura Freeman, SERC Research Council Member, Director of the Hume Center's Intelligent Systems Lab and Assistant Dean for Research in the College of Science, and Research Associate Professor in the Department of Statistics at Virginia Tech

Tentative Dates – Wednesdays at 1 PM ET: October 27 | December 1, 2021

CONTACT

Webinar Coordinator: Ms. Mimi Marcus, Stevens Institute of Technology – mmarcus@stevens.edu

Please visit the SERC Talks page to register and for more information and updates.

SYSTEMS ENGINEERING RESEARCH CENTER

SERC TALKS

THANK YOU FOR JOINING US!

Please check back on the SERC website for today's recording and future SERC Talks information.

Subscribe and follow SERC on our social channels

SYSTEMS ENGINEERING RESEARCH CENTER

CONTACT US:

serc@sercuarc.org

www.sercuarc.org/contact-us/