Best Practices for Irregular Warfare (IW) Data Quality Control

Jeff Appleget & Fred Cameron

29th ISMOR

August 2012

Agenda

- Irregular Warfare (IW)
- Background
  - Physics-Based Combat Modeling
  - IW Modeling Validation Best Practices
- IW Data Quality Control Research
- IW Data Challenges
- IW Data QC Best Practice Recommendations
- Models, Complexity and Error:
  Implications for Modeling Irregular Warfare
- Conclusion

Irregular Warfare (IW)
US Department of Defense

References

- Irregular Warfare (IW) Model Validation Best Practices Guide (TRAC, 11 Nov 2011)
- Irregular Warfare (IW) Data Quality Best Practices Guide (TRAC, 31 Dec 2011)
- DoD Directive 3000.07 Irregular Warfare (DEC 2008)
- DoD Instruction 5000.61 DoD Modeling and Simulation (M&S) Verification, Validation, and Accreditation (VV&A) (DEC 2009)
- Joint Pub 3-0 w/Change 1 (FEB 2008)
- IW Joint Operating Concept Version 2.0 (MAY 2010)
- FM 3-24/MCWP 3-33.5 Counterinsurgency (DEC 2006)

IW Definition

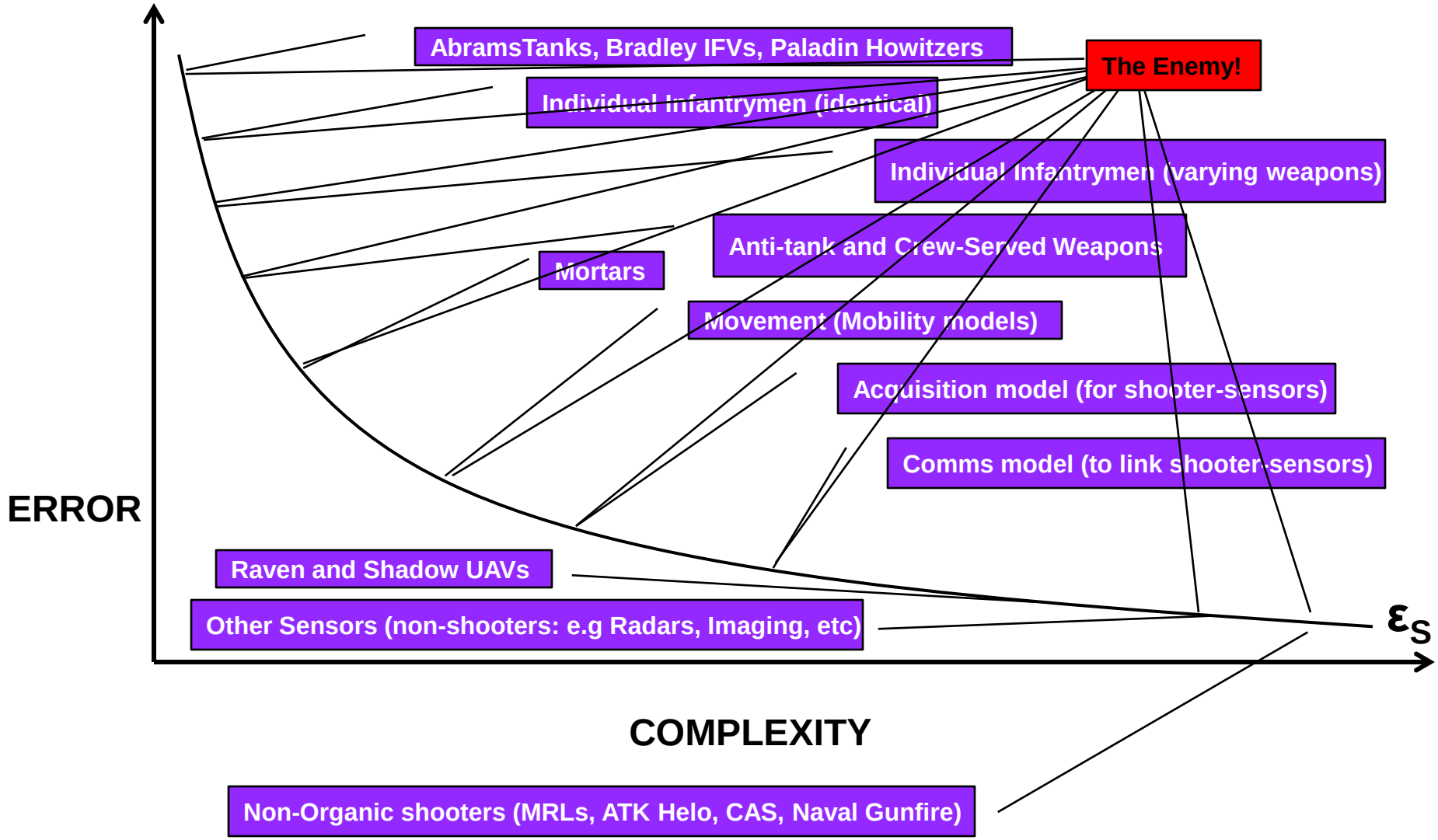
Irregular warfare. A violent struggle among state and non-state actors for legitimacy and influence over the relevant populations. Irregular warfare favors indirect and asymmetric approaches, though it may employ the full range of military and other capabilities, in order to erode an adversary's power, influence, and will. (JP 1-02)

The focus of IW is the relevant populations, not the enemy's military capability.

Background:
Physics-Based Combat Modeling

Models, Complexity and Error

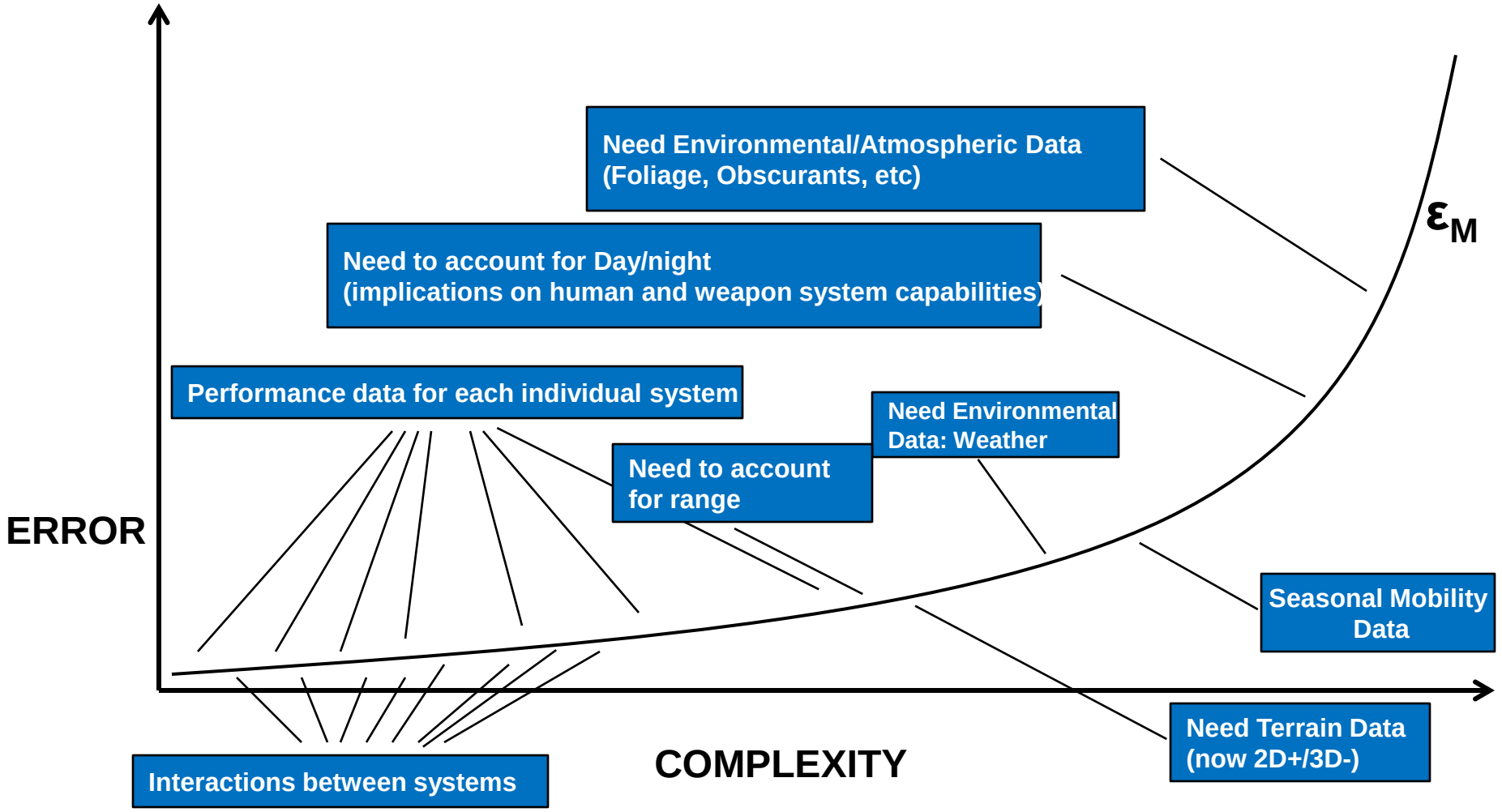
Example: Entity-level Modeling of a US Heavy Brigade Combat Team (BCT)



The error of specification, ε_S, decreases as more of the systems in the BCT are represented.

Models, Complexity and Error

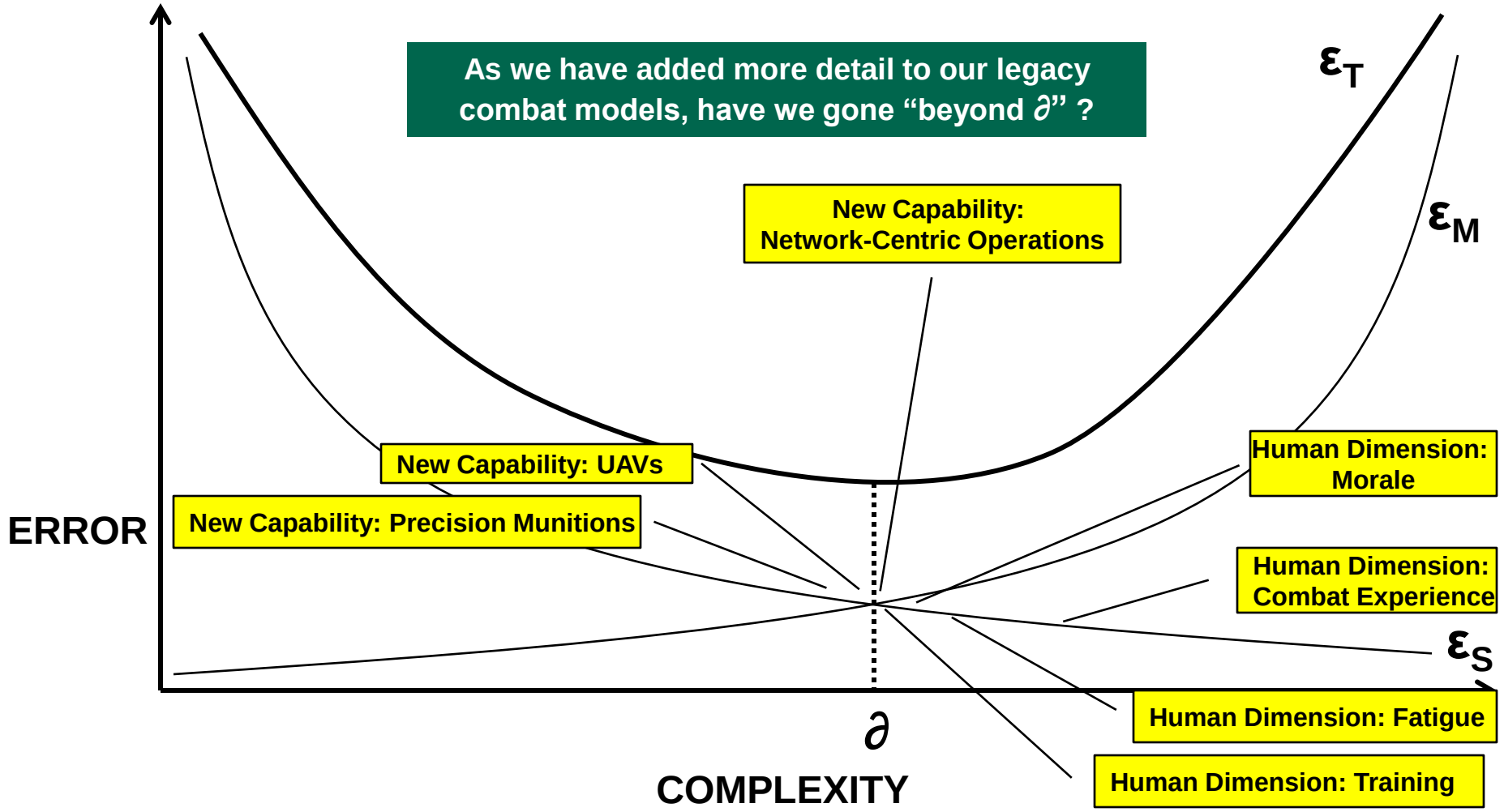
Example: Entity-level Modeling of a US Heavy Brigade Combat Team (BCT)



The error of measurement, ε_M, increases as each system added requires representational data and interactional data with many of the other systems (friendly and enemy).

Models, Complexity and Error

Example: Entity-level Modeling of a US Heavy Brigade Combat Team (BCT)



Total error: There exists a point ∂ where, beyond which, adding more detail to your model actually *increases* the overall error of the model.

Fred Cameron suggested using the following paper that uses this rubric to describe economics, energy and environmental factors
Leinweber, David. "Models, Complexity, and Error," A Rand Note prepared for the Department of Energy, N-1204-DOE, June 1979.

Physics-Based Cold War Legacy
Ground Models

- **Attrition:**
  - **Strategic/Theater-level Models: Processes incorporating modified Lanchester approaches for attrition.**
  - **Tactical level models: Entity based models using variants of the ACQUIRE algorithm and performance data generated from engineering level models (SSPK, P(Hit), P(Kill/Hit), …) for attrition.**
- **Purposes:**
  - **Force Structure**
  - **Force Design**
  - **Acquisition**
  - **Operational Planning & Assessments**
  - **Training**

***Assertion*: As we got into the next-generation Physics-Based combat models, we started with existing attrition modeling as the foundation.**

Background:
Irregular Warfare Modeling Validation Best Practices

Physics-based combat modeling vs. IW combat modeling

Physics-based Modeling

Representation:	Small combat unit force-on-force lethal engagements.
Conceptual Model:	Describes the interactions that must be accounted for when two entities (e.g. a red and a blue tank) exchange fire.
Referent:	Laws of physics that represent target searching, target acquisition, and engagement of targets, accounting for lines of sight, weapons ballistics, and assessing damage.

Referent is implicit in force-on-force combat modeling and adequate for underpinning models. It comes from the laws that we use to represent combat.

IW Modeling

Representation:	Specific multi-layered conflict ecosystem, to include interaction between population and combat actors.
Conceptual Model:	Describes the interaction (kinetic and non-kinetic) of actors (e.g. insurgents and counter-insurgent forces) with each other and civilian populace.
Referent:	Social science theories that account for human behavior interaction, laws of physics representing combat.

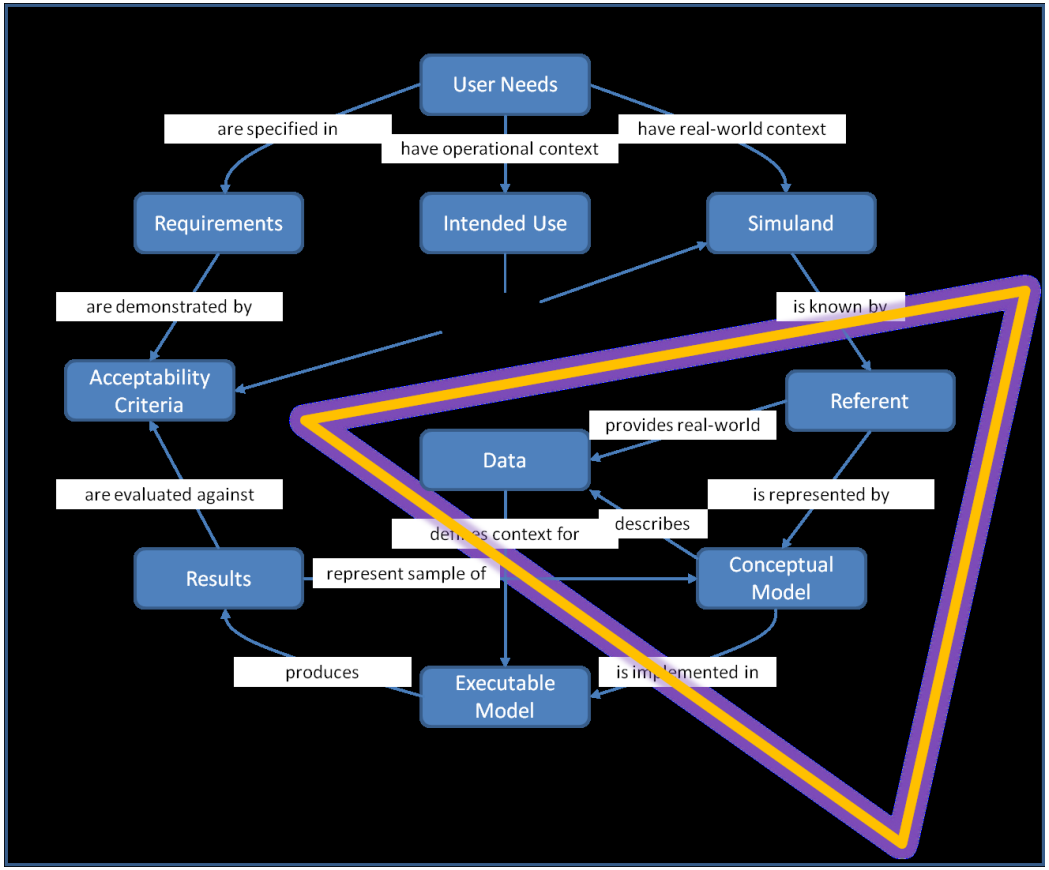
Referent must be explicitly defined, accounting for how the actors will interact within the modeling environment. A far less familiar modeling domain.

The referent for our force-on-force combat models has been the laws of physics—social science model referents are typically <u>theoretical</u>.

Validation Framework Concept Map
"The Validation Triangle"

Having developers provide a detailed <u>conceptual model</u>, a <u>referent</u> that describes each social science theory that will be modeled (including alternate theories and why the candidate theory was chosen), and a description of the <u>data</u> that the model requires, and the source(s) of the data will be vital to producing a model that can be validated.

- **User Needs**
  - The developer needs to obtain a succinct and clear statement of the problem the M&S is expected to address.
- **Requirements**
  - Develop specific functional or quality statements that can be directly and explicitly assessed to determine requirement.
- **Acceptability Criteria**
  - Develop a requirements traceability matrix relating each specified requirement with acceptability criteria applicable to the intended use.
- **Results**
  - The acceptability criteria identify what the model needs to do to satisfy or meet the set of respective requirements pertinent to the intended use.
- **Data**
  - The greater the specificity in the data requirements for a model, the greater the ability to collect the data needed to populate the model.



<u>Modeling Best Practice</u>: Validation starts before the first line of code is written!

- **Intended Use**
  - Obtain a clear, succinct statement of intended use from the user representatives.
- **Simuland**
  - The simuland is the real-world system of interest, including the objects, processes, or phenomena to be simulated.
- **Referent**
  - identify the social science theory (or theories, if multiple competing theories will be represented in the model for comparison) that explains that phenomena.
- **Conceptual Model**
  - Develop the conceptual model using tools and techniques that create machine-readable specifications of the data and logic of the model.
- **Executable Model**
  - Design the model implementation to be as transparent as possible to permit analysis of execution paths and computed outcomes.

Irregular Warfare Data Quality Control (versus ‘data validation’)

Problem Statement & Research Team
(Work sponsored by OSD-CAPE & JDS through the IW-SCG)

- **In FY10, as follow on work to the IW Model Validation Best Practices Guide (TRAC, 11 Nov 2011), JDS asked us to delve into "IW Data Validation."**
- **Task: "TRADOC Analysis Center (TRAC) will provide a report that assesses, at a minimum, the validation of IW data, to include an examination of data requirements, data sources, and data availability as well as derivation of data."**
- **Team:**
  - **MAJ Ricky Brown and MAJ Joe Vargas, TRAC-Monterey;**
  - **Dr. Jeff Appleget, Mr. Curt Blais, Dr. Mike Jaye, NPS;**
  - **Dr. Eric Weisel, Weisel Science & Technology Corporation.**
- **Reviewers:**
  - **Mr. Howard Body and Dr. George Rose, [dstl]**
  - **Mr. Fred Cameron, CORA**
  - **Ms. Robin Griffen, Complex Operations Data Development Activity, TRAC-FLVN**
  - **Dr. Dean Hartley, Hartley Consulting**
  - **Mr. Don Hodge, AMSAA**
  - **Mr. Steve Stephens, MCCDC**
  - **Mr. Ed Weinberg, OSD-CAPE (contractor)**

OSD-CAPE: Office of the Secretary of Defense - Cost Assessment and Program Evaluation
JDS: Joint Data Support
IW-SCG: Irregular Warfare Senior Coordinating Group
[dstl]: Defence Science and Technology Laboratory, Ministry of Defence (MoD), United Kingdom
CORA: Centre for Operational Research and Analysis, Department of National Defence (DND), Canada
AMSAA: Army Materiel Systems Analysis Activity
MCCDC: Marine Corps Combat Development Command

Data Quality Control Not Well Defined

- **DoD policy[1] references data indirectly**
  - **Models, simulations, and associated data used to support DoD processes, products, and decisions shall undergo verification and validation (V&V) throughout their lifecycles.**
  - **Models, simulations, and associated data used to support DoD processes, products, and decisions shall be accredited for an intended use.**
- **A review of the available literature finds discussions of verification and validation to be focused almost exclusively on models and simulations. For the few papers where verification or validation of data is discussed, it is almost exclusively focused on numerical data.**
- **The Army organization with the mission to provide systems performance data to M&S users, uses the term certification and not verification, validation, nor accreditation.**

During the conduct of our research, it was not apparent that DoD organizations understood verification, validation, and accreditation to be distinctly different and separable processes that were to be applied to data.

1. DODI 5000.61, 9 December 2009

IW Data Challenges

IW Versus Physics-Based Data

Representational Challenge

In physics-based models, the underlying assumptions about how things work are well-known and widely accepted.

For most US Army and USMC physics-based models, AMSAA provides performance data that has undergone a QC process called "certification."

In the parts of IW models that represent the civilian population:

- **Many theories on individual and group behaviors exist**
- **[illegible]**
- **[illegible]**

What is more scientifically rigorous: a theory or a hypothesis?

  - **Many IW modeling development teams list no social scientists as team members or even consultants.**
  - **Many IW modeling proposals do not cite any relevant social science theories or models to explain the foundation of their modeling concepts.**
- **Simple Aggregation techniques do not apply to many social science disciplines**
  - **Complicated versus Complex systems.**
  - **Micro versus macro economics.**
  - **Individual versus group behavior.**

On Theories and Hypotheses…

- **A hypothesis is an educated guess, based on observation. Usually, a hypothesis can be supported or refuted through experimentation or more observation. A hypothesis can be disproven, but not proven to be true.**
- **A scientific theory summarizes a hypothesis or group of hypotheses that have been *supported with repeated testing.* A theory is valid as long as there is no evidence to dispute it. Therefore, theories can be disproven. Basically, if evidence accumulates to support a hypothesis, then the hypothesis can become accepted as a good explanation of a phenomenon. One definition of a theory is to say it's an accepted hypothesis.**

"Repeated testing" implies to me that there should be a record of that testing. Could that record be called…data!?!

By Anne Marie Helmenstine, Ph.D., About.com Guide

Additional IW Data Challenges…

- **Data Types:**
  - **Intangible**
  - **Transient**
  - **Non-numerical**
- **Data Sources:**
  - **Non-DoD**
  - **Non-governmental**
  - **Dependent on Subject Matter Experts**
  - **‘Pay to play’**
- **Data Responsibilities:**
  - **Responsibility for IW data has not been assigned**

Initial Consent Matrix*

		Faction				
		Militia A	Militia C	Militia E	Govt	UNSFOR
Ethnic Group	Ethnic A					
	Ethnic B					
	Ethnic C					
	Ethnic D					
	Ethnic E					

How is data developed for this? How do these data change over time? What are the threshold values to transition between states? Can you jump from "Supports" to "Tensions Exist" without ever being neutral?

- Supports
- Neutral towards
- Tensions exist

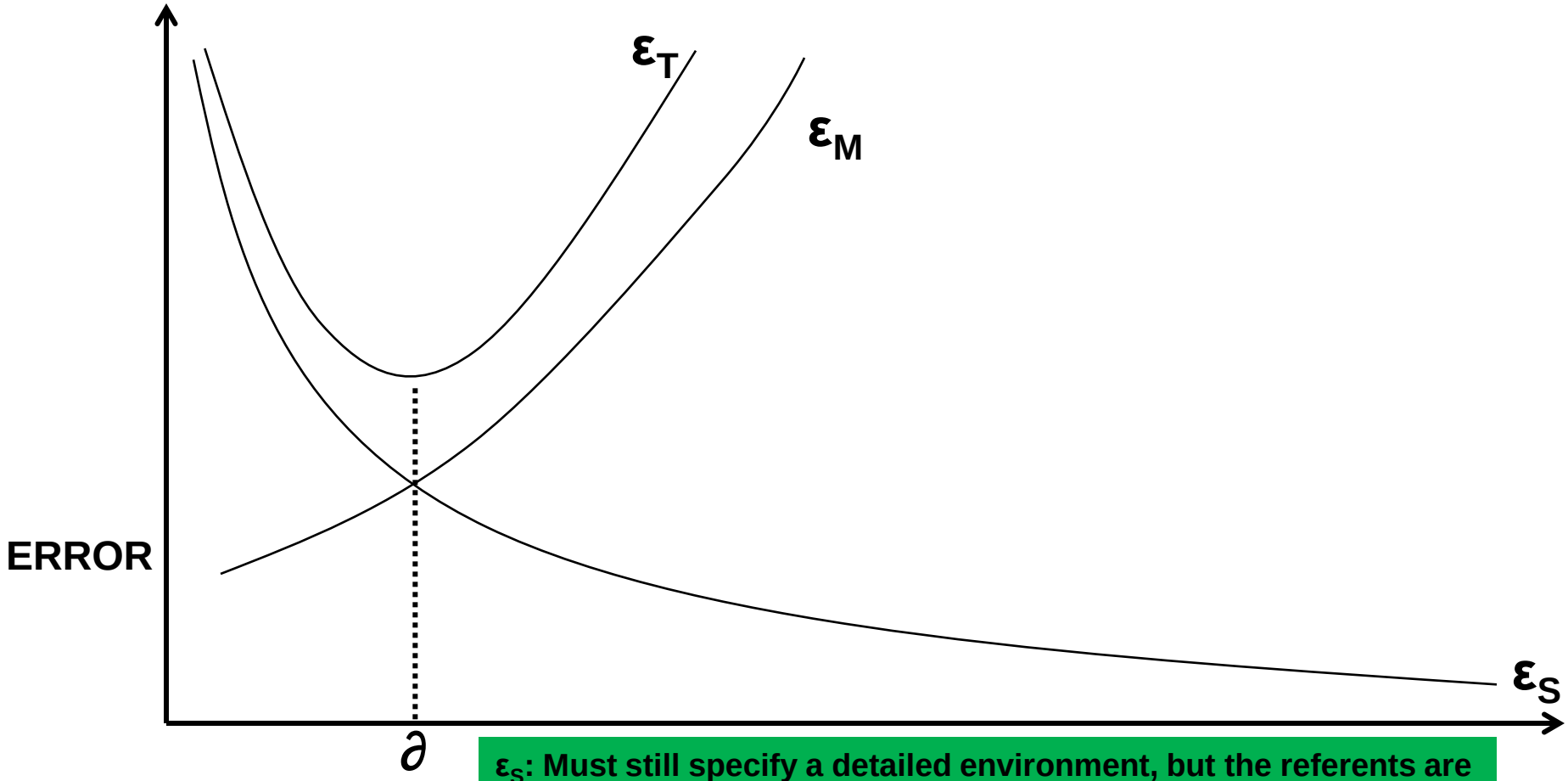
** Example from [dstl], PSOM Yellowstone Scenario, but there are similar matrices in other IW models.*

Best Practice Recommendations

IW Data Quality Control Best Practices

Best Practice	***Rationale for Best Practice***
• **DoD:**	
– designates a single organization to serve as the IW Data Clearinghouse.	– Provides a focal point for data reuse.
– directs a rewrite of DoDI 5000.61 to clarify data management practices.	– Codifies how DoD will manage IW data.
• **DoD organizations using IW Methods, Models, and Tools (MMTs):**	
– certify the data used in the MMT as being "fit for purpose."	– Puts the onus on the study directors to ensure the data is good enough.
– document data sources, data development methodology, and data risk assessment.	– Provides other DoD IW data users insight into what's available, what it was used for, and the user's quality assessment.
• **DoD IW Data Clearinghouse:**	
– specifies a standard metadata set for describing data supporting IW MMT.	– Provides DoD IW data users a standard format to document IW data sources.
– maintains a repository of metadata for IW data sources used in DoD.	– Provides DoD IW data users a place to "shop" for IW data.
– specifies and enforces data quality and data risk assessment entries for IW data sources catalogued in the IW metadata database.	– Allows DoD IW data users an understanding of the data's quality.
– coordinates the procurement of IW data from sources requiring formal agreements, usage restrictions, additional certifications, and/or fees for usage.	– Provides a focal point for the procurement and management of IW data.

Models, Complexity and Error: Implications for Modeling Irregular Warfare



ε_S: Must still specify a detailed environment, but the referents are social science theories instead of physics laws.

ε_M: Because we're dealing with human interactions, the data will not have nearly the same fidelity as our physics-based models.

ε_T: Any modeling of Irregular Warfare, COIN, Stability Ops, Peace Support ops needs to be simple for it to be useful.

Leinweber, David. "Models, Complexity, and Error," A Rand Note prepared for the Department of Energy, N-1204-DOE, June 1979.

Conclusion

- **IW Data Quality Control big challenge.**
- **DoD has a big role to play.**
- **Best Practices good start, but will morph as IW models mature.**
- **Users of IW data must accept responsibility for the data they use.**
- **There will be no equivalent AMSAA that provides the DoD IW data, at least not for the non-kinetic data requirements.**