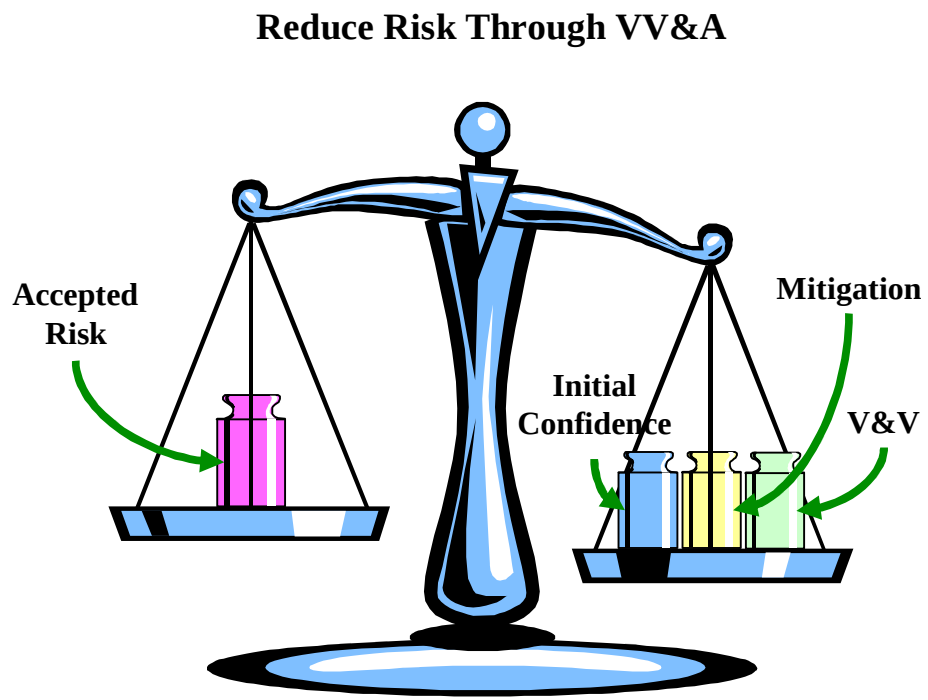
VV&A of DIME/PMESII Models



Dr. Dean S. Hartley III

What Is DIME/PMESII?

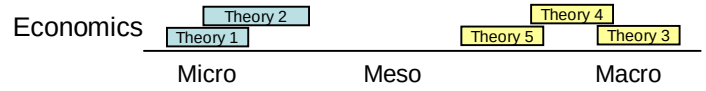
- "PMESII" represents the state of the situation and "DIME" represents the actions that a nation can take to change that state
- PMESII is actually a decomposition of the entire state into
  - Political state variables
  - Military state variables
  - Economic state variables
  - Social state variables
  - Information state variables
  - Infrastructure state variables
- DIME is a decomposition of national elements of power into
  - Diplomatic actions
  - Information actions
  - Military actions
  - Economic actions
- Decomposition methodologies
  - There are other possible decompositions than DIME/PMESII
  - The proper choice depends on what is useful and on the nature of the world
  - For general purpose discussions, DIME/PMESII is adequate to convey the division between state variables and actions and a rough concept of decomposition

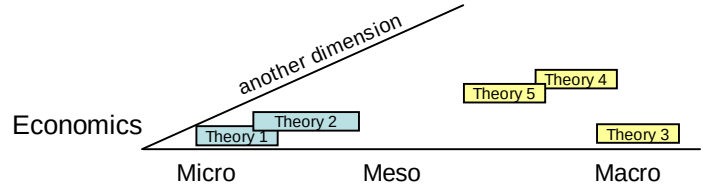
Why DIME/PMESII Models?

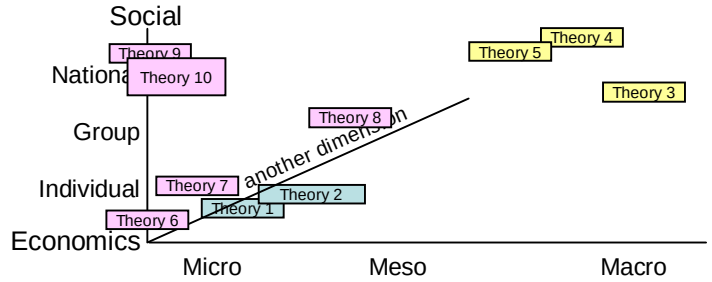
- Situations for which the DIME/PMESII concepts are relevant typically involve bad or poor situations at a national or global level
- Doing something (including the “do nothing” option) involves real risks to real people
- The decision of what to do almost always involves the use of models (though usually these have been mental models), in which some person (or group) guesses what will happen for each of several choices
- The DIME/PMESII paradigm supports a more detailed examination of options and results
- The complexity of the possibilities suggests advantages for codifying the model in a computer model

What's the Problem?

- We don't know very much about how the world really works
- Viewing Economics as a single dimension, there are several theories that cover the Micro level and several that cover the Macro level, but fewer in the Meso area. Some of the theories disagree.
- Economics actually has several dimensions. This resolves some apparent conflicts; however it exposes more gaps.
- Social theory has some theories that relate to economics and some that don't (shown on the "Social" axis). More gaps and more contradictions are exposed.
- The full DIME/PMESII space would involve an unknown number of dimensions and would show large gaps and multiple contradictions.
- How do you define a good DIME/PMESII model?







What's the Solution?

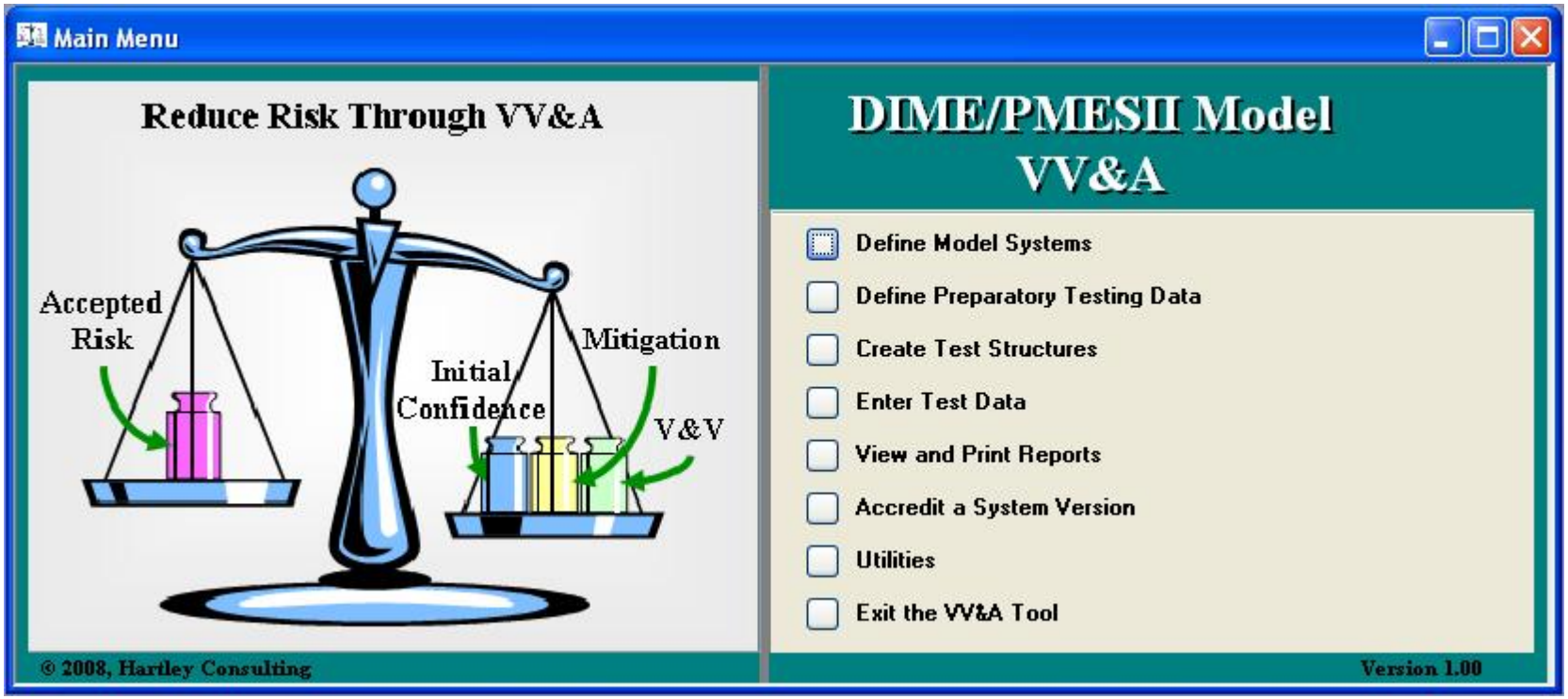
- People will continue to build DIME/PMESII models (both mental and computer models) because the need is real
- We need to identify what we don't know
  - What gaps exist
  - What contradictions exist in our theories
- Until we know more about the real world, we need to assess the models
- We need to understand what questions we are asking of the models
- We need to understand the risks involved (with a fair amount of detail)

Central VV&A Concepts

- The purpose of V&V is to generate information to permit the best uses of a model
- The verification process generates information to support improvements to the model and bug fixes
- The validation process generates information to support
  - improvements to the model
  - mitigation of risks
    - modified model operation techniques
    - modified model output interpretations
    - restrictions on types of use
    - supplemental model and manual analysis needs
- Accreditation is an official determination to use a model
  - may be as simple as a “go” for a study
  - may include caveats based on incomplete bug fixes and validation mitigations

The Prototype

- We have a prototype VV&A support tool
- Provides
  - Disciplined approach
  - Embodies current best thinking on how process should work
  - Ease of use
  - Minimum data entry and optional increased information entry, where desired
- Will be modified as experience directs



Standard VV&A:
Commercial & Kinetic Military Models

- Best reference for processes: ***Simulation Validation***, Peter L. Knepell & Deborah C. Arangno, IEEE Computer Society Press, 1993.
- Organization, discipline, & attention to details are key concepts for successful VV&A
- We will assume general familiarity with the following concepts

Standard Test Types*

- Informal
  - Audit
  - Desk checking
  - Documentation checking
  - Face validation
  - Inspections
  - Reviews
  - Turing test
  - walkthroughs
- Static
  - Cause-effect graphing
  - Control analysis
    - Calling structure analysis
    - Concurrent process analysis
    - Control flow analysis
    - State transition analysis
  - Data analysis
    - Data dependency analysis
    - Data flow analysis
  - Fault/failure analysis
  - Interface analysis
    - Model interface analysis
    - User interface analysis
  - Semantic analysis
  - Structural analysis
  - Symbolic evaluation
  - Syntax analysis
  - Traceability assessment
- Dynamic
  - Acceptance testing
  - Alpha testing
  - Assertion testing
  - Beta testing
  - Bottom-up testing
  - Comparison testing
  - Compliance testing
    - Authorization testing
    - Performance testing
    - Security testing
    - Standards testing
  - Debugging
  - Execution testing
    - Execution monitoring
    - Execution profiling
    - Execution tracing
  - Fault/failure insertion testing
  - Field testing
  - Functional (black box) testing
  - Graphical comparisons
  - Interface testing
    - Data interface testing
    - Model interface testing
    - User interface testing
  - Object-flow testing
- Dynamic (continued)
  - Partition testing
  - Predictive validation
  - Product testing
  - Regression testing
  - Sensitivity testing
  - Special input testing
    - Boundary value testing
    - Equivalence partioning testing
    - Extreme input testing
    - Invalid input testing
    - Real-time input testing
    - Self-driven input testing
    - Stress testing
    - Trace-driven input testing
  - Statistical techniques
  - Structural (white box) testing
    - Branch testing
    - Condition testing
    - Data flow testing
    - Loop testing
    - Path testing
    - Statement testing
  - Submodel/module testing
  - Symbolic debugging
  - Top-down testing
  - Visualisation/animation
- Formal
  - Induction
  - Inductive assertions
  - Inference
  - Lambda calculus
  - Logical deduction
  - Predicate calculus
  - Predicate transformation
  - Proof of correctness

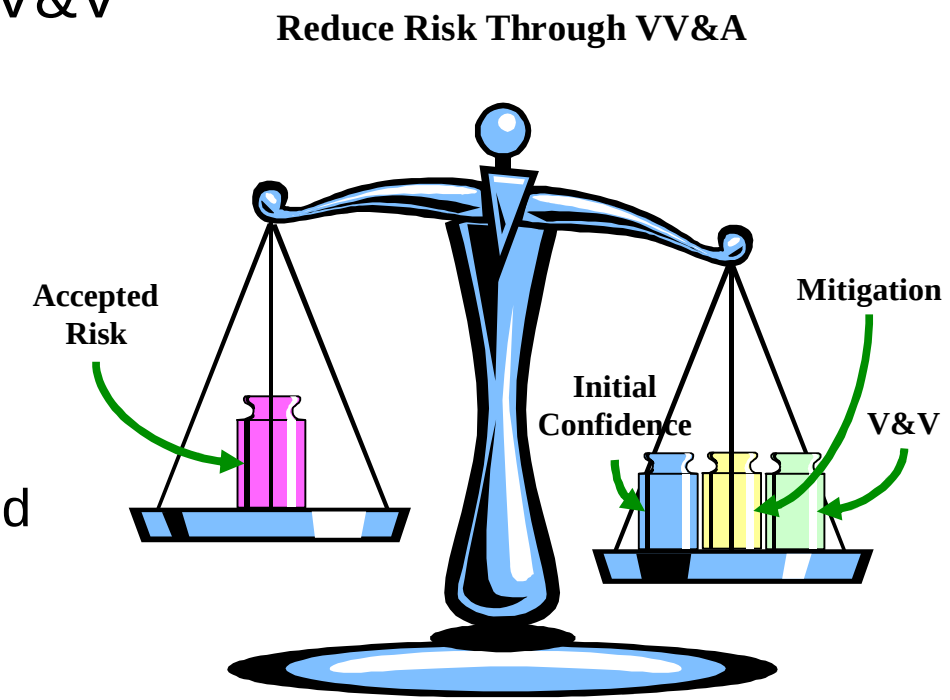
* Encyclopedia of Operations Research and Management Science

Standard DMSO-Recommended V&V Plan Elements

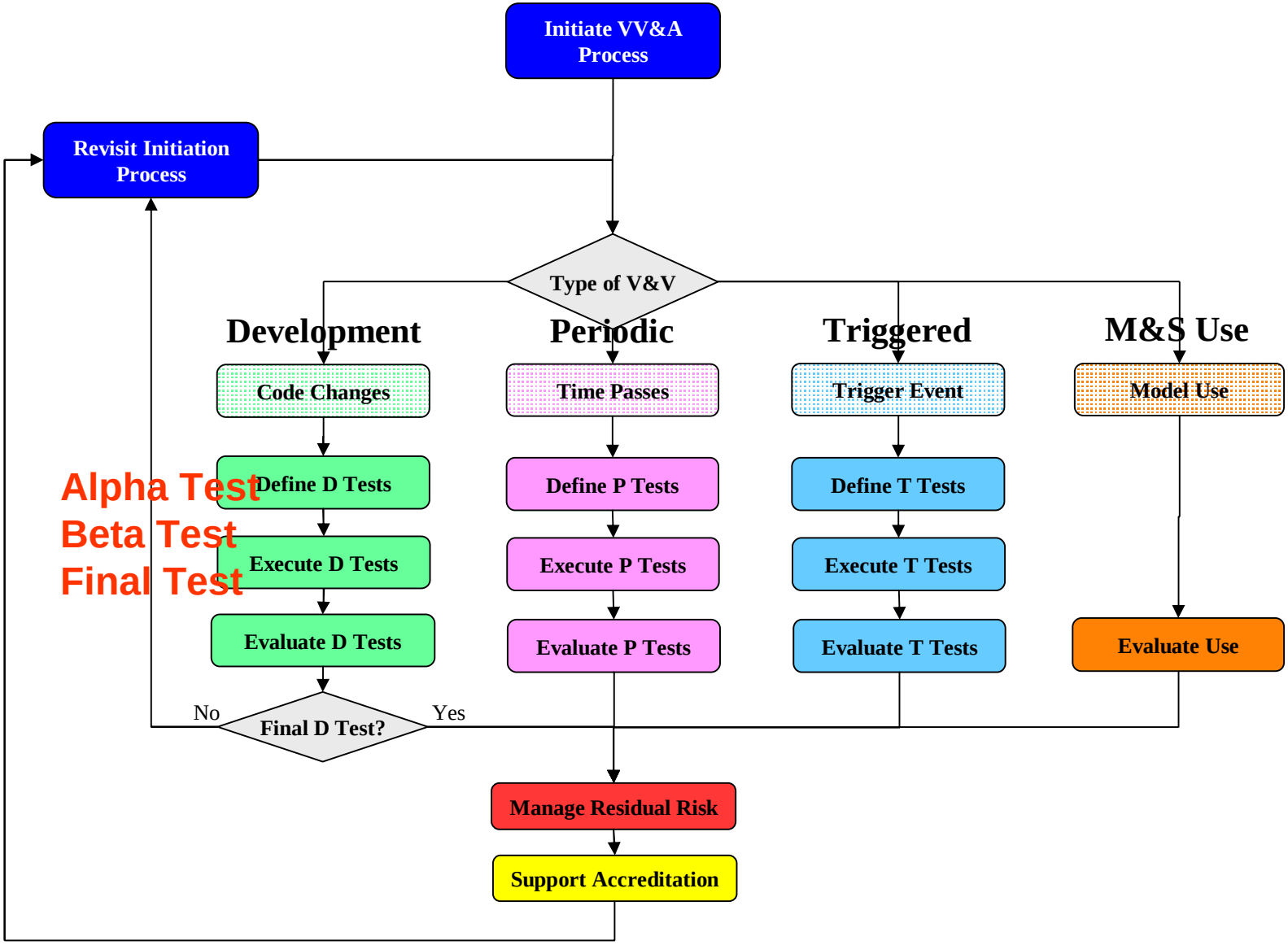
- V&V PLAN EXECUTIVE SUMMARY
- PROBLEM STATEMENT
  - Intended Use
  - M&S Overview
  - M&S Application
  - Accreditation Scope
  - V&V Scope
- M&S REQUIREMENTS AND ACCEPTABILITY CRITERIA
- M&S ASSUMPTIONS, CAPABILITIES, LIMITATIONS, & RISKS/IMPACTS
  - M&S Assumptions
  - M&S Capabilities
  - M&S Limitations
  - M&S Risks/Impacts
- V&V METHODOLOGY
  - Planned Data V&V Tasks/Activities
  - Planned Conceptual Model Validation Tasks/Activities
  - Planned Design Verification Tasks/Activities
  - Planned Implementation Verification Tasks/Activities
  - Planned Results Validation Tasks/Activities
  - Planned V&V Reporting Tasks/Activities
- V&V ISSUES
- KEY PARTICIPANTS
  - Accreditation Participants
  - V&V Participants
  - Other Participants
- PLANNED V&V RESOURCES
  - Planned V&V Tasking and Funding
  - Planned V&V Timeline
- APPENDIX A: M&S DESCRIPTION
  - M&S Overview
  - M&S Development and Structure
  - M&S Capabilities and Limitations
  - M&S Use History
  - Data
  - Configuration Management
- APPENDIX B: M&S REQUIREMENTS TRACEABILITY MATRIX
- APPENDIX C: BASIS OF COMPARISON
- APPENDIX D: REFERENCES
- APPENDIX E: ACRONYMS
- APPENDIX F: GLOSSARY
- APPENDIX G: V&V PROGRAMMATICS
- APPENDIX H: DISTRIBUTION LIST
- APPENDIX I: ACCREDITATION PLAN

Risk Based VV&A Philosophy: Extension of Standard Usage

- Identify problems through V&V
  - correct the problems or
  - mitigate the problems
- Identify residual risks
- Mitigate the risks
- Decide on accreditation
  - accept the remaining risk and accredit
  - do not accredit
  - (accreditation need not be for the entire original use)



Entrenched VV&A Practice: Extension



Baseline VV&A Process* = Standard + Risk-Based + Entrenched

- Initiate the VV&A Process
  - Define and assemble the V&V Team(s)
  - Characterize the model
  - Perform stakeholder assessment: roles, responsibilities, relationships
  - Define the Documentation Standards
  - Define the Referents for V&V
  - Define the General Risks
  - Define the Metrics
  - Define the V&V Criteria
- Perform actions appropriate to type of V&V over complete model life-cycle
- Manage Risk
- Support Accreditation

*Note: The baseline VV&A process is designed to address a problem set for which adequate time and resources are available

Compressed & Hyper-Compressed VV&A: New Concepts

- In Compressed VV&A, the time available for changing the model or responding to a changed situation is measured in weeks (including time for VV&A)

- In Hyper-Compressed VV&A, the time available for changing the model or responding to a changed situation is measured in hours or days (including time for VV&A)

- Three techniques are offered that can support this time compression

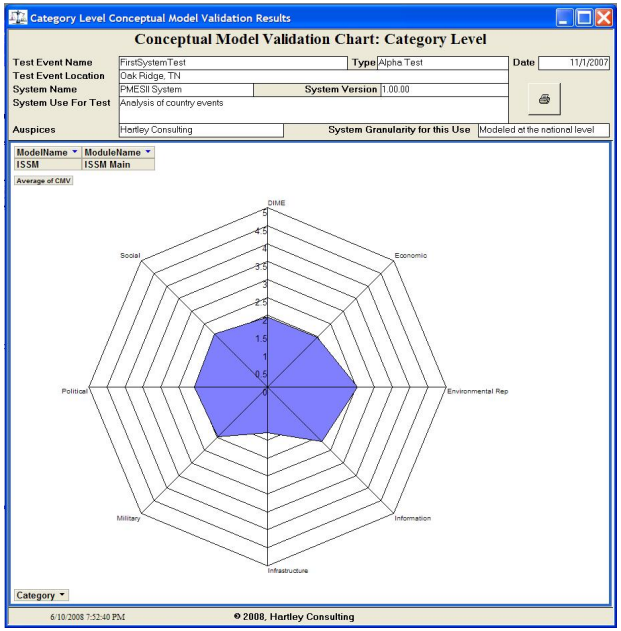
DIME/PMESII VV&A

- Verification standards are the same as for any other model
- Achievable validation levels will be lower than for physics-based models
- That means valid uses will be more restricted

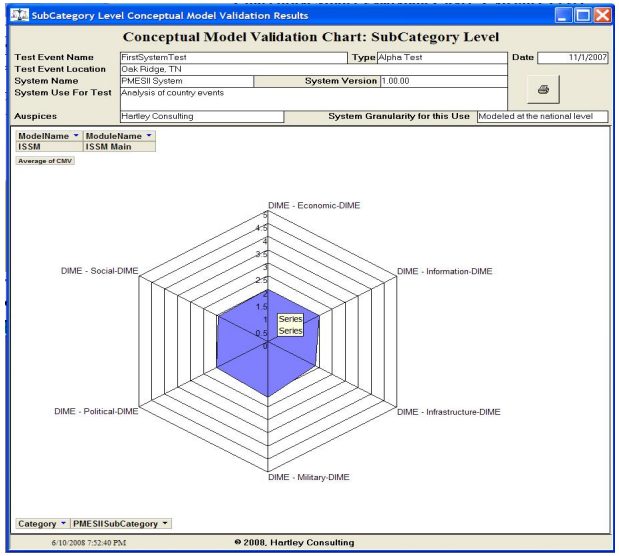
- However, don’t stint on validation because a particular use doesn’t require it
  - Most models are ultimately pushed to new uses over time
  - Think of the validation process as gathering information on what the model does and doesn’t do

- Record everything

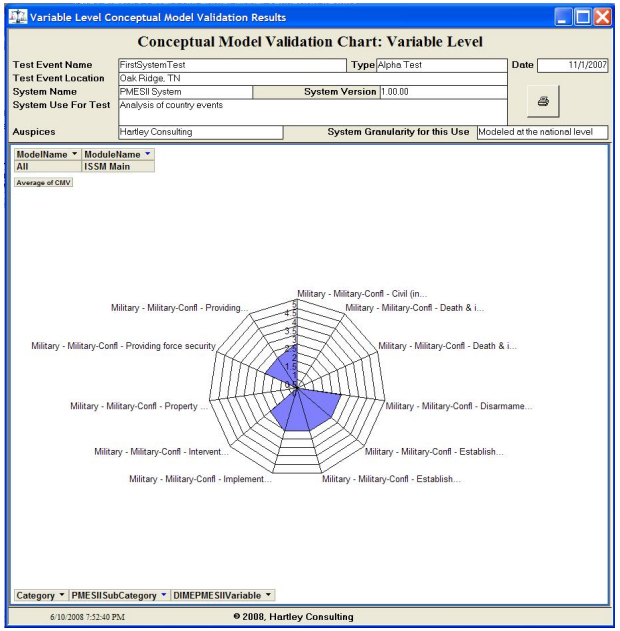
Talking to the Sponsor, e.g. Conceptual Model Validation Charts



Category Level Tests



SubCategory Level Tests



Variable Level Tests

Where are the gaps in coverage of the Conceptual Model?
What is the maximum level of validity that can be expected?

Conclusion

- V&V are difficult processes
- DIME/PMESII complexity adds to the difficulty

- However, with proper support
  - V&V is possible
  - The V&V supports Accreditation

- Proper support is now available
  - The DIME/PMESII Model VV&A Tool
  - http://home.comcast.net/~dshartley3/VVATool/VVA.htm

- A tremendous amount of detail has been omitted from this presentation!

From the hills of East Tennessee

Dr. Dean S. Hartley III

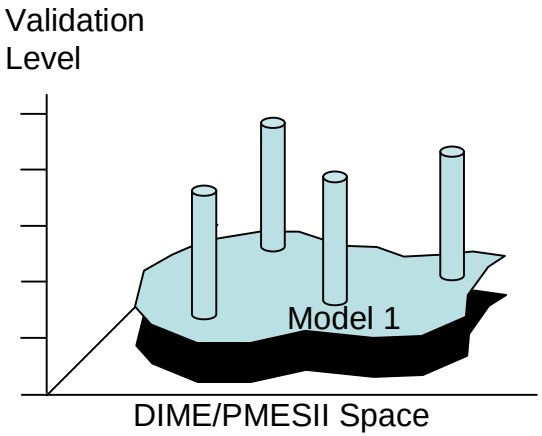
Hartley Consulting

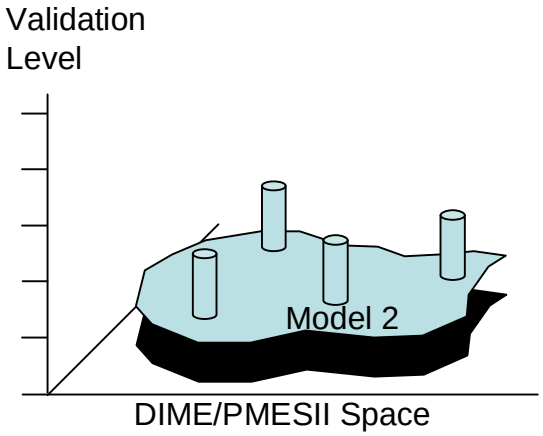
DSHartley3@comcast.net

Back up slides

Ensemble Validation Metrics

What is the overall validation level?





Is it really different from the second model?

Validation Metrics - Conceptual Models

- Labels for individual model sub-components and for entire ensemble of models
- “5” label meets the most stringent standards, “4” next most stringent, etc.
- In general, the label for an ensemble will be lower than the labels for its components
- Expectations for IW components and ensembles shaded in yellow

Label	Component	Ensemble
5	Expresses fully validated theory, e.g., Newtonian physics with caveats on operations near light speed or in regimes subject to quantum effects	Expresses multiple Level 5 theories with fully engineered interfaces, e.g., fly-out model of ground-to-air rocket, involving, chemical reactions of propellants, ballistics, air flow, electronics, etc.
4	Expresses well researched theory involving considerable data checking and peer review, e.g., economic theory earning Nobel prize	Expresses multiple Level 4 or Level 5 theories with well researched interfaces, e.g., economic model ensemble used by Federal Reserve in setting U.S. interest and discount rates
3	Expresses theory supported by data and published in peer-reviewed literature	Expresses multiple Level 3, 4 or 5 theories with considerable peer-reviewed interfaces, e.g., some U.S. combat models
2	Expresses theory with rational basis, accepted by some experts as plausible (SWAG)	Expresses multiple Level 2 - 5 theories with plausible interfaces (SWAG)
1	Expresses a codified theory (WAG)	Expresses multiple Level 1 - 5 theories with codified interfaces (WAG)
0	Uncodified, mental model of uncertain consistency and completeness	Uncodified processes for connecting models of uncertain consistency and completeness

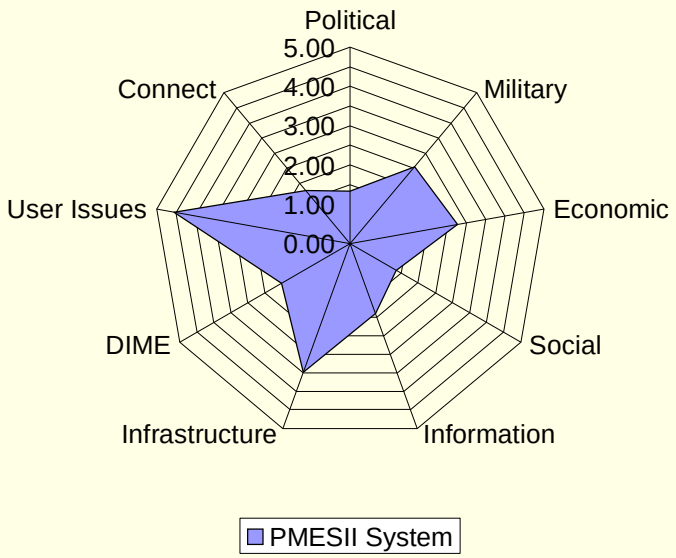
Coded Model Validation Data Entry

Name 2 **Use A**

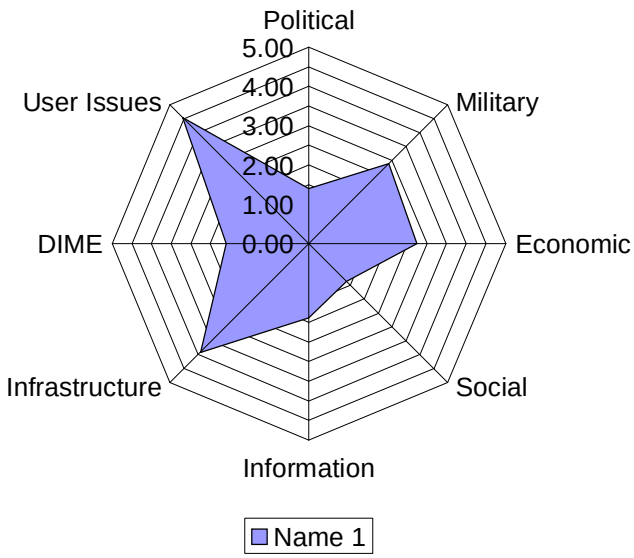
PMESII	DIME	PMESII Variables and DIME Functions	Relevant for this model? (0=no, 1=yes)	Theory (1=WAG, 2=SWAG, 3=Peer Rvw, 4=Accepted, 5=Proved)	Model Raw Score	Suitability for Use Fraction from Testing (granularity match, model match, right direction of change, right order of magnitude change)	Model Use Score
Political-Gov	DIME	Conducting nationwide elections	1	1	0.20	0.90	0.18
Political-Gov	DIME	Create local governments	1	1	0.20	1.00	0.20
Political-Gov	DIME	Educate local governments	1	1	0.20	0.90	0.18
Political-Gov	DIME	Establishing a mechanism for constitutional reform	1	1	0.20	0.90	0.18
Political-Gov	DIME	Establishing, staffing & funding effective transition national govt	1	1	0.20	0.90	0.18
Political-Gov	DIME	Providing advisors to national govt officials	1	1	0.20	1.00	0.20
Political-Gov	DIME	Supply local governments	1	1	0.20	0.90	0.18
Political-Gov	DIME	Training newly elected national political leaders	1	1	0.20	0.90	0.18
Political-Gov		Central authority is effective	1	1	0.20	0.90	0.18
Political-Gov		Central government exists	1	1	0.20	0.95	0.19
Political-Gov		Social services are adequate	1	1	0.20	0.90	0.18
		Sums	11		2.20		2.03
		Scaled Scores			1.00		0.92
Political-Pol	DIME	Maintaining compliance with peace accord milestones & conditions	1	1	0.20	0.90	0.18
Political-Pol	DIME	Mediating & negotiating w/ conflicting parties	1	1	0.20	0.90	0.18
Political-Pol	DIME	Monitoring government powersharing arrangements	1	1	0.20	0.90	0.18
Political-Pol	DIME	Transferring control of government functions to host nation officials	1	1	0.20	1.00	0.20
Political-Pol		Opposition party does not espouse force	1	1	0.20	0.90	0.18
Political-Pol		Opposition party doesn't attempt to dominate by force	1	1	0.20	1.00	0.20
Political-Pol		There are charismatic leaders advocating peace and stability	1	1	0.20	0.90	0.18
Political-Pol		There are no charismatic leaders advocating dissension	1	1	0.20	0.90	0.18
Political-Pol		There are no external forces advocating conflict	1	1	0.20	0.90	0.18
		Sums	9		1.80		1.66
		Scaled Scores			1.00		0.92
Political-ROL	DIME	(Re)building & monitoring new police force	0	3	0.00	0.00	0.00
Political-ROL	DIME	Assisting in establishing humane penal systems	0	1	0.00	0.00	0.00
Political-ROL	DIME	Assisting in establishing/reforming legitimate legal system	0	1	0.00	0.00	0.00
Political-ROL	DIME	Conducting constabulary operations	0	1	0.00	0.00	0.00
Political-ROL	DIME	Conducting war crimes investigations, tribunals, etc.	0	1	0.00	0.00	0.00
Political-ROL	DIME	Monitoring and reporting on corruption by govt officials	1	1	0.20	1.00	0.20
Political-ROL	DIME	Monitoring human rights practices	1	1	0.20	0.90	0.18

Coded Model Validation Visualization

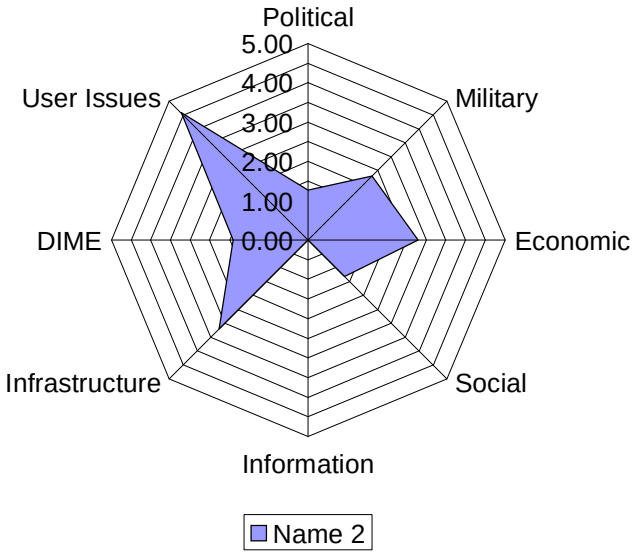
System Validation Metrics



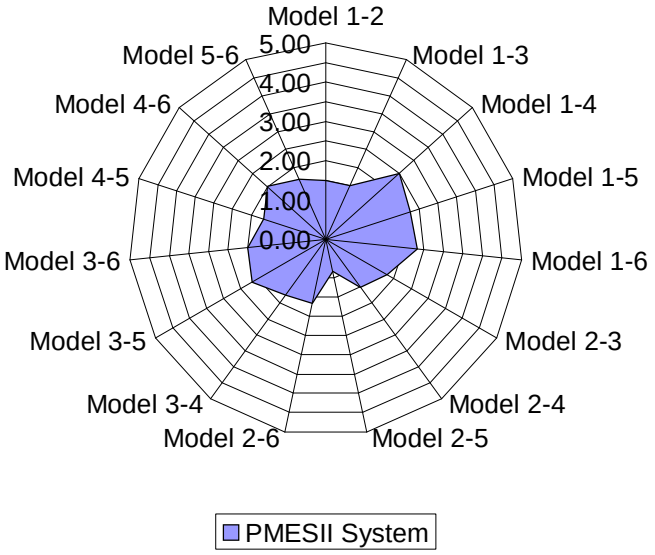
Model Validation Metrics



Model Validation Metrics



Inter-Model Connections



Risk Mitigation with IW Model

Type of Risk

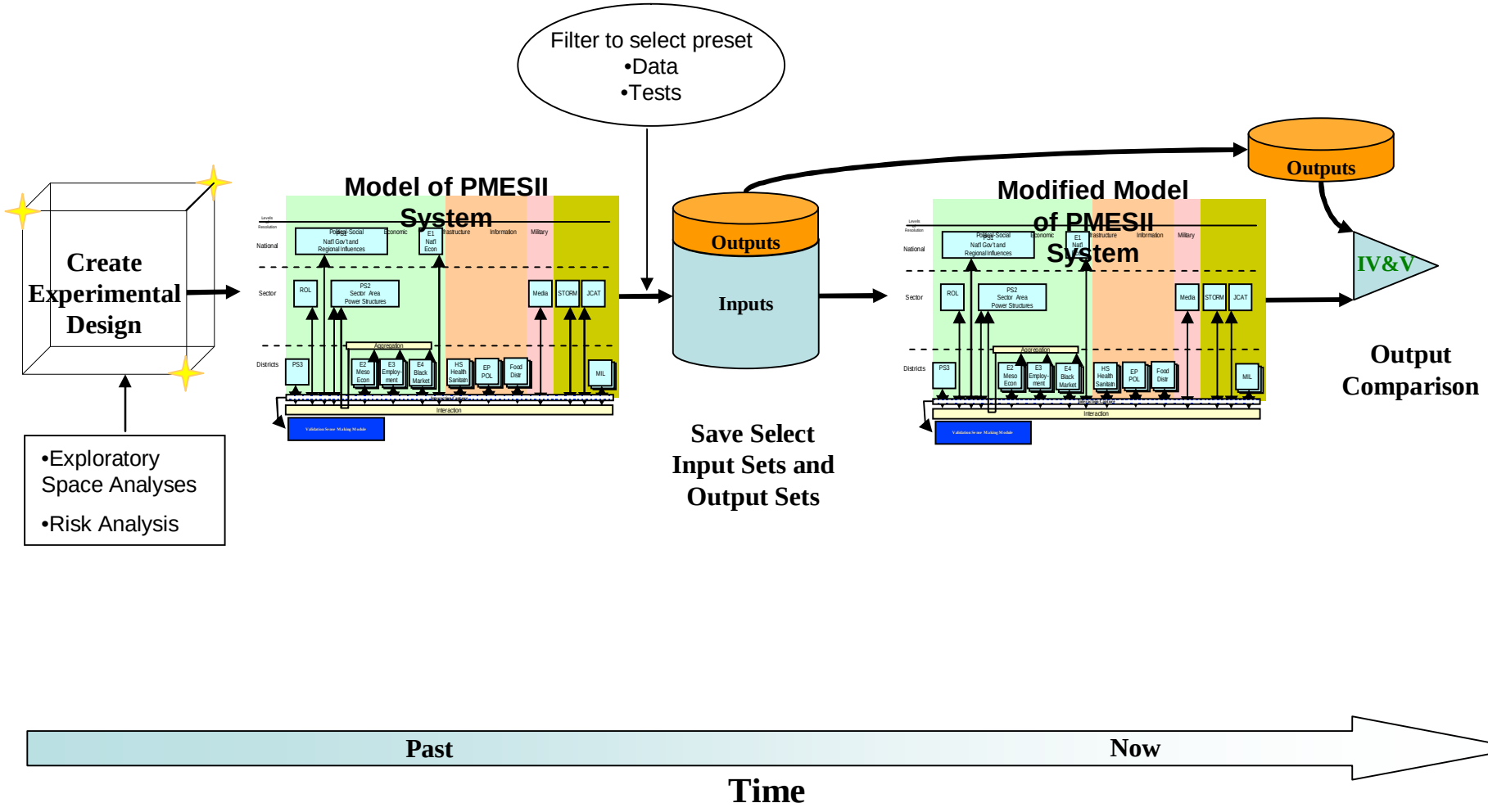
Validation Level

Risk	Methodology		Intellectual Capital			Data			
						Particular Data		General Data	
CM Val Lev	Intended use	CONOPS maturity	User experience	SME availability	SME quality	Bad input data	Mis-interpreting results	Wrong format	Content nature
Level-5	Medium M: Train users	Medium M: Train users	Minor	Minor	Minor	Major M: V&V data	Major M: Train users; Use peer review	Medium M: Train users	Minor
Level-4	Medium M: Train users	Medium M: Train users	Medium M: Train users	Medium M: Train users	Medium M: Train users	Major M: V&V data	Major M: Train users; Use peer review	Medium M: Train users	Minor
Level-3	Major M: Enforce caveats	Major M: Train users; enforce caveats	Medium M: Train users	Medium M: Train users	Medium M: Train users	Medium M: Use parametric data exploration	Medium M: Use peer review	Medium M: Train users	Medium M: Use parametric data exploration
Level-2	Major M: Enforce caveats	Major M: Train users; enforce caveats	Major M: Train users	Major M: Train users	Major M: Train SMEs; train users	Medium M: Use parametric data exploration	Medium M: Use peer review	Medium M: Train users	Major M: Use parametric data exploration
Level-1	Major M: Enforce caveats	Major M: Train users; enforce caveats	Major M: Train users	Major M: Train users	Major M: Train SMEs; train users	Medium M: Use parametric data exploration	Medium M: Use peer review	Medium M: Train users	Major M: Use parametric data exploration; use peer review

Validation Level

Risk	Tool										
	Individual Model			Model Interactions				General Tool			
CM Val Lev	Model is biased	Model is incomplete	Model is wrong	Interactions are not traceable	Interaction is biased	Interactions are incomplete	Interaction is wrong	Maturity of component	Maturity of composed M&S	Prior V&V	Configuration management
Level-5	Minor	Minor	Minor	Minor	Minor	Minor	Minor	Minor	Minor	Minor	Medium M: Institute Conf Mgmt
Level-4	Minor	Medium M: Add SME opinions	Minor	Minor	Minor	Medium M: Add SME opinions	Minor	Minor	Minor	Medium M: Institute V&V	Medium M: Institute Conf Mgmt
Level-3	Medium M: Use parametric data exploration	Major M: Add SME opinions	Medium M: All to the left	Medium M: Use parametric data exploration	Medium M: Use parametric data exploration	Major M: Add SME opinions	Medium M: All to the left	Medium M: All to the left	Medium M: All to the left	Medium M: Institute V&V	Major M: Enforce Rigorous Conf Mgmt
Level-2	Major M: Use parametric data exploration; try weighting results	Major M: Add SME opinions	Major M: All to the left	Major M: Use parametric data exploration	Major M: Use parametric data exploration; try weighting results	Major M: Add SME opinions	Major M: All to the left	Major M: All to the left	Major M: All to the left	Major M: Enforce entrenched V&V	Major M: Enforce Rigorous Conf Mgmt
Level-1	Major M: Use parametric data exploration; try weighting results; use peer review	Major M: Add SME opinions	Major M: All to the left	Major M: Use parametric data exploration	Major M: Use parametric data exploration; try weighting results ; use peer review	Major M: Add SME opinions	Major M: All to the left	Major M: All to the left	Major M: All to the left	Major M: Enforce entrenched V&V	Major M: Enforce Rigorous Conf Mgmt

Exploratory Design

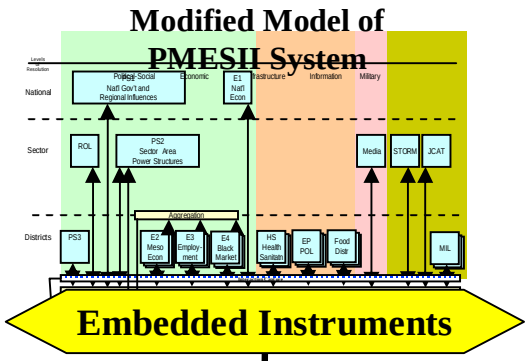


Exploratory Design: Highlights

- Methods:
  - Use Exploratory Space Analysis and Risk Analysis to identify those input data cases that possess the highest Return on Investment
  - Preset data and tests in testing module before the model changed
  - Assess based on the detection of outliers and the examination of causes to determine whether change is valid
  - Note: all the old data may not be used & new data may have to be added; other inputs may have to adjusted as well
- Challenges
  - Run time and assessment process limit the number of runs
  - May require extreme computing power
  - Automation to support the assessment process necessary

Illustrative Logic Tracing Results

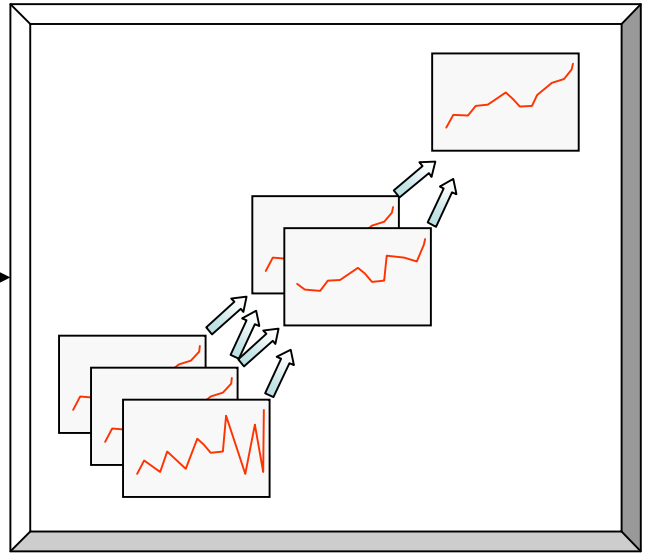
More Detail



Extract data during model execution to support later traceback of the results

Internal, Time-Coded Data

Automated Results-Traceback Program



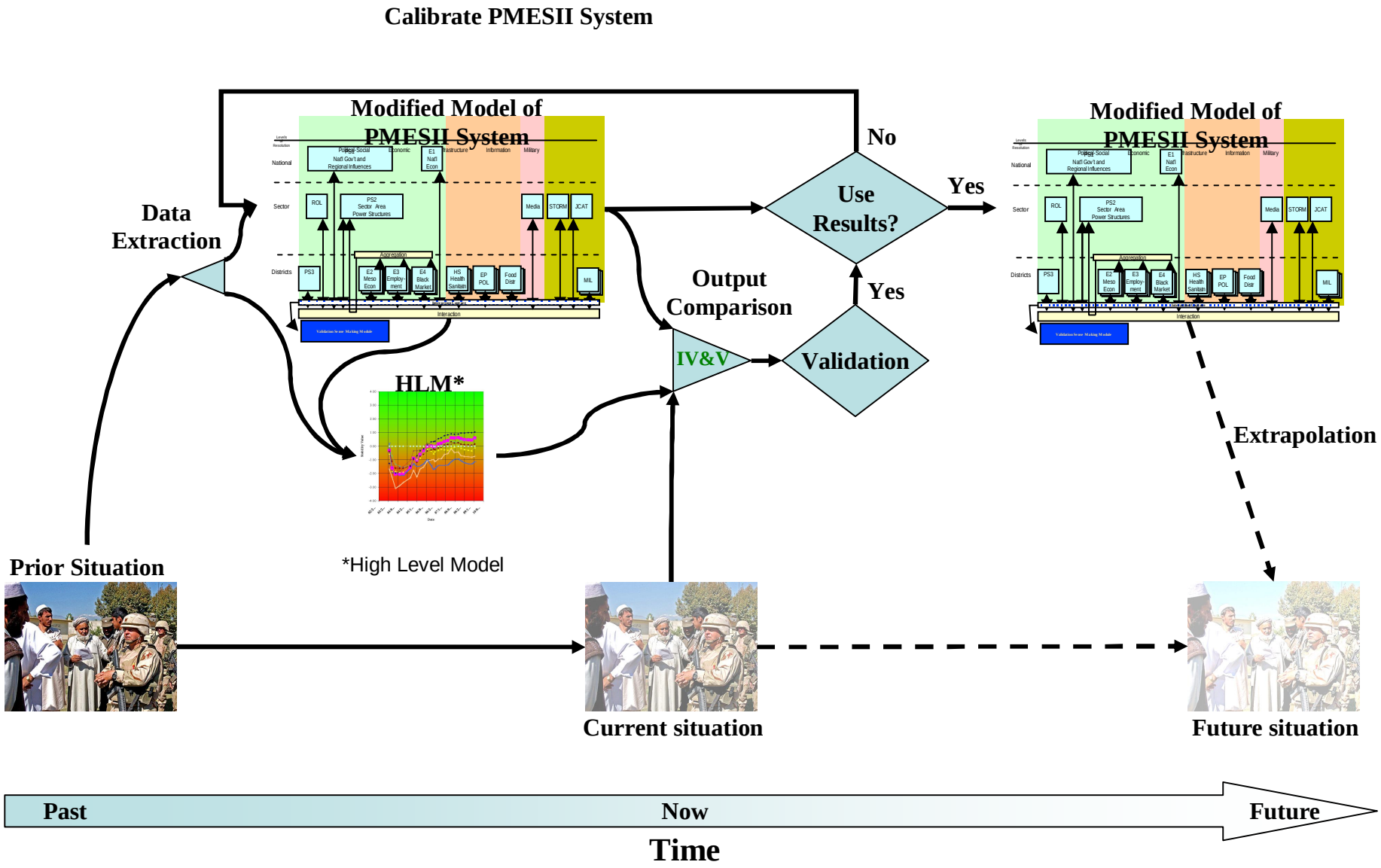
Now

Time

Logic Tracing: Highlights

- Methods:
  - Capture the data transformations with code blocks in each module (a “flight recorder”)
  - Organize the traces in a meaningful way (Automatic “logic tracing” module)
- Challenges:
  - In complex situations, cause-effect relationships are actually *causal networks* (i.e., multiple causes of multiple effects)
  - Tracing combinatorics may be extensive, requiring significant time even when fully automated
  - Automatically organizing traces in a meaningful way will be difficult
  - Creative visualization will be required to adequately support users

Illustrative Comparison



Comparison Highlights

- Methods:
  - Employ high level analytical model and contemporaneous real-world data to be continually generating Measures of Merit (MoMs)
    - Model of general problem (e.g., ISSM)
    - Model for specific subsets of problem (e.g., behavior of group)
  - Generate comparable input data using the IW Model and use the High Level Model to generate MoMs
  - Compare the outputs of the models in real time
    - Test agreement of behavior of revised IW Model through general trends (time series)
    - Attempt to calibrate IW Model in order to reach agreement
    - Use calibrated IW Model for forward projections
    - If cannot get agreement, delve deeper into potential causes
- Challenges
  - Measuring degree of agreement and level of confidence
  - **Agreement doesn't guarantee validity**

Comments on Compressed & Hyper-Compressed VV&A

- Exploratory Design: May still be too slow for hyper-compressed situations
- Logic Tracing: Can be used in either compressed or hyper-compressed situations
- Comparisons: Use with ISSM is known to be feasible for compressed & hyper-compressed situations
- Combinations: Using combinations of approaches would be beneficial for compressed and hyper-compressed situations
- Baseline VV&A: Using all of the above approaches, singly and in combination, provide useful information for later compressed and hyper-compressed situations
  - Periodically exercise compressed and hyper-compressed VV&A processes while conducting the baseline VV&A process
  - Compressed & hyper-compressed VV&A can only be performed in the context of a disciplined baseline VV&A process