



NAVAL
POSTGRADUATE
SCHOOL



Simulation
Experiments &
Efficient
Designs
Center for Data Farming
Naval Postgraduate School
<http://harvest.nps.edu>



Improving Risk Assessment in U.S. Navy Campaign Analysis

By Thomas W. Lucas, Mary L. McDonald, Captain Brian L. Morgan, Paul J. Sanchez,
Susan M. Sanchez, CDR Harrison C. Schramm, Jerry R. Smith, and Stephen C. Upton

Operations Research Department
Naval Postgraduate School
Monterey, CA.

- **Complex, high-dimensional computational models are of increasing importance in scientific research and public policy studies.**
- **Recent breakthroughs in the science for exploring these models enables much more powerful large-scale simulation studies**
- **The SEED Center is working to:**
 - **Spread the word and improve the practice!**
 - **Develop better methods (advance the theory)**
 - **Partner with organizations in DoD to quickly obtain better information from their models**
 - **Case study with STORM**



Project Overview

- **Decisions and investments made today determine the assets and capabilities comprising the U.S. Navy for decades to come. How best should we equip, organize, supply, maintain, train, and employ our naval forces? These decisions involve large sums of money and impact our national security.**
- **OPNAV N81 provides warfighting analysis to senior Navy and DoD leadership ... a modeling environment that underpins many important N81 and joint studies is the Synthetic Theater Operations Research Model (STORM).**
- **SEED Center's mission: Making modeling and simulation more effective for decision makers. (see: <http://harvest.nps.edu>)**
- **Research Objective: Develop tools and processes to reduce the amount of manpower and time required to complete STORM output post-processing for a set of replications...and create the capability to boost the speed and accuracy with which analysts gather insights.**



Background and Objective

Background

- **STORM is a stochastic, theater-wide, multi-Service campaign model widely used throughout the Department of Defense (DoD)**
- **Typical scenarios take a few hours to run per replication and contain approximately 40+ MB of input distributed over 150 input files and can generate well over 50 GB of output**

Challenge

- **Analysts can struggle in time pressure studies to integrate thousands of interactions across multiple runs to identify insights and trends**

Objective

- **Enhance understanding—help analysts quickly understand “what happened and why” by identifying consistent trends/themes, illustrating temporal sequencing, and highlighting areas worthy of detailed scrutiny**
- **Increase performance through automation and parallel processing**

Experimental objectives...

Traditional: “The three primary objectives of computer experiments are:

- (i) **Predicting** the response at untried inputs,
- (ii) **Optimizing** a function of the input factors, or
- (iii) **Calibrating** the computer code to physical data.”

--Sacks, Welch, Mitchell, and Wynn (1989)

Our goals are:

- (i) **Developing a basic understanding** of a particular model or system;
 - Identifying dominant factors and significant interactions.
 - finding regions, ranges, and thresholds where interesting things happen.
- (ii) **Finding robust** decisions, tactics, or strategies;
- (iii) **Comparing** the merits of various decisions or policies

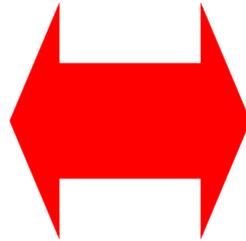
--Kleijnen, Sanchez, Lucas, & Cioppa (2005)

“Models are for thinking” – Sir Maurice Kendall

Response assumptions...

Common DOE Assumptions

- Small/moderate # of factors
- Univariate response
- Homogeneous error
- Linear
- Sparse effects
- Higher order interactions negligible
- Normal errors
- Black box model



Assumptions for Defense & Homeland Security Simulations

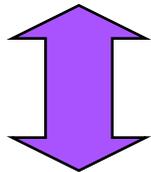
- Large # of factors
- Many output measures of interest
- Heterogeneous error
- Non-linear
- Many significant effects
- Significant higher order interactions
- Varied error structure
- Substantial expertise exists

“The idea behind [Monte Carlo simulation]...is to [replace] theory by experiment whenever the former falters—Hammersley and Handscomb

“We use simulations to avoid making Type III errors—working on the wrong model”—Mitroff and Featheringham (1974)



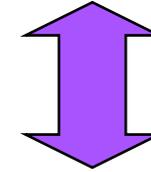
Real World



Simulation

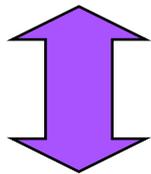
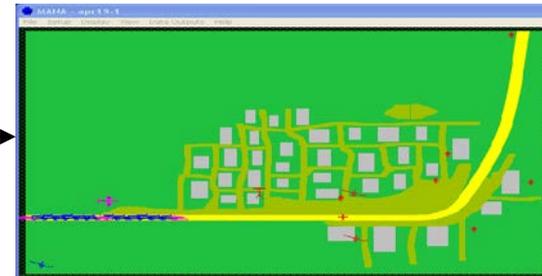
Input Factors (X)

1	6	1	4	6
3	2	8	5	3
2	3	5	3	2
4	5	7	2	1
7	4	6	1	4
5	1	2	8	7
6	2	3	2	8
8	8	4	8	5

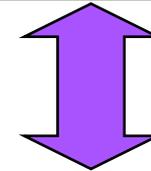


Output Responses (Y)

4	6
5	3
3	2
2	1
1	4
6	7
7	8
8	5



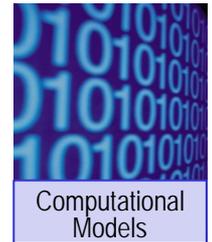
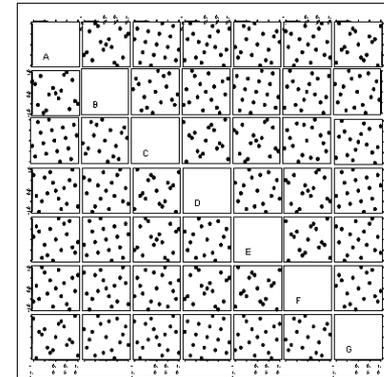
Meta-Model
“model of a model”



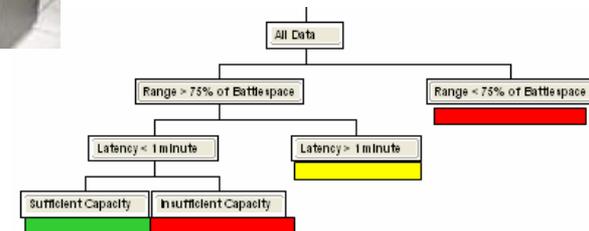
$$y = \beta_0 + \underbrace{\sum_{j=1}^k \beta_j X^j}_{\text{Linear Effects}} + \underbrace{\sum_{j=1}^k \beta_{jj} (X^j)^2}_{\text{Non-linear Quadratic Effects}} + \underbrace{\sum_{i=1}^{k-1} \sum_{j>i}^k \beta_{ij} X^i X^j}_{\text{Two-Way Interaction Effects (Synergies)}} + \underbrace{\epsilon}_{\text{Random Noise}}$$

SEED Center Mission: Making modeling and simulation more effective for decision makers

- Harnessing Enabling Technologies**
 - New Design of Experiments (DOE)
 - Computational models
 - High-performance computation
 - Data mining and visualization
- Revolution in analysis capabilities**
 - Quickly yield broad insight
 - Address uncertainties
 - Assist in VV&A
 - Enable massive sensitivity analysis



**Enable rapid and efficient
 computational
 experimentation and analysis**




 Simulation
 Experiments &
 Efficient
 Designs
 Center for Data Farming
 Naval Postgraduate School
<http://harvest.nps.edu>

Petaflop machines like Roadrunner have the potential to **fundamentally alter science and engineering...[allowing scientists to] perform experiments that would previously have been impractical.** *The New York Times, June 9, 2008*

Moore's Law is not enough!

The “curse of dimensionality” will not be solved by hardware alone.



Petaflop = 1000 trillion ops/second
Cost of “Roadrunner” = \$133 million

“2¹⁰⁰ is forever”

—Maj Gen Jasper Welch

Even with a petaflop computer and a simulation that runs as fast as a single operation, one replication would take over 40 million years...but efficient experimental designs could let us study all factors for longer-running simulations with multiple replications over a weekend...



STORM 2.4 Campaign Model

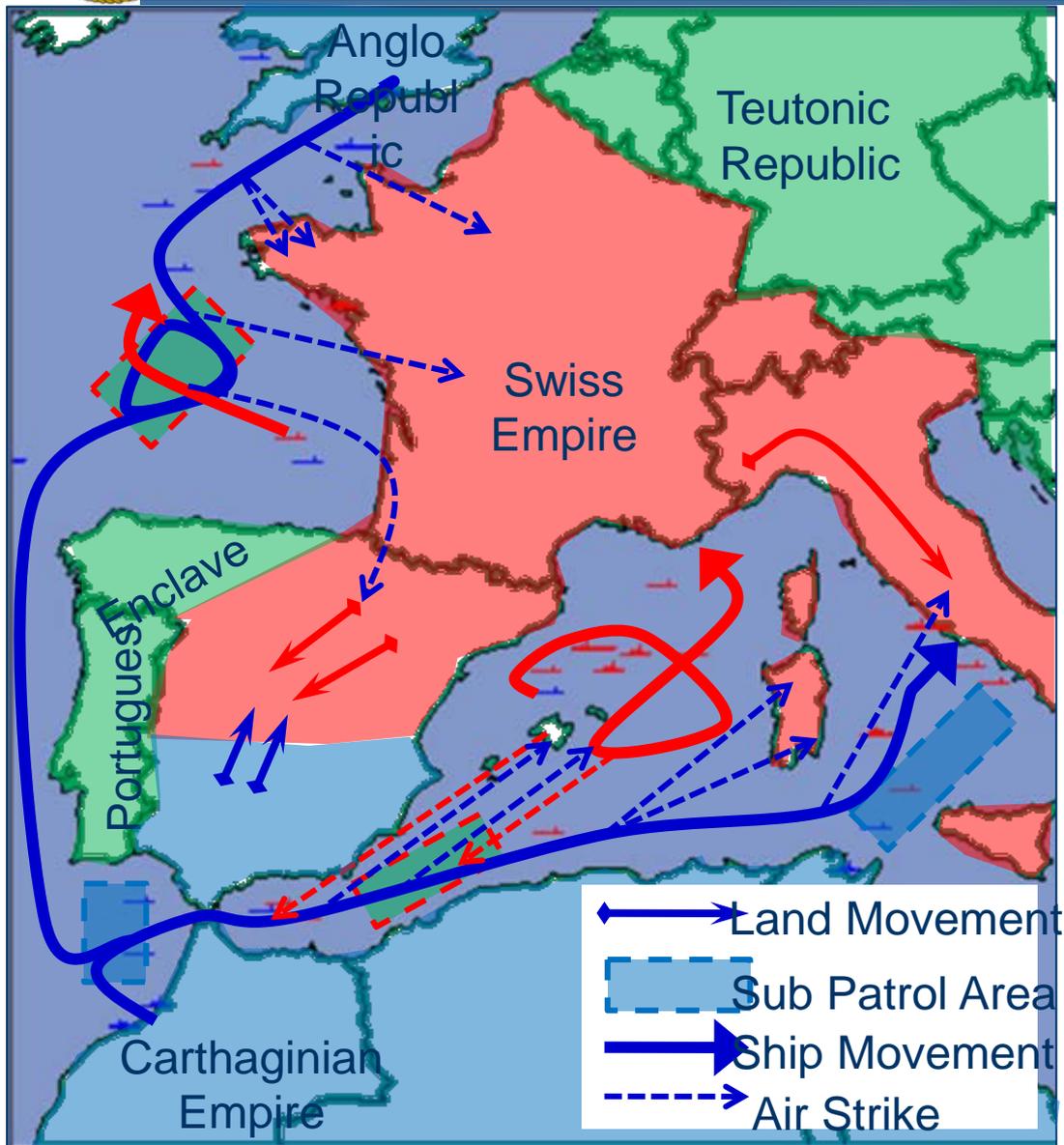


- **Multi-sided, stochastic simulation of air, space, ground and maritime planning and execution - the new DoD standard**
 - Includes options for logistics, maintenance, ISR, and weather
 - Perception-based interactions – auto-planning shaped by user priorities and ISR feeds
 - Auto-generated or user-defined flight paths
 - Territorial control determines who uses, repairs, builds bases
- **Data-driven, closed form – no analyst-in-the-loop**
- **Rule-based OPLAN / CONOPS**
- **Capable of executing and tracking**
 - Multiple cases: User access to full range of input data
 - A range of outcomes within a case: Stochastic interactions
- **Key STORM users include: OSD, USAF, USN, MCCDC, J8, PACOM, USFK**
 - Maritime and expeditionary functionality complete





Test Scenario: Punic War II



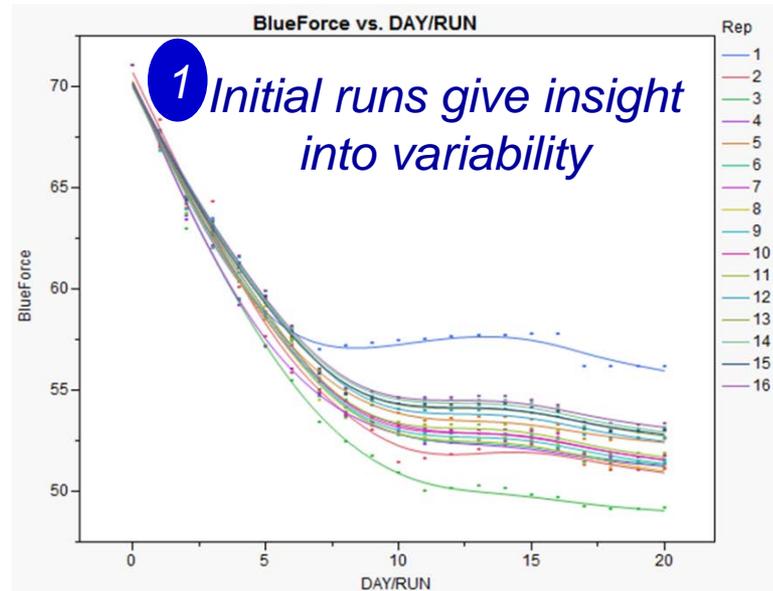
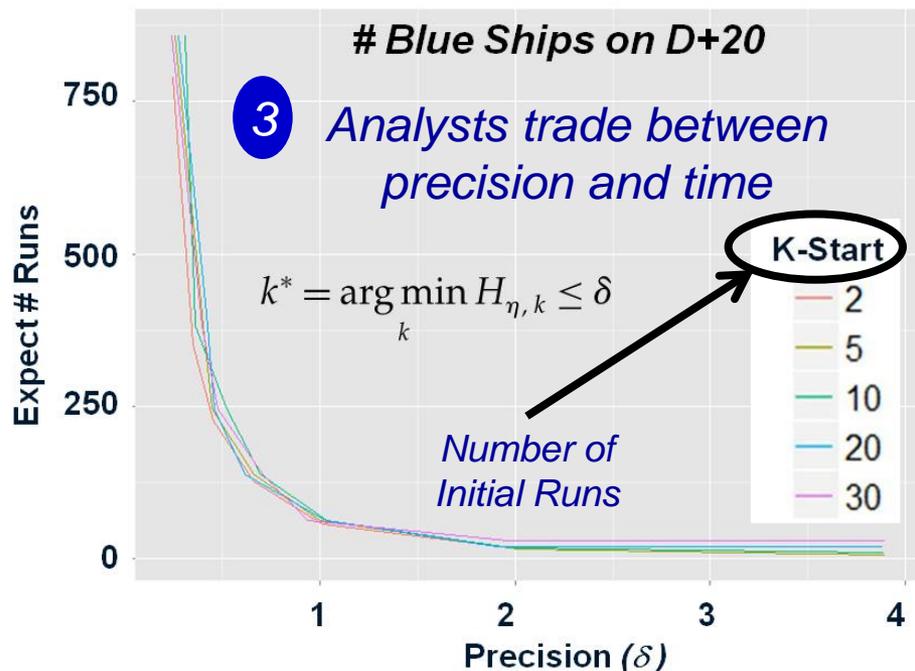
- Background: Swiss Empire (Red) seizes financial accounts and restricts exports of cheese, chocolate, and time pieces
- In response Carthaginian Empire (Blue) plans invasion of Swiss Empire
- Anglo Republic (Blue) swings CSG to Med and escorts ESG from Casablanca
- Red Subs and SAG attack in Bay of Biscay
- Red and Blue CSGs attack in Med
- Blue amphibious invasion near Rome
- Portuguese Enclave and Teutonic Republics remain Neutral with no fly-over rights

All presented data in this presentation were generated from the unmodified, unclassified STORM training scenario (Punic II)

Analysts Determine When “Enough is Enough”



- Calculation performed after initial runs
- Enables trades between more precision in estimates and cost of runs
- Normality of underlying data is not



- 2** Analysis yields estimates of Runs vs. Precision requirements

$$H_{\eta, k} = t_{\eta, k-1} \sqrt{\frac{S_k^2}{k}}$$

$$k^* = \min_{k \geq 2} \{k : H_{\eta, k} \leq \delta\}$$

[1] “Optimal Absolute Precision Stopping Rules for Simulation Experiments,” by Dashi Singham and Lee Schruben



Example: Killer-Victim Scoreboard



<i>Killer/Victim</i>	Blue Carrier	Blue Sub	Blue Surface	Blue Other	Red Carrier	Red Sub	Red Surface	
Blue Surface	0/0	0/0	0/0	0/0	0.33/0.47	1.29/0.75	7.40/2.26	3.01/3.43
Blue Sub	0/0	0/0	0/0	0/0	0/0	4.19/1.40	3.66/2.49	3.93/2.03
Blue Other	0/0	0/0	0/0	0/0	0/0	3.13/1.27	0/0	3.13/1.27
Red Surface	0/0	0.08/0.28	2.11/1.57	0/0	0/0	0/0	0/0	1.10/1.52
Red Sub	0.06/0.25	4.82/1.13	5.70/2.94	0.32/0.60	0/0	0/0	0/0	2.73/3.02
Red Other	0/0	1.77/1.05	0.69/0.76	0/0	0/0	0/0	0/0	1.23/1.07
	0.06/0.25	2.22/2.16	2.83/2.88	0.32/0.60	0.33/0.47	2.87/1.68	5.53/3.02	

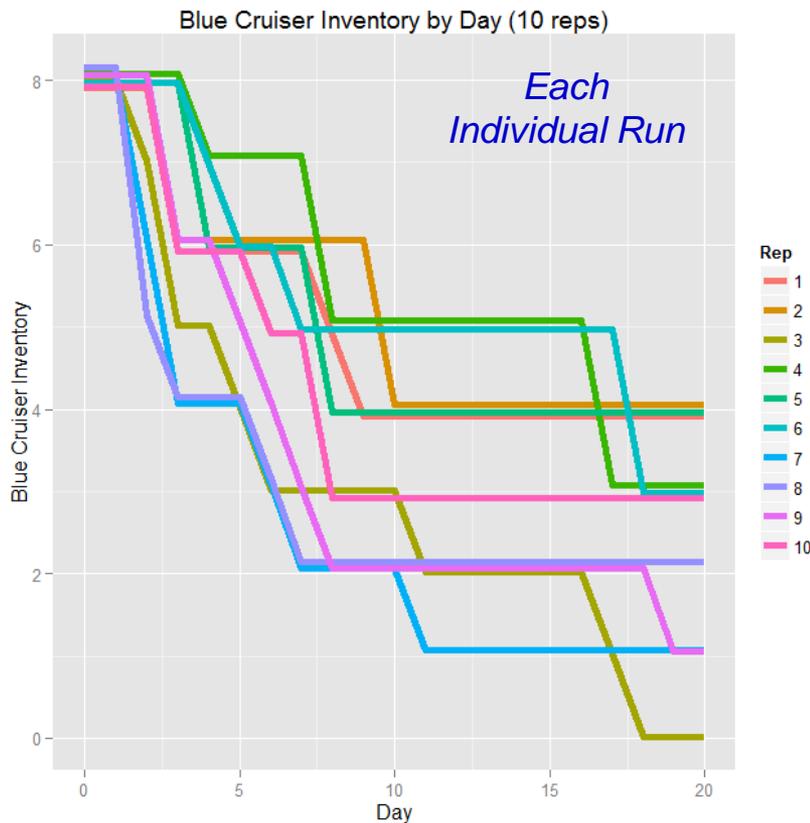
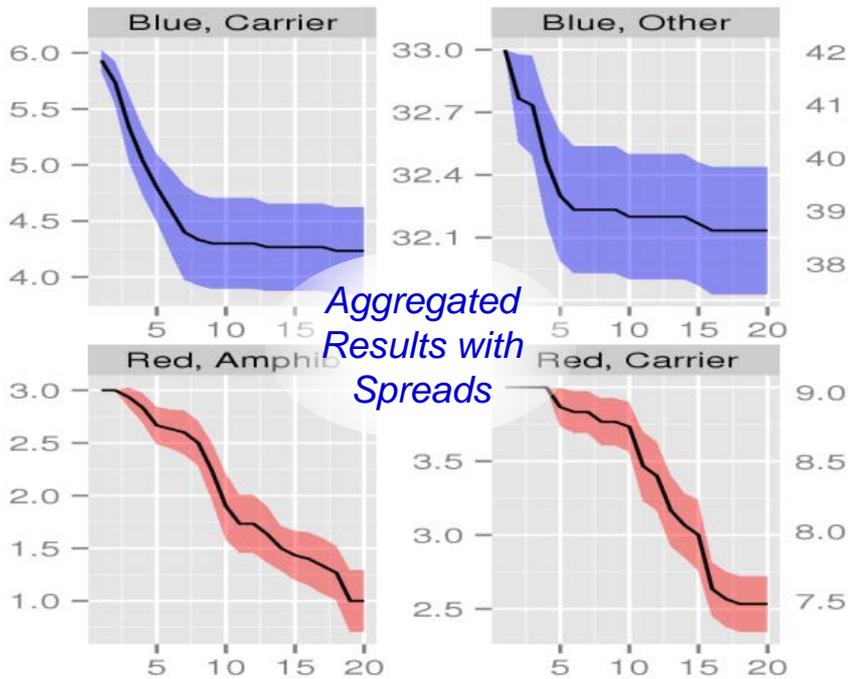
Biggest threat to Blue Surface Combatants was Red Subs

Each cell contains: Mean / SD (over the reps)

With Row/Column summaries

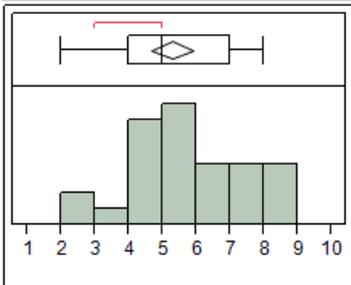


Variability and Summary Statistics



Distributions

Blue_Cruiser_killed



Quantiles

100.0%	maximum	8
99.5%		8
97.5%		8
90.0%		8
75.0%	quartile	7
50.0%	median	5
25.0%	quartile	4
10.0%		3.1
2.5%		2
0.5%		2
0.0%	minimum	2

Summary Statistics

Mean	5.3
Std Dev	1.684616
Std Err Mean	0.3075674
Upper 95% Mean	5.9290459
Lower 95% Mean	4.6709541
N	30

Analysts Pick which Metrics to Examine and Track over the Scenario



Dashboard: User-Defined Success Criteria



Criteria Name links to Metric's Definition

Metric's Value at End of Run

Run Number



Fail

Pass

In Jeopardy

variable	1	8	16	25	2	7	33	46	28	31	39	35	40	3	32	24	45	5	13	20	44	22	38	11	15	23	9	
Carrier Losses	0	0	0	0	0	0	2	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
SurfaceShip Losses	6	6	14	10	12	5	6	6	8	8	12	14	12	12	13	4	11	9	10	6	9	7	4	2	5	6	2	
Amphib Losses	1	0	0	0	0	0	0	2	0	1	0	1	0	0	3	0	0	0	1	0	0	3	0	0	0	0	1	
Sub Losses	7	7	5	7	8	6	5	7	7	9	5	6	7	6	6	7	5	6	6	5	6	7	7	6	5	5	6	7
BlueAirSupremacy	0	4	5	7	3	5	0	5	0	5	6	0	0	6	5	0	0	2	5	6	5	5	1	9	10	6	0	0
RedAirSupremacy	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
RedAirSuperiority	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	
isRedCarrierDead	4	0	0	7	0	0	0	6	0	0	3	17	0	0	0	4	0	0	0	6	0	15	9	13	6	10		
carthageCBGHasEnteredMed	30	31	31	30	30	30	31	30	30	30	30	30	30	31	30	30	30	31	30	30	30	31	30	30	30	31	30	
carthageESGHasEnteredMed	29	30	30	29	30	29	30	30	30	29	29	29	29	29	30	29	29	30	30	29	29	29	30	29	30	30	30	
carthageESGHasArrivedOffRome	6	6	6	7	12	6	6	6	6	6	6	11	6	6	6	6	6	6	6	6	6	6	6	6	10	6	6	
expeditionaryOpsHaveBegun	5	5	5	5	11	5	5	5	5	5	5	9	5	5	5	5	5	5	5	5	5	5	5	5	9	5	5	
gibraltarMinesweepingHasStarted	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	
seabaseIsComplete	37	37	37	37	37	37	0	0	37	37	37	37	37	37	37	37	37	37	37	37	37	37	37	37	37	37	37	
cruisersHaveArrivedAtSeabase	37	37	37	37	37	37	37	37	37	37	37	37	37	37	37	37	37	37	37	37	37	37	37	37	37	37	37	
seabaseNWScreenIsComplete	32	0	0	0	0	20	0	0	0	0	0	0	0	0	0	0	33	30	0	0	0	31	32	33	33	32	34	
seabaseNEScreenIsComplete	37	37	37	37	37	37	0	37	0	0	0	0	0	0	0	37	37	37	37	37	37	37	37	37	37	37	37	
seabaseSWScreenIsComplete	32	27	0	0	0	0	0	0	33	30	0	0	0	0	26	0	33	30	0	18	30	0	0	0	33	32	34	
seabaseSEScreenIsComplete	0	0	0	0	0	0	37	37	37	37	37	37	37	0	37	37	37	37	37	37	37	37	37	37	37	37	37	

Color's indicate which Criteria are met at End of Scenario

Quickly Identifies which Objectives are Frequently Not Met

* Presented data from the unclassified STORM training scenario

Heat Maps Depict Time and Space Relationships



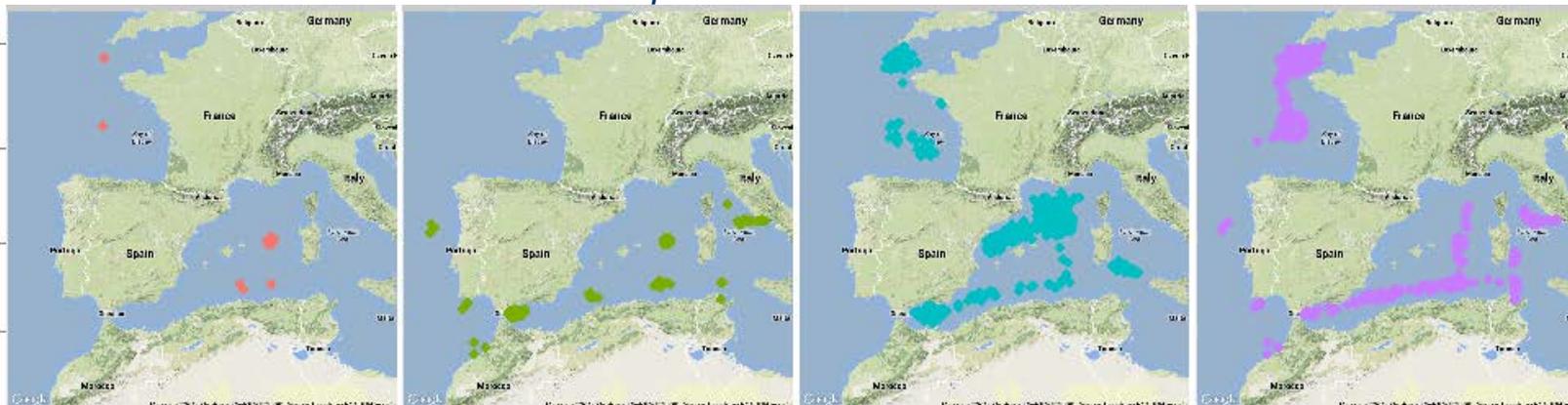
Carrier

Aux Ship

Submarine

Combatant

Unit Type – Space



Damage
to
Specific
Blue Type
at Any
Time

110
Runs

D+0 to D+3

D+3 to D+5

D+5 to D+10

D+10 to D+15

Time – Space



Damage
to Any
Blue Unit

Quickly Enable Analysts to Identify Illustrate Where and When Activities are Occurring to Which Units



Illustrative Examples (1)

Partition for Blue_Advanced_Carrier_killed

RSquare	RMSE	N	Number of Splits	AICc
0.825	0.3633053	30	2	33.9856

All Rows			
Count	30	LogWorth	Difference
Mean	1.3333333	8.0801369	1.5873
Std Dev	0.8840866		

Sum_GAR_Cnx<4			
Count	21	LogWorth	Difference
Mean	0.8571429	4.6424604	0.8625
Std Dev	0.4780914		

Sum_GAR_Cnx>=4			
Count	9	LogWorth	Difference
Mean	2.4444444		
Std Dev	0.5270463		

Blue_Future_Naval_MultiRole_Fighter_killed<67			
Count	5	LogWorth	Difference
Mean	0.2		
Std Dev	0.4472136		

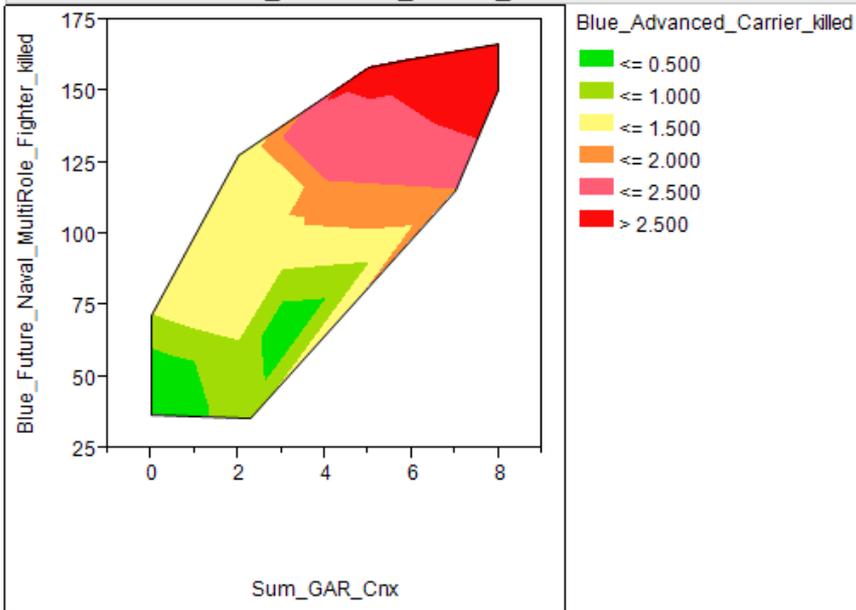
Blue_Future_Naval_MultiRole_Fighter_killed>=67			
Count	16	LogWorth	Difference
Mean	1.0625		
Std Dev	0.25		

GAR = Ground Alert Air Refueling

Simple meta-models with high explanatory power for key campaign metrics

Graphs that support main insights

Contour Plot for Blue_Advanced_Carrier_killed

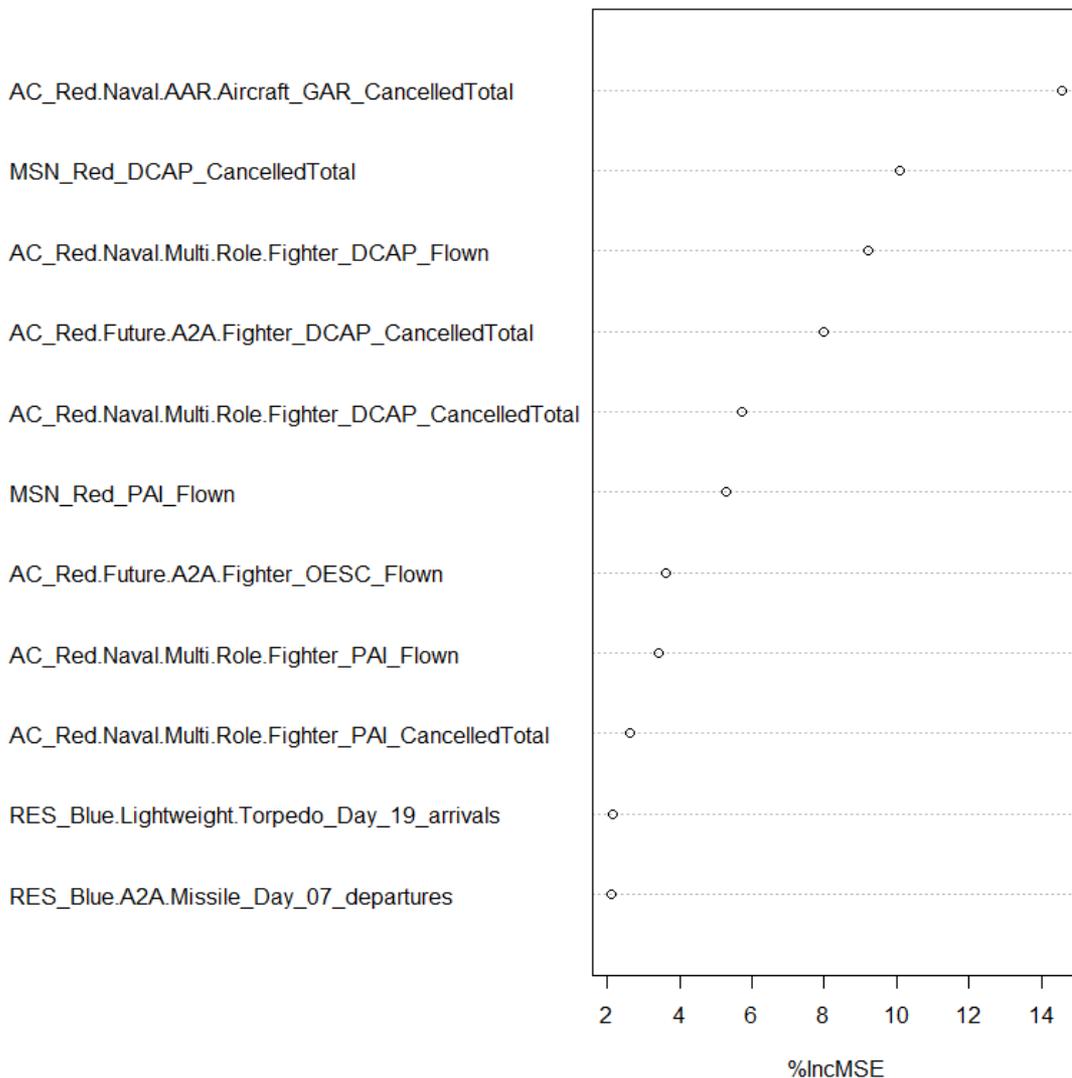




Example: Variable Importance for RedCarrierDead

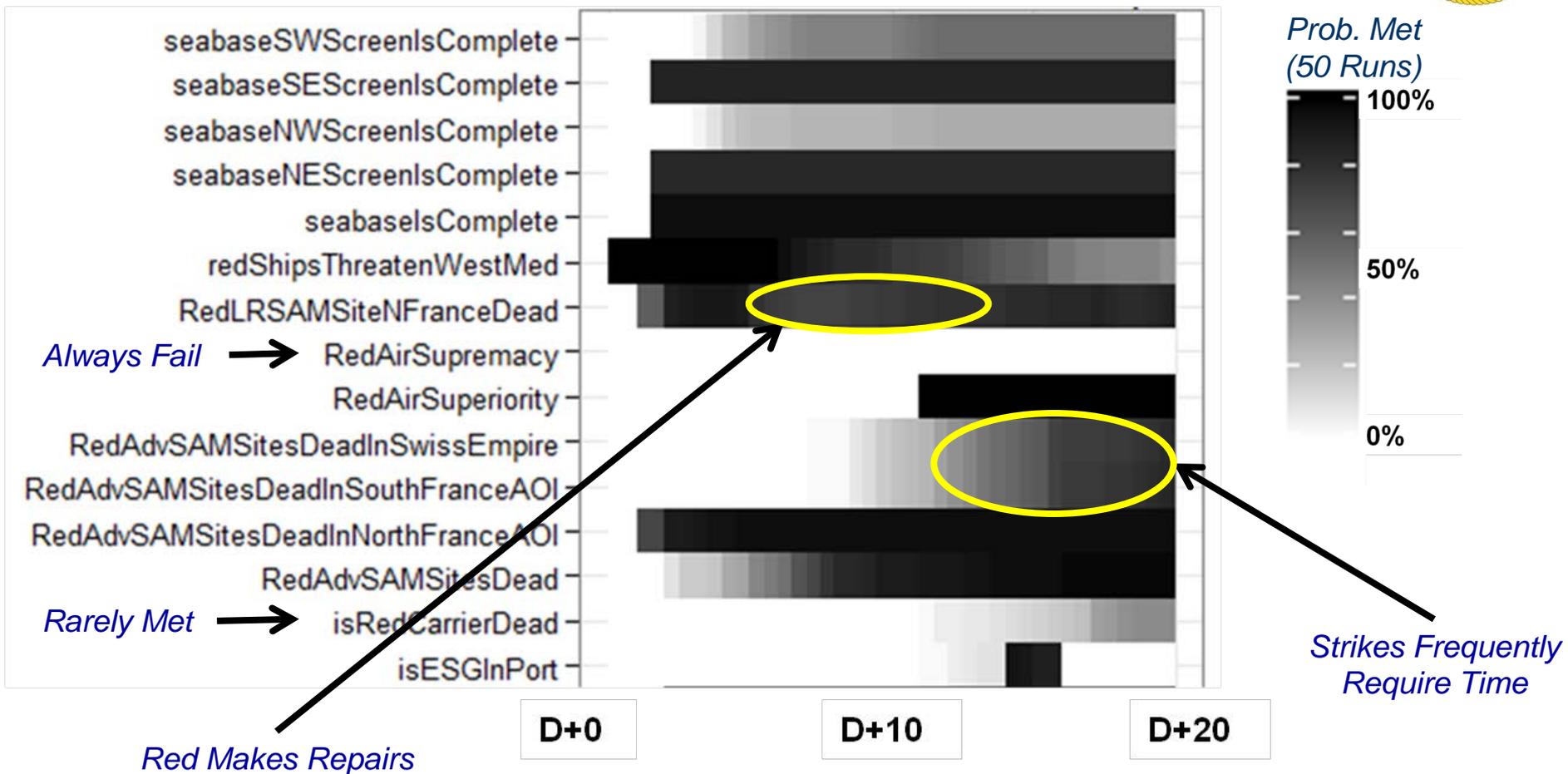


Variable Importance for RedCarrierDead



- Variable Importance results from running a Random Forest of Partition Trees
- Top variables will be depicted in a single partition tree
- Partial dependence plots will be generated for top variables
- Top two variables will be displayed in a contour plot

Time Plots: Time Distributions of Meeting Objectives



Time Plots Enable Analysts to Identify when Objectives are Met and Illustrate the Statistical Ebb and Flow of the Scenario

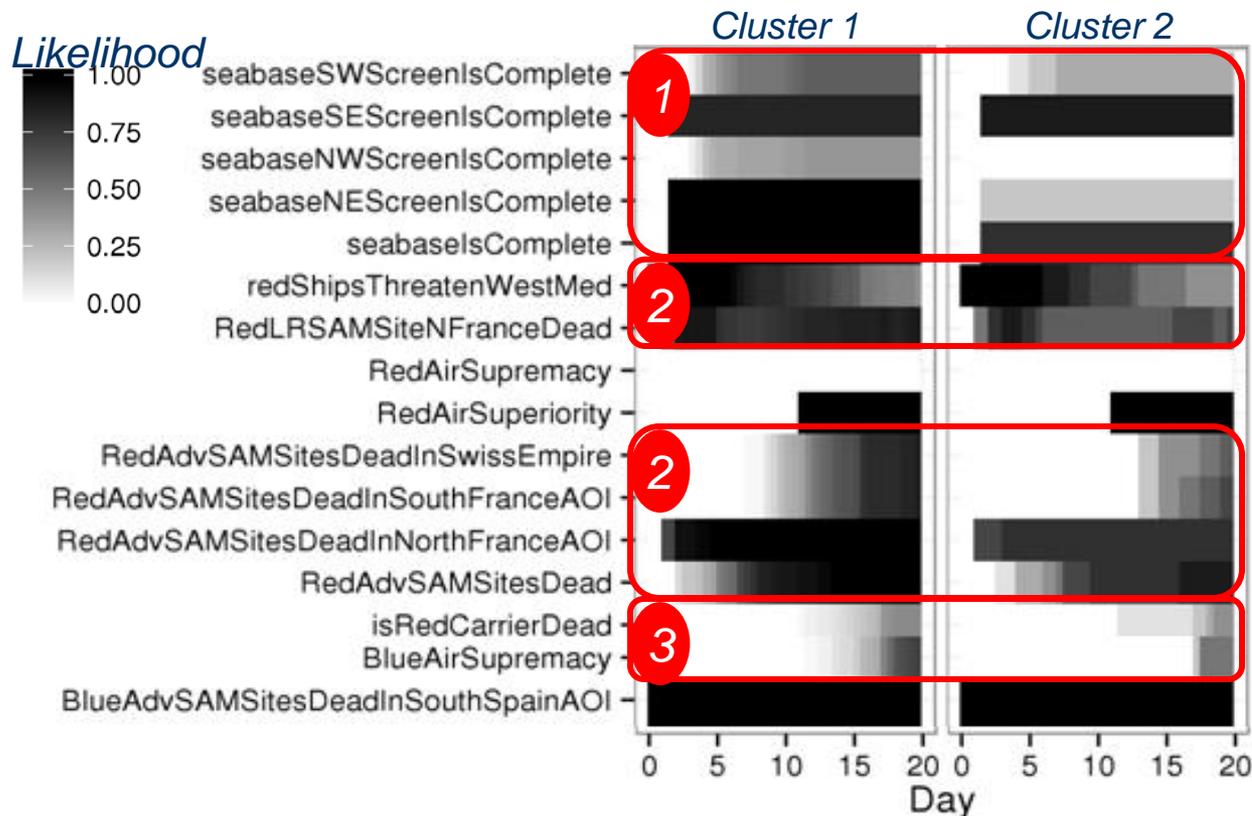
* Presented data from the unclassified STORM training scenario



Examining Cluster in Detail Identifies Cause of Branching



Continuing from last slide, ...



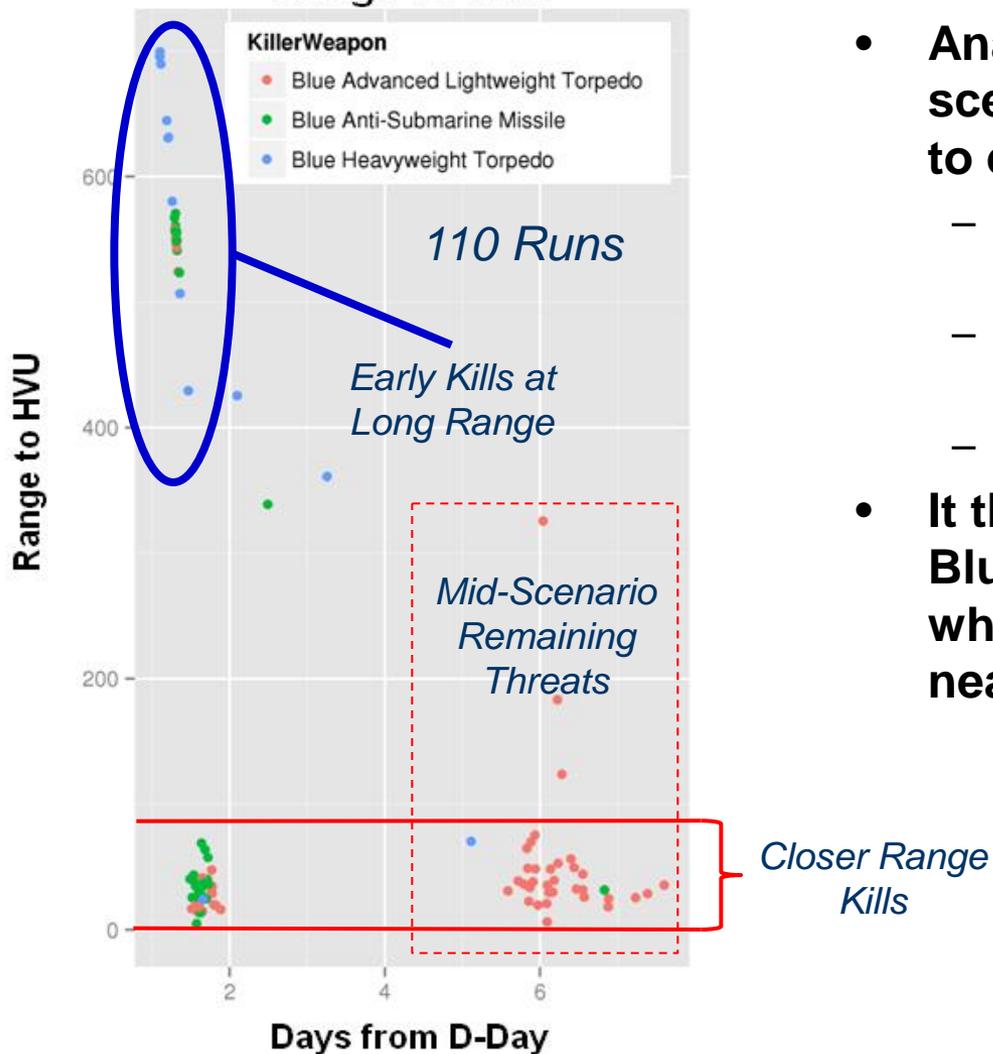
1. Not completing Screens increases Blue CV losses which means ...
2. Fewer Blue CVs to destroy the Red Adv SAMs (there are fewer available Blue A/C), which means ...
3. Blue does not gain Air Supremacy to support the landing near Rome

Although Scripts cannot Find Causality, they can Guide Analysts

Users May Define Details for Casualties of Interest



Red Submarine Kills Range vs HVU



- Analysts can identify additional, scenario specific metrics and events to examine in greater detail
 - Uses standard print-statements within STORM log files
 - Add log statements to existing post-processing scripts inputs book
 - User plots data with any graphic tool
- In this case, user examined which Blue weapons killed Red subs, at what time, and at what range to the nearest High Value Unit (HVU)

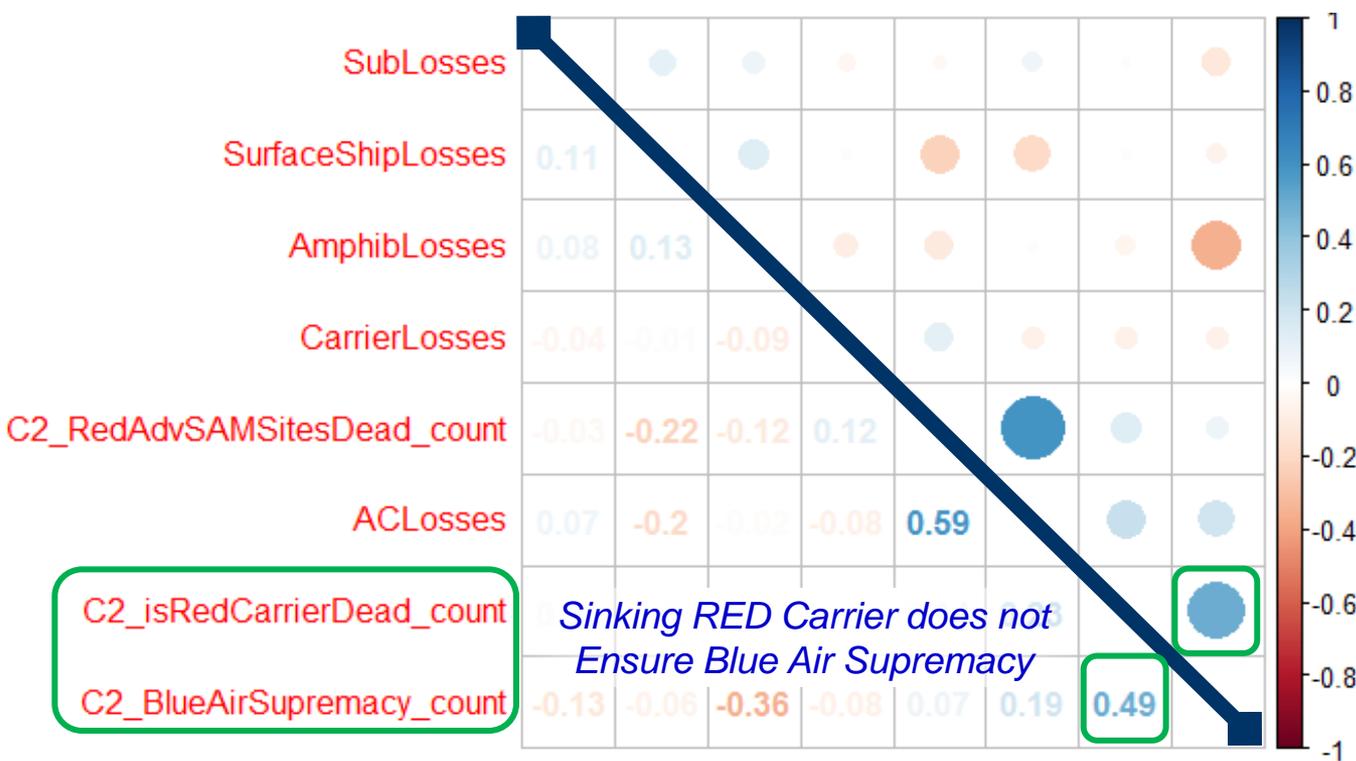
*Automatically Quantifies
Time-space Threat Metrics of
Interest*



Correlations of Success Criteria



- With thousands of interactions over dozens of runs, it is difficult to identify which events and outcomes are strongly interrelated
- Though many correlations are obvious (sortie rate and fuel reserves), many may not be immediately evident

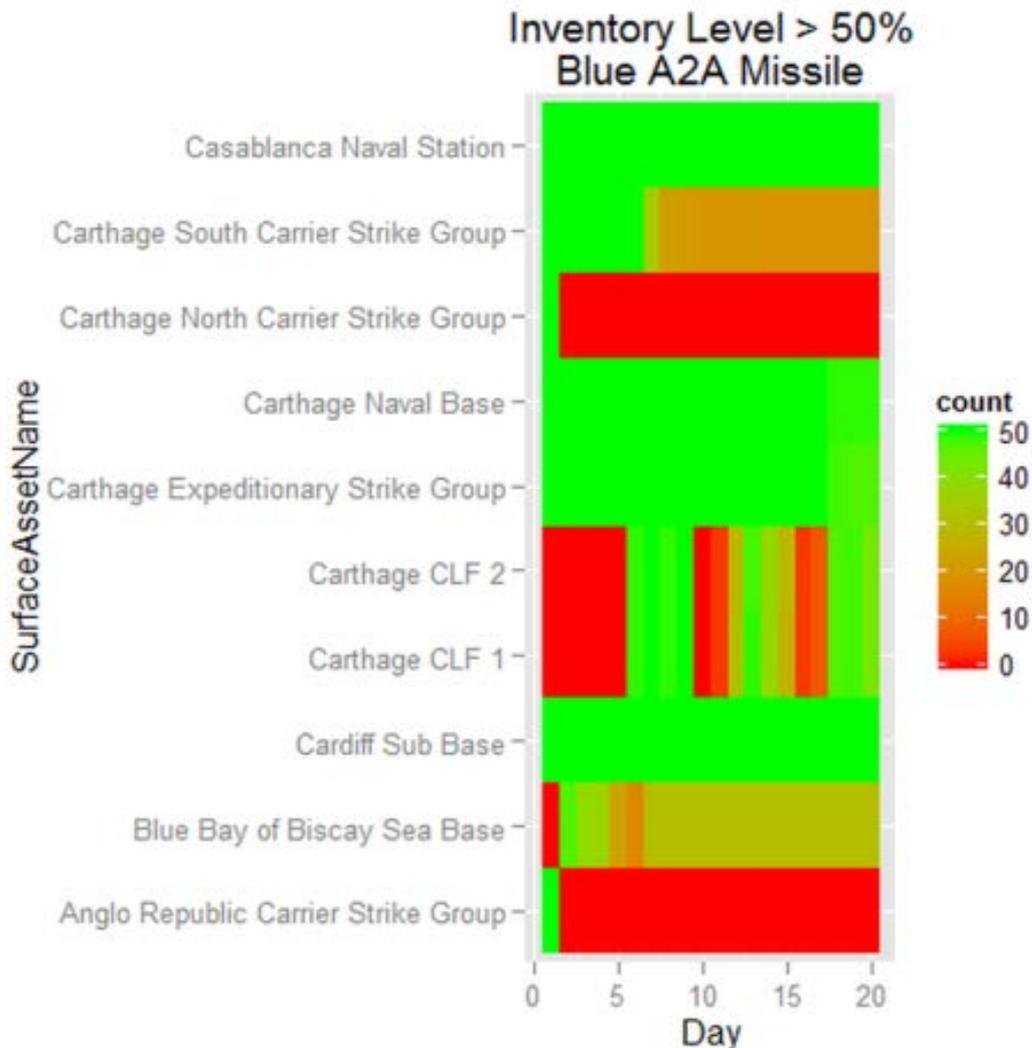


- Columns correspond with the rows
- Matrix is symmetric
 - $A(i, j) = A(j, i)$
- Correlation values in lower half with visuals (circles) in upper half
- Circle sizes depict strength of correlation
- Colors denote type of correlation
 - Blue for positive
 - Red for negative

Analysts Select Data Fields Included in Correlation Checks



Example: Ammo Inventory > 50% Blue A2A Missile, by Resource Carrier



Legend:

- Count is over the 50 reps
- Green means that the Inventory Level was over 50% for every rep
- Red means that the Inventory Level was never over 50%, for any rep
- In-between colors represent variability over the reps

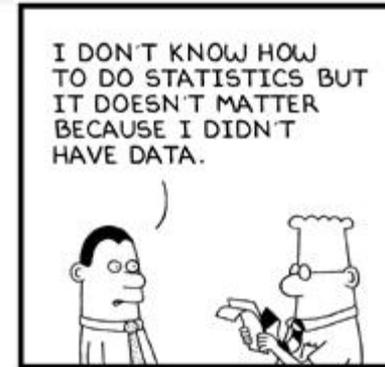
- **OPNAV: “[T]his approach has dramatically reduced the time and costs of studies...[which] are now approximately **33% faster and 16% less expensive**. These methods have also increased our analytic maturity across the organization.”**
- **Next steps: Make STORM data farmable. That is, embed STORM in an environment that facilitates using sophisticated design of experiments to run experiments on a computing cluster.**

<http://harvest.nps.edu>

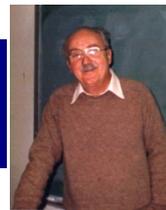
Check here for updates on:

- Lists of student theses (available online)
- Spreadsheets & software
- Pdf files for several publications, reference info for the rest
- Links to other resources (e.g., *Scythe*)

Or, contact Tom Lucas (twlucas@nps.edu) or Susan Sanchez (ssanchez@nps.edu) for more information



All models are wrong, but some are useful—George Box



It's the analyst, not the model, that produces important useful results. Improve the former before the latter — Seth Bonder



Simulation
Experiments &
Efficient
Designs
Center for Data Farming
Naval Postgraduate School
<http://harvest.nps.edu>