IMPLICATIONS OF MULTI-RESOLUTION MODELING (MRM) AND EXPLORATORY ANALYSIS FOR VALIDATION

James H. Bigelow, RAND Corp

Paul K. Davis, RAND Corp & RAND Graduate School

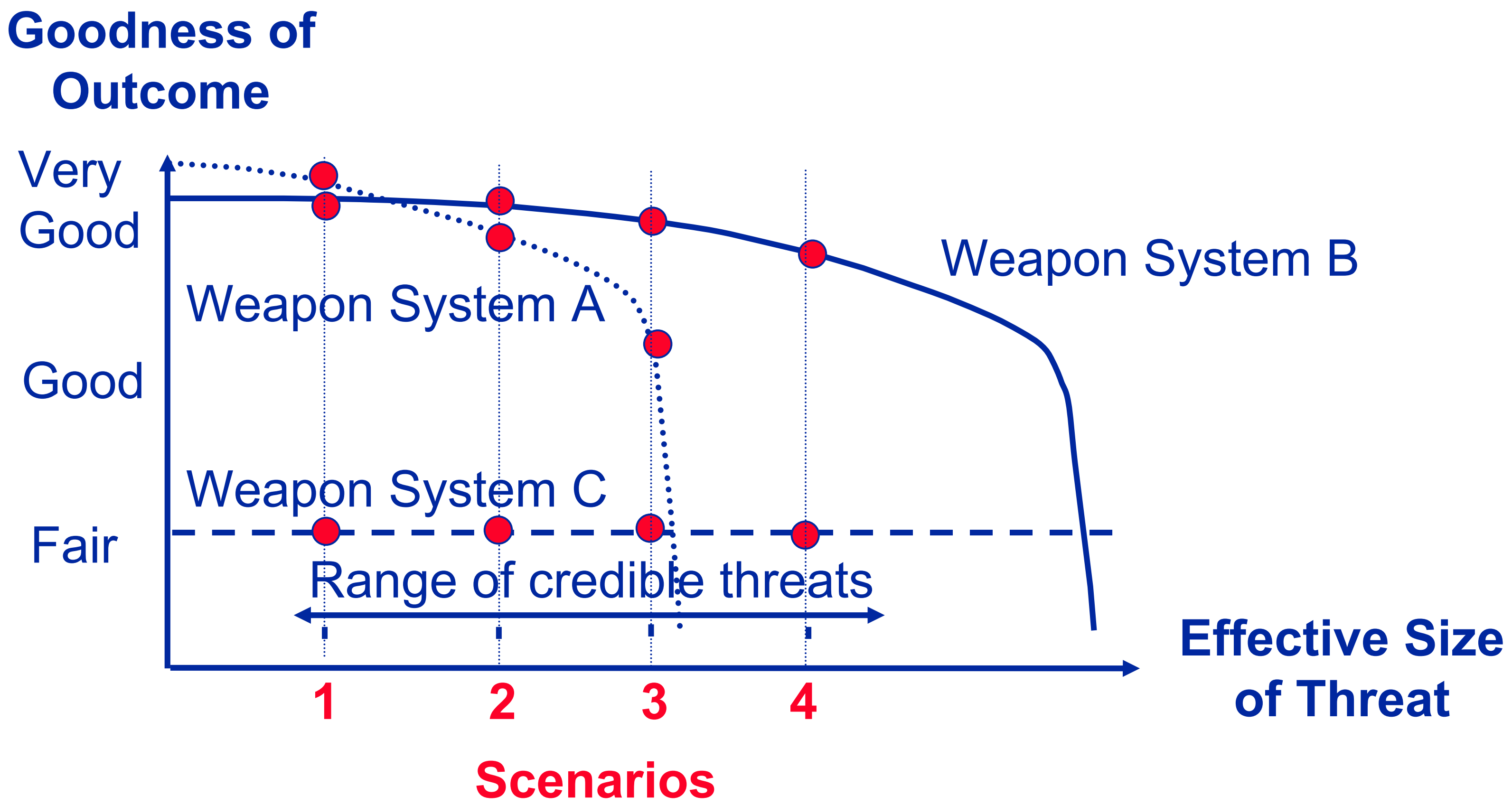
Foundations ‘02

October 23, 2002

Common Examples of Multi-Resolution Thinking

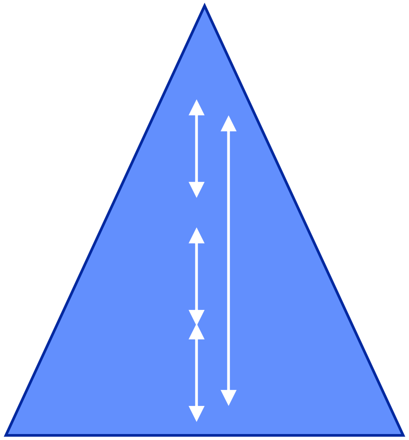
- Back-of-the-envelope calculation to check a more detailed computation
  - Estimate aircraft cost by component (e.g., engine, airframe, avionics, etc.); check against historical cost per pound
- Stories to explain model behavior
  - Used to debug model or data
- Simple displays for presenting, explaining, and generalizing results to the client
  - Model shows the analyst a few trees; analyst shows the client the forest

Show Client a Simple Display, Generalize From High-Resolution Results



Multi-Resolution Modeling to Link Levels of Analysis

Level	Data
Strategic	→
Operational	→
Tactical	→
Entity	→



MRM is building families of mutually consistent models at different levels of resolution

- Design family together (integrate from start)
- Use all relevant info.
- Seek mutual calibration
- Construct “good seams”

Exploratory Analysis, Massive Uncertainty, and Validation

- Many problems we work on have massive uncertainty, e.g.,
  - Technology and threats will change between now and 2030
  - We can't build predictive models
- Use an exploratory analysis strategy
  - Look for robust and adaptive policies
  - Requires:
    - Agile models (few variables)
    - Unbridled use of extrapolation
- When is model & data for exploratory analysis "valid"?
  - Tentative suggestion:
    - Model is structurally correct
    - Domain for analysis includes all cases of interest
  - What practical tests can we apply?

Outline

- Example of working at multiple levels of resolution
- Consistency and validation
- Motivated metamodels, or the importance of a good story

An Example Use of MRM: Comparing Weapon Systems

Given a specified amount of funding, should we buy weapon system A or B or C?

To support a recommendation to buy system B, the analyst argues:

1. Scenarios 1, 2, 3, and 4 constitute an adequate test
2. System B is more combat effective than A or C in scenarios 1, 2, 3, and 4
3. System B has adequate non-combat performance (addressed in the paper, but not in this presentation)

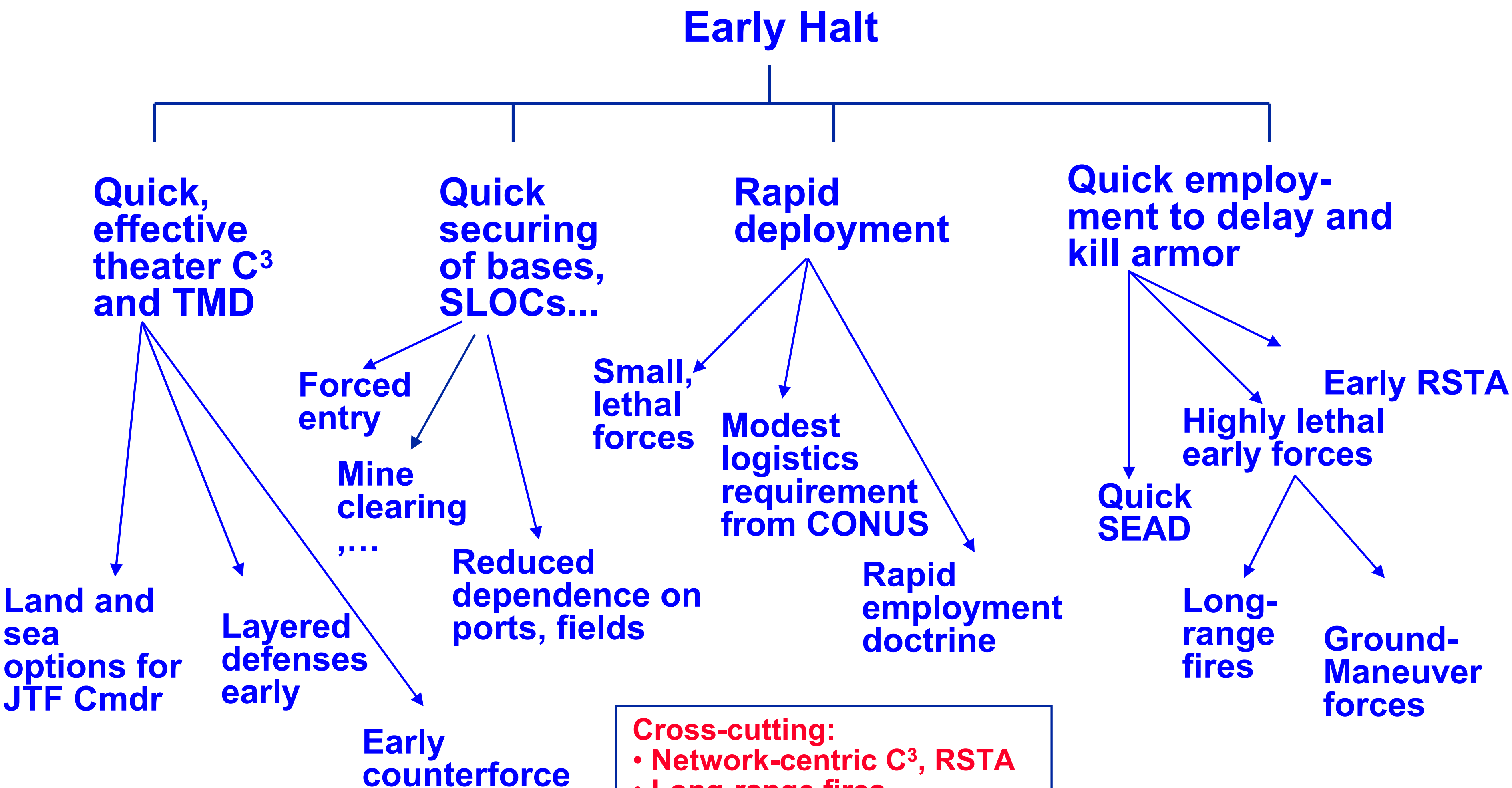
Recommended SecDef Operational Challenges

- *Early halt* of classic armored invasion given depth (e.g., in Kuwait or Northern Saudi Arabia)
- *Early shallow-halt* of fast parallel-operation invasion without depth (e.g., Korea) and with multifaceted opponent
- *Early offensive actions* without first building massive force
- Effective *low-risk* early intervention in “next Bosnia”
- Effective *low-risk* peacemaking in urban environments

Select Challenges Based on Analytic Importance

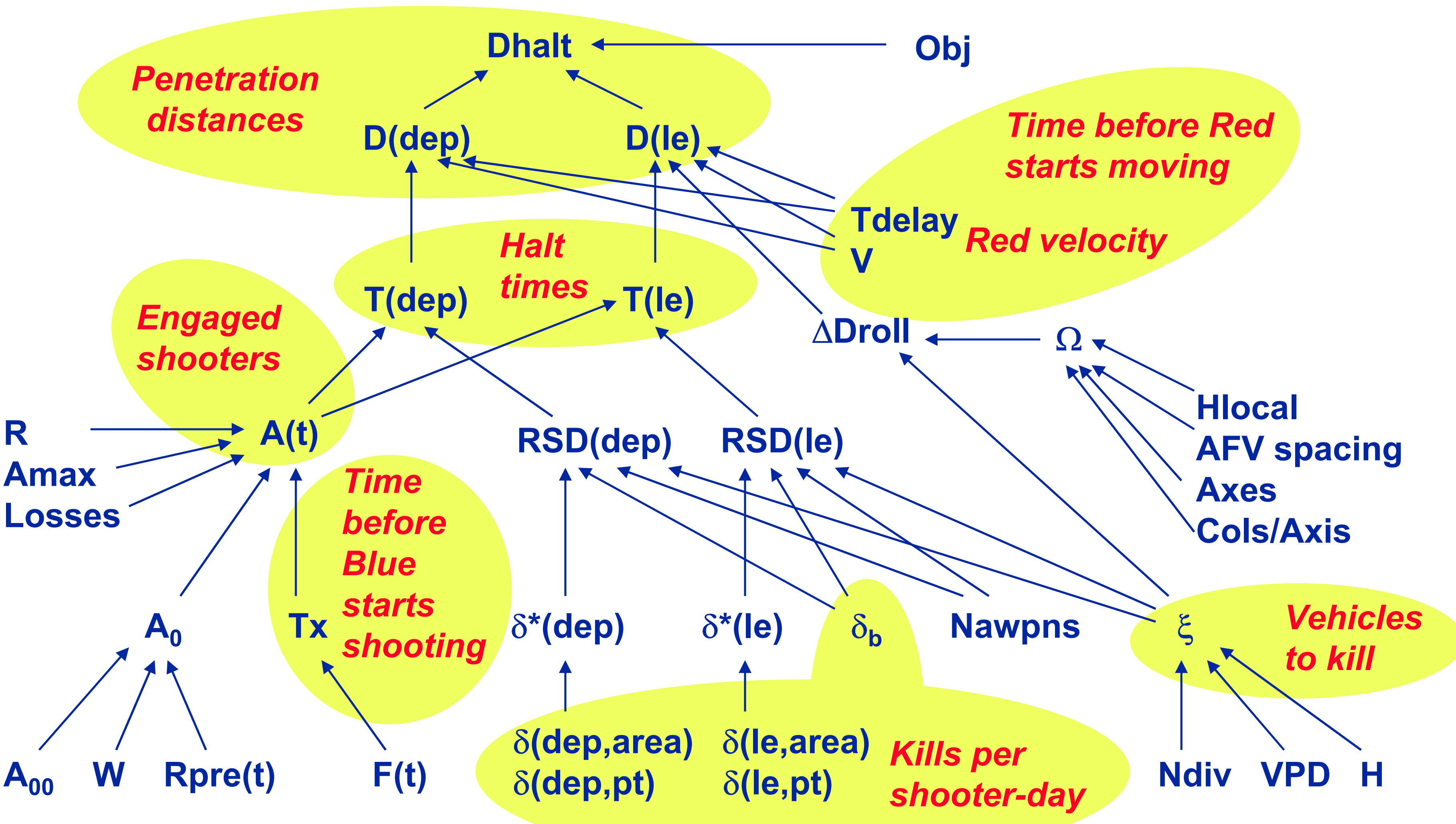
- Categorize challenges in three dimensions
  - Likelihood (low - high)
  - Consequence (small - great)
  - Difficulty (easy - hard)
- Choose hard challenges-but not too hard
- Balance likelihood versus consequence

Decompose Challenges Hierarchically

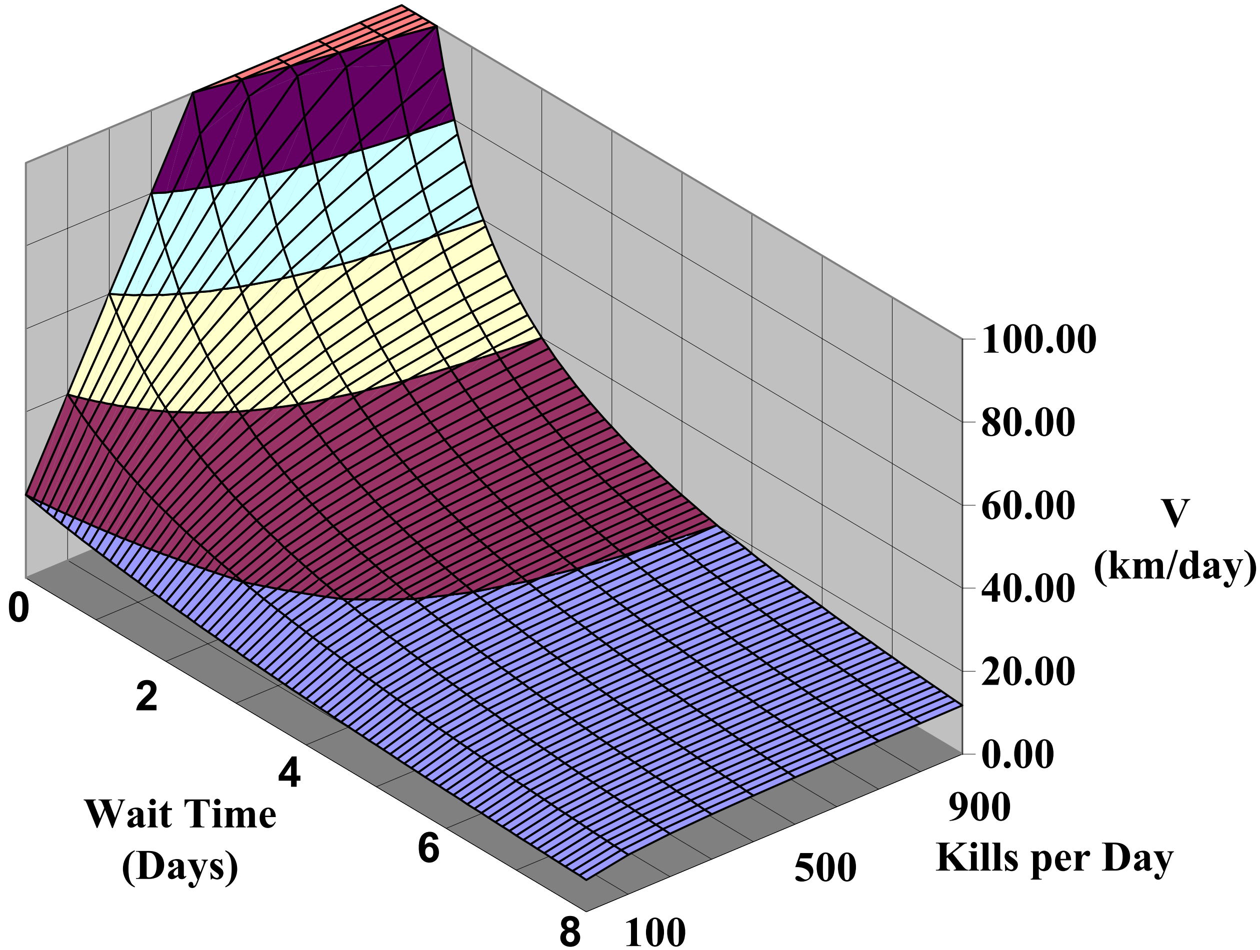


Cross-cutting:
- Network-centric C^3, RSTA
- Long-range fires
- Effective ops with allies
- Presence and forward-lean at warning
- Intra-theater mobility

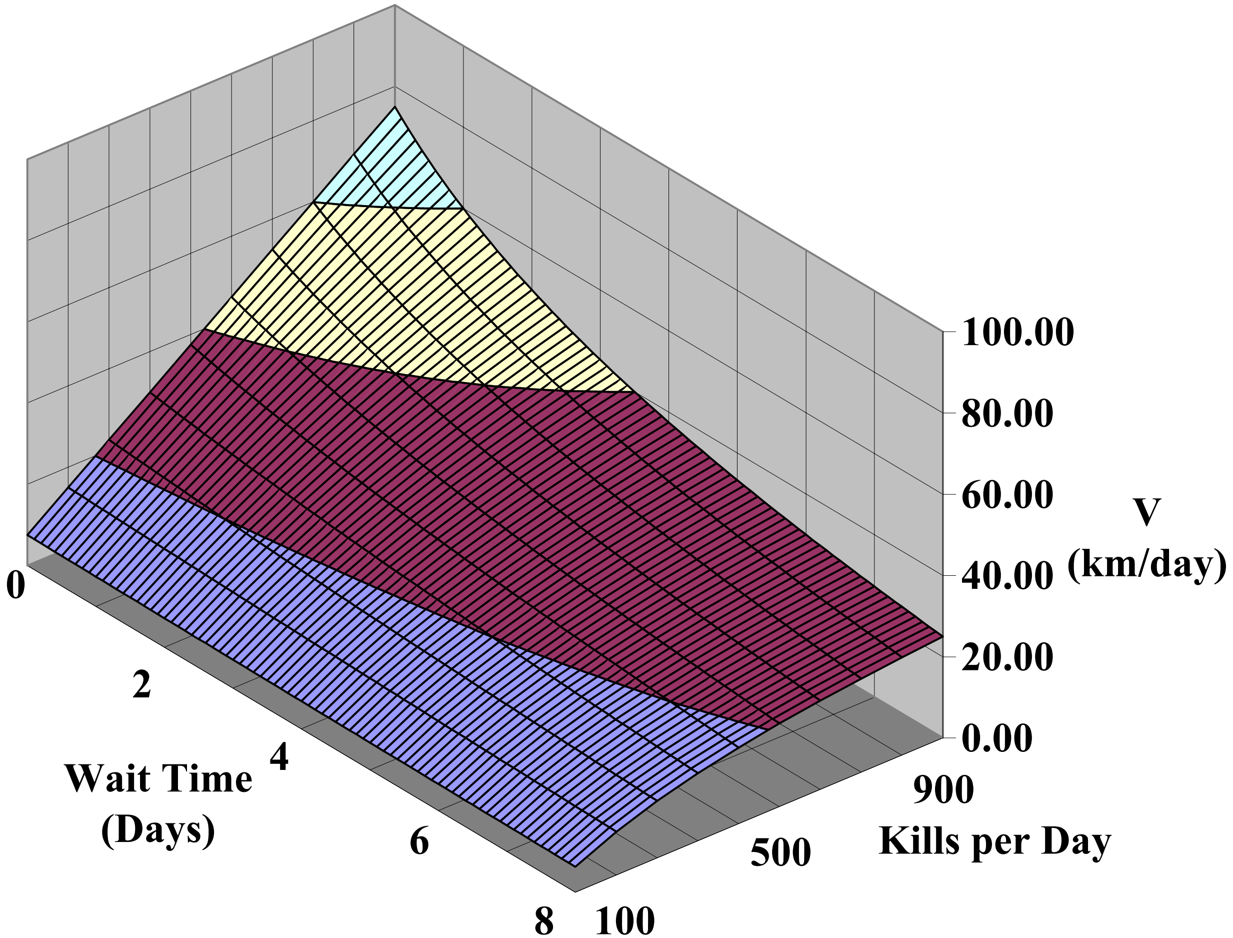
EXHALT-CF:
Use Long-Range Fires to Halt an Armored Column



Input Combinations that Kill 500 Armored Vehicles Within 100 km



Input Combinations the Kill 4000 Armored Vehicles Within 300 km



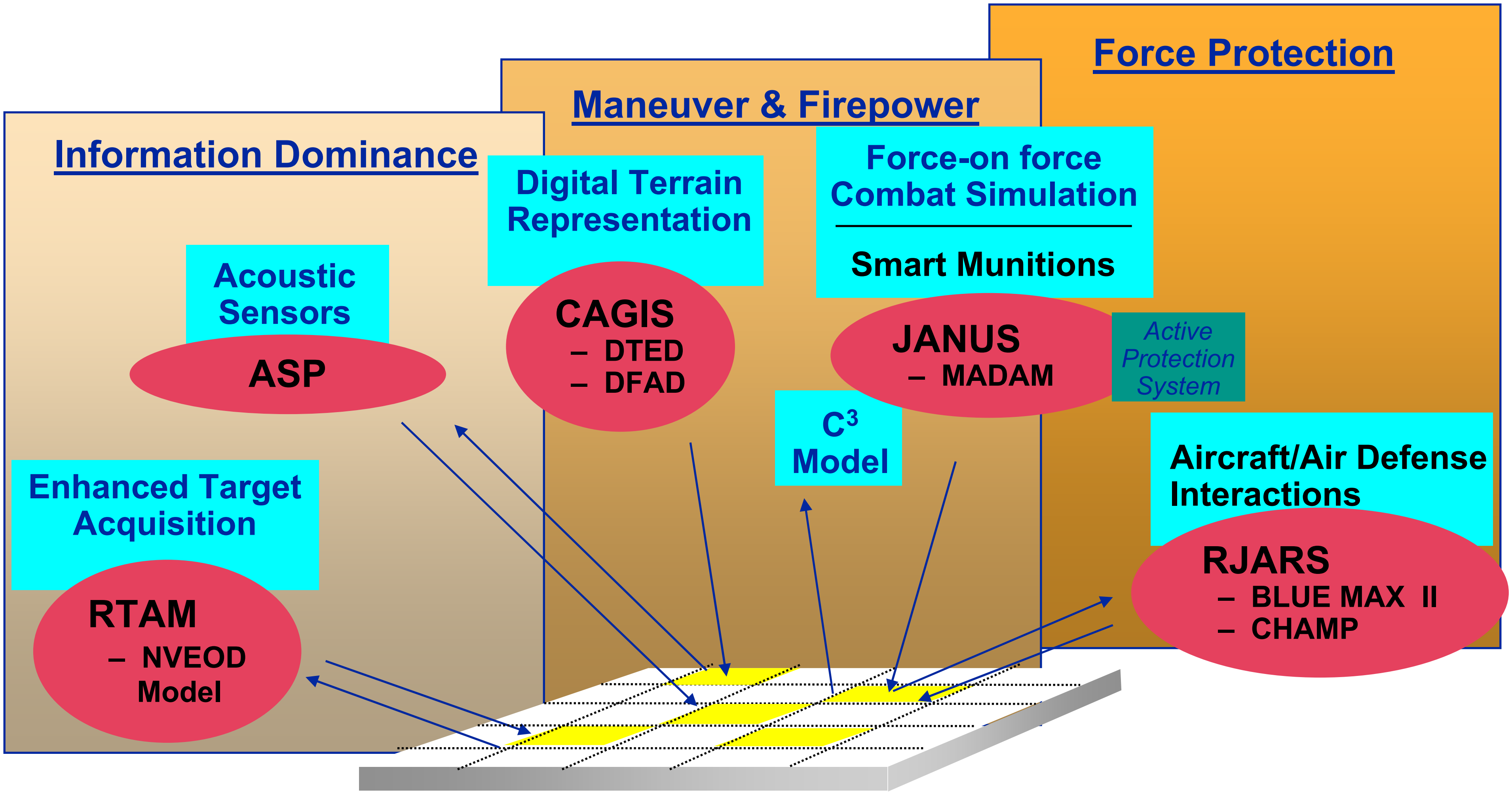
An Example Use of MRM: Comparing Weapon Systems

Given a specified amount of funding, should we buy weapon system A or B or C?

To support a recommendation to buy system B, the analyst argues:

1. **Scenarios 1, 2, 3, and 4 constitute an adequate test**
2. **System B is more combat effective than A or C in scenarios 1, 2, 3, and 4**
3. **System B has adequate non-combat performance (addressed in the paper, but not in this presentation)**

RAND's Force-On-Force Modeling Suite Provides A Unique Capability for High Fidelity Analysis



Use of JANUS Suite for Long Range Precision Fires

- JANUS simulates movement of each Red vehicle across the terrain (scripted)
- Periodic snapshots of Red vehicle positions are provided to a man-in-the-loop[1]
  - 5 minute intervals
  - See a vehicle with probability P_{TREE}, P_{OPEN}
- Man-in-the-loop[1] selects aim points, impact times for long range fires
  - Aim only at open areas
  - $T_{IMPACT} \geq T_{SNAPSHOT} + \Delta T_{DELAY}$
- MADAM simulates kills of Red vehicles
  - Monte Carlo, run 20 trials to obtain "stable" estimates

[1] For some weapons, the function of the man-in-the-loop has been automated.

Effectiveness of ATACMS/BAT Can Vary by an Order of Magnitude

1996 DSB Summer Study

- **3 kills per ATACMS/BAT**

1998 DSB Summer Study

- **0.35 kills per ATACMS/BAT**

Explanations and Plausible Counterarguments

1996 DSB Summer Study
- 3 kills per ATACMS/BAT
- No foliage
- 458 of 504 Red vehicles were AFVs
- 50+ AFVs/packet, 50-100m separation between AFVs

1998 DSB Summer Study
- 0.35 kills per ATACMS/BAT
- Tree cover
- 104 of 543 Red vehicles were AFVs
- 3-10 AFVs/pkt, 150-600m separation between AFVs

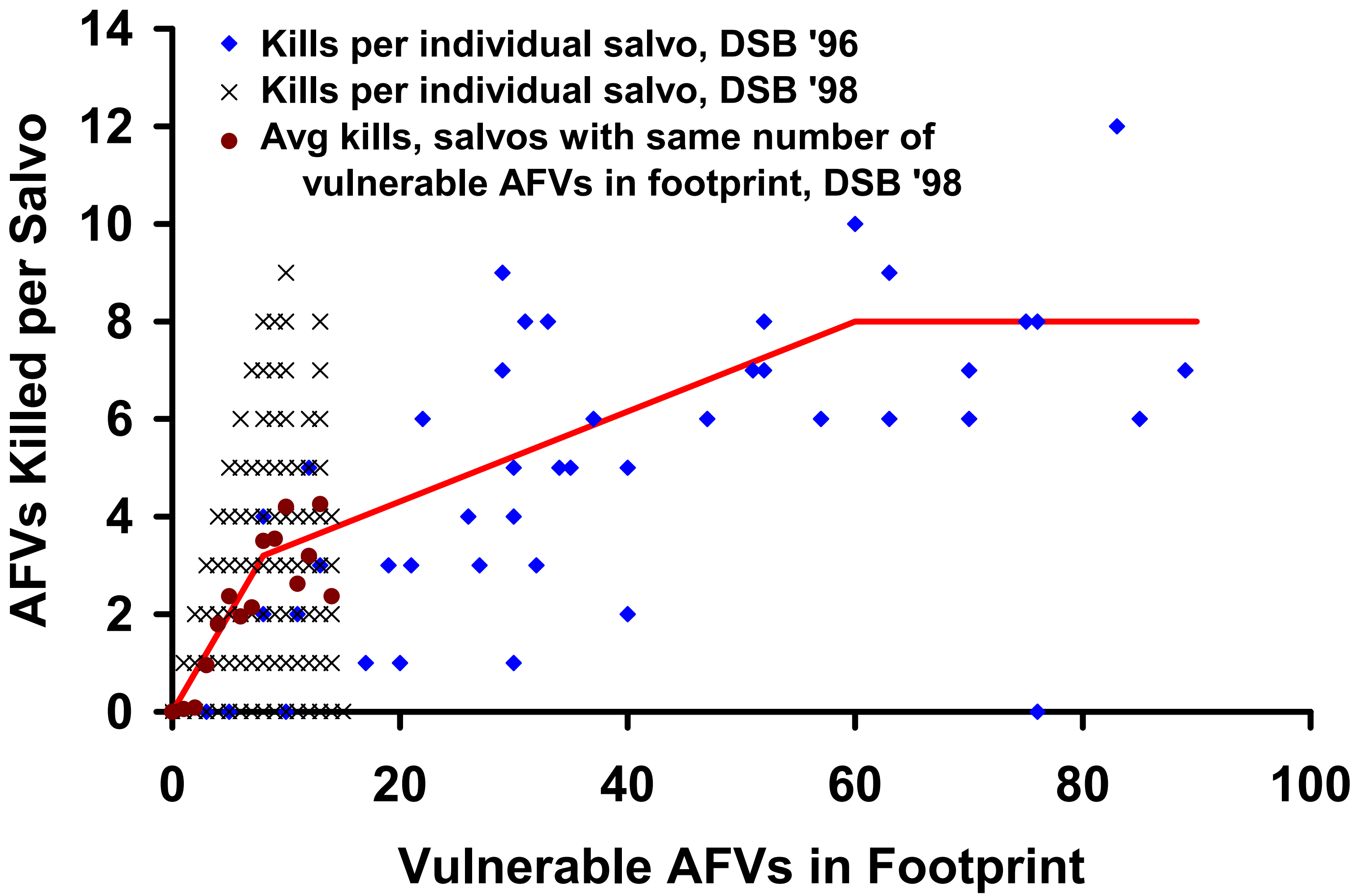
Man-in-the-loop aimed only at clearings

BAT submunitions only "hear" combat vehicles

ATACMS/BAT has a huge footprint

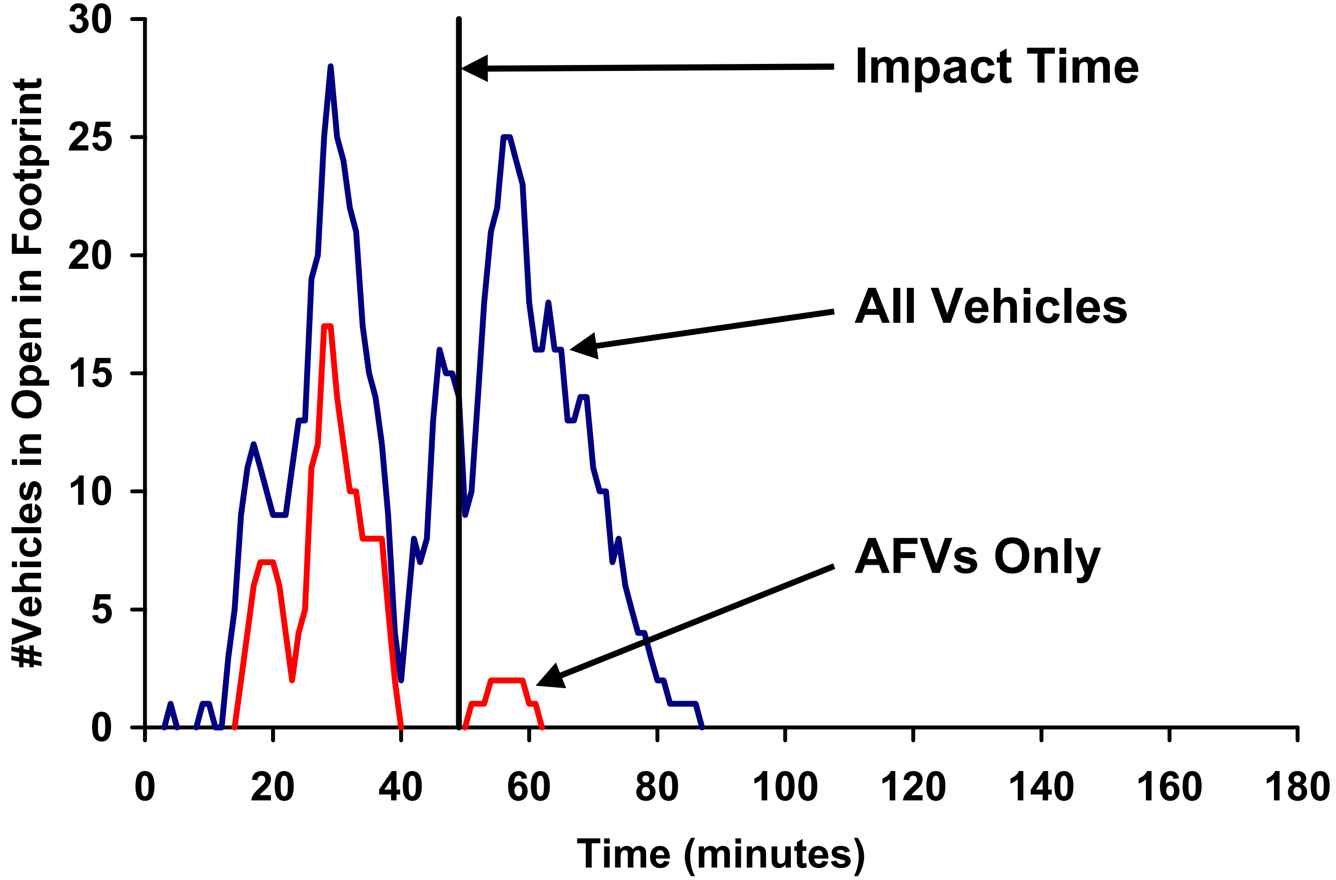
SO, IS IT THE MODEL, OR IS IT THE WEAPON???

Kills per Salvo Versus Vulnerable AFVs in Footprint



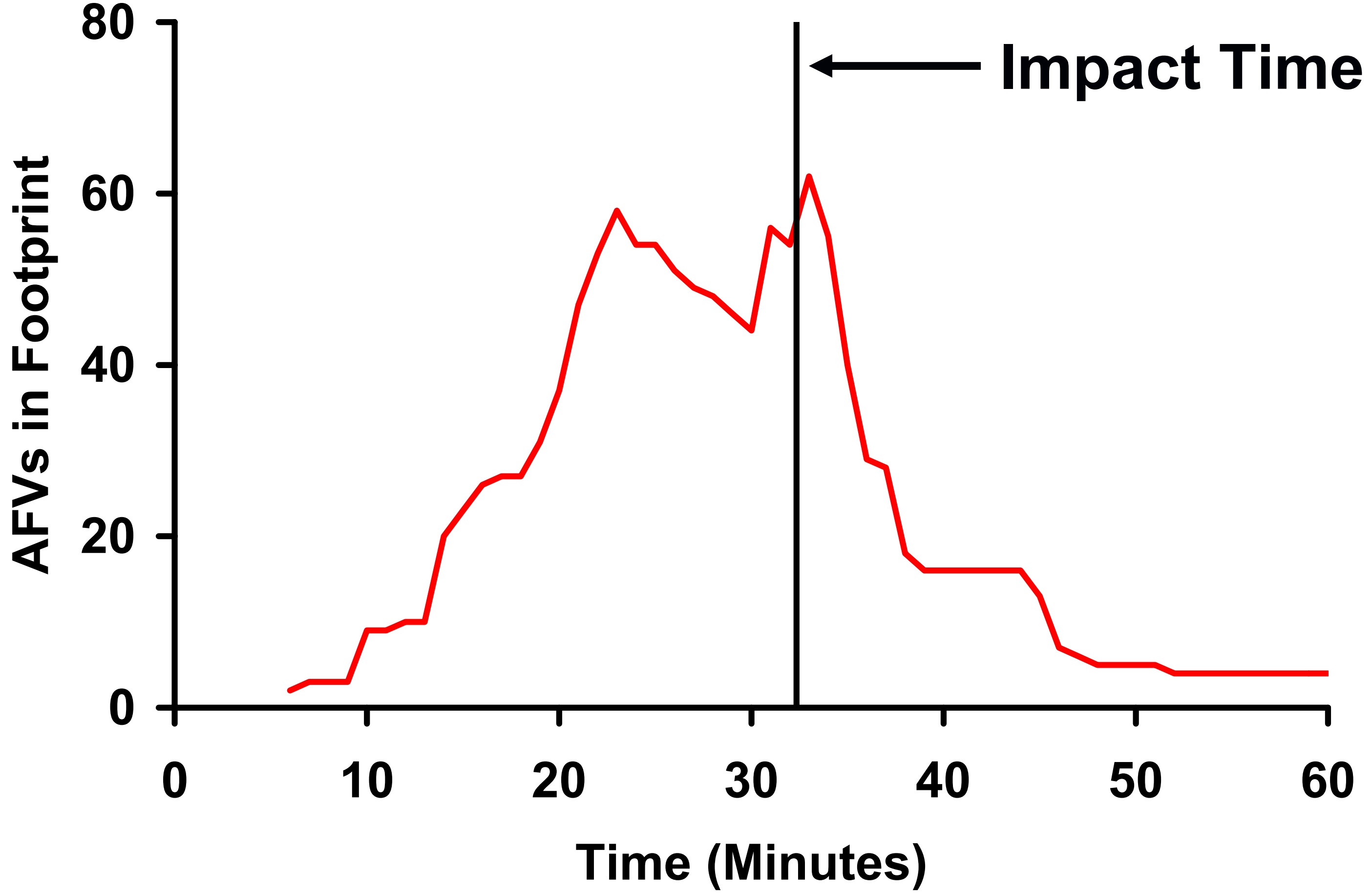
Against Sparse Red Formations, Can't Time Shots Well Enough to Target Specific Packets

Numbers of vulnerable AFVs in a typical footprint, DSB '98

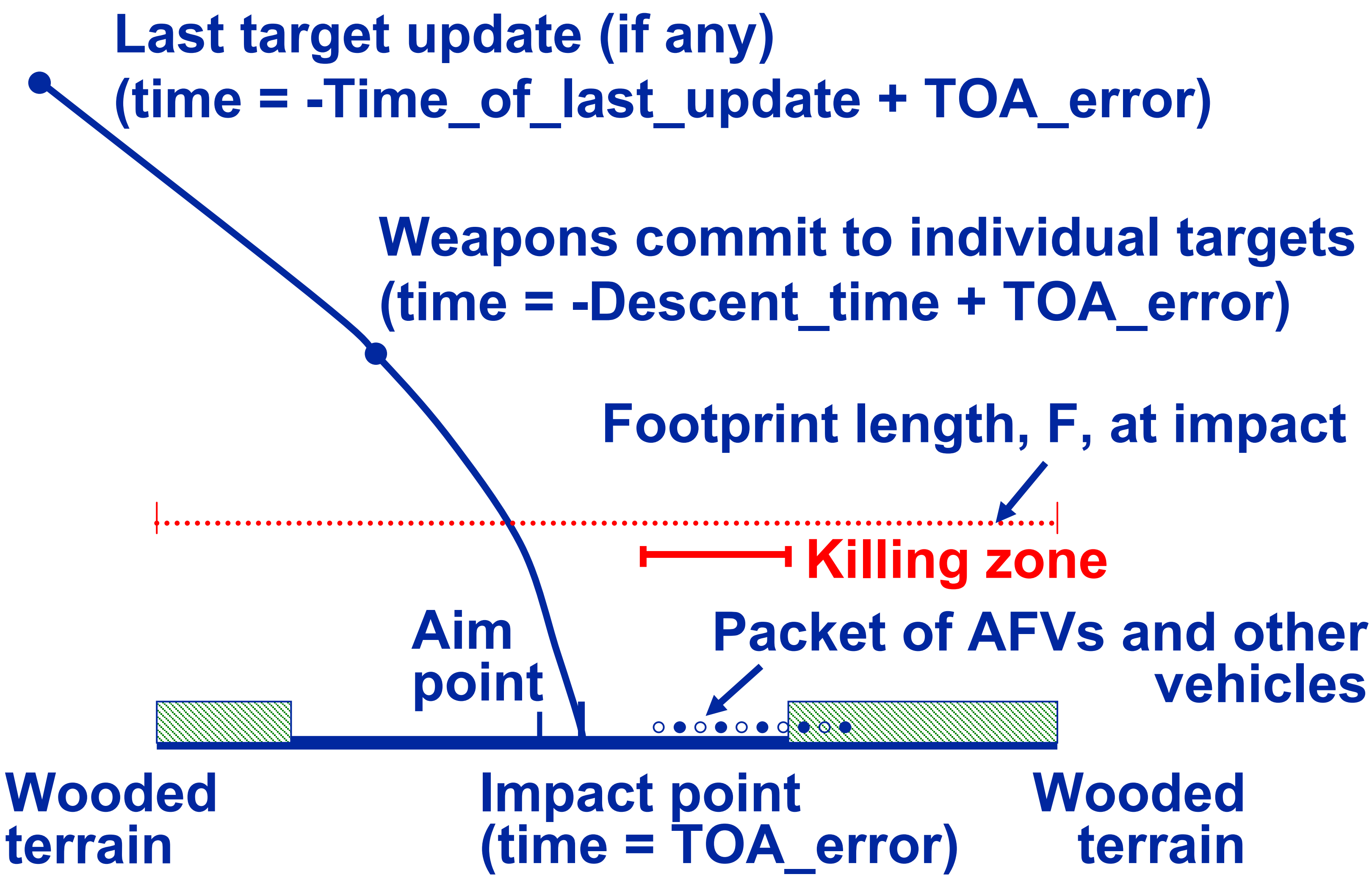


Precise Timing Unnecessary Against Large, Dense Red Formations

Numbers of vulnerable AFVs in a typical footprint, DSB '96

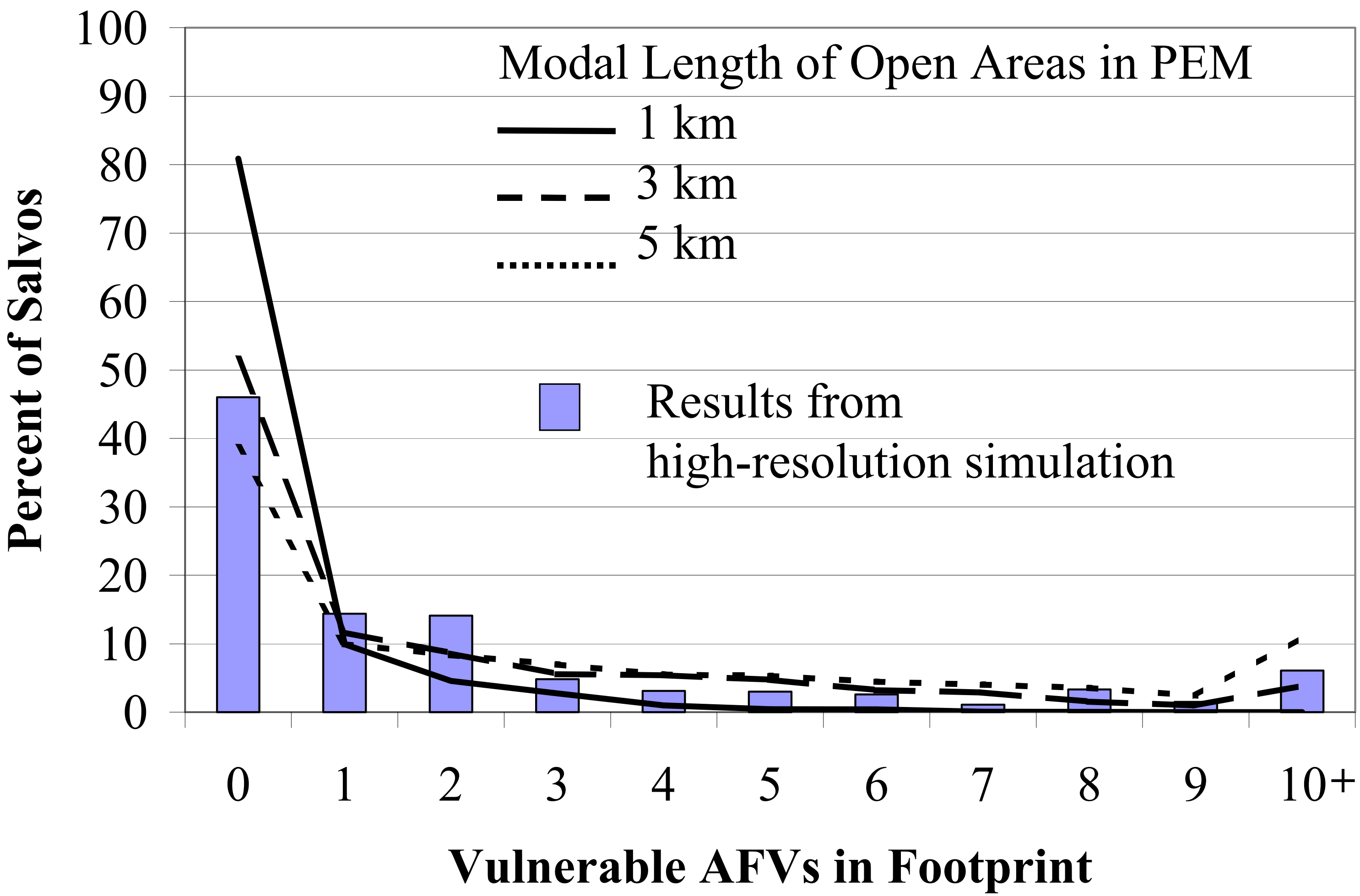


Estimating Number of Vulnerable AFVs in the Footprint



Most Footprints Contain Few Targets

Comparing Data from DSB ’98 with PEM estimates



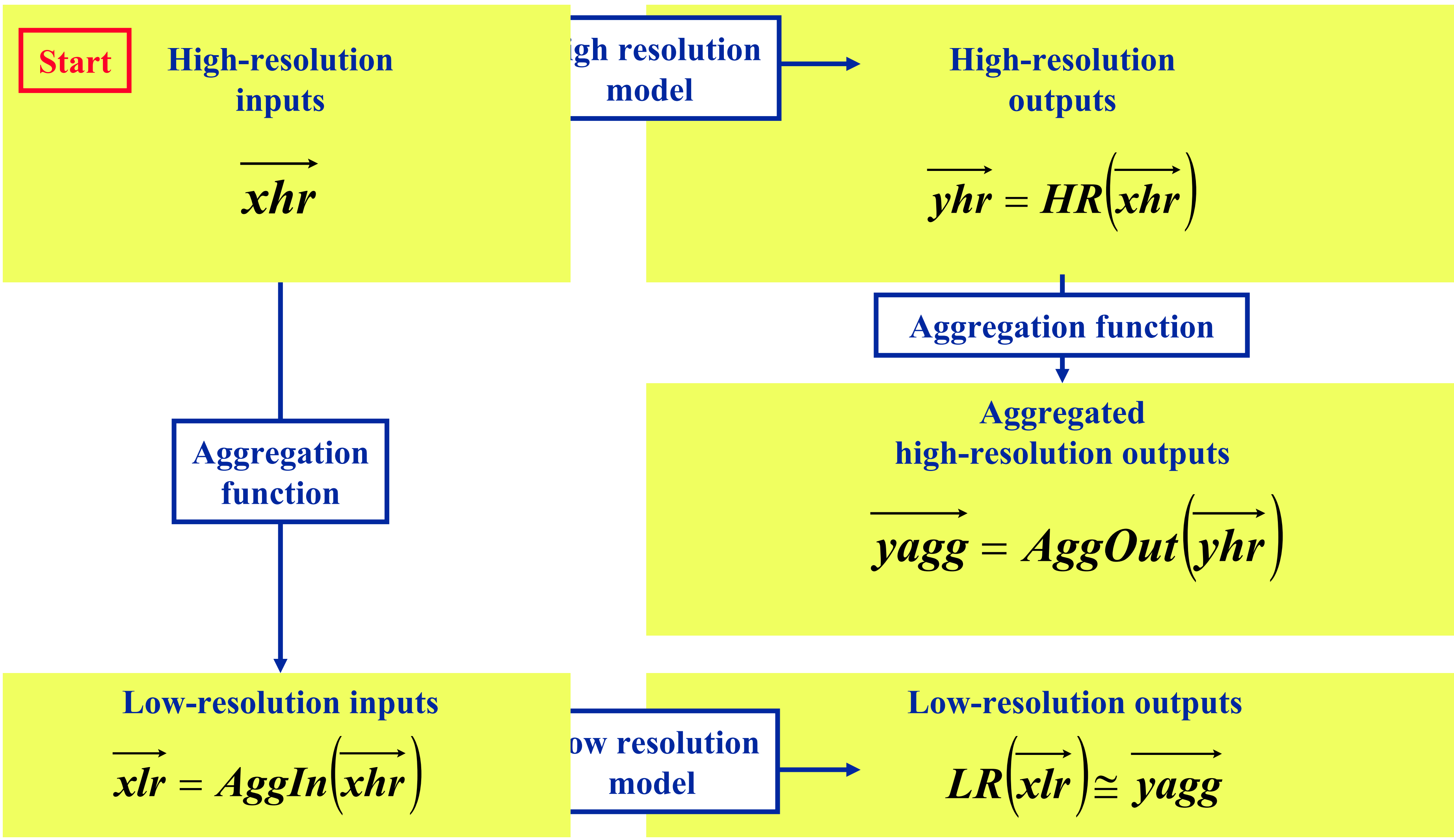
PGM Effectiveness Model (PEM)

- Low-resolution model
- Scales effectiveness of long range precision fires for influence of multiple factors
  - Weapon characteristics
  - Impact time error
  - Terrain
  - Density of Red formation
- Uses of PEM
  - Exploratory analysis
  - Subroutine to incorporate above factors in other models
  - Establish face validity of high-resolution model
  - Generalize high-resolution results

Outline

- Example of working at multiple levels of resolution
- Consistency and validation
- Motivated metamodels, or the importance of a good story

The Usual Definition of Consistency Between High- and Low-Resolution Models



A More General Definition of Consistency Between High- and Low-Resolution Models

Ensemble of high-resolution inputs

$$E_{in}\left(\overrightarrow{xlr}\right) = \left\{\overrightarrow{xhr} \,\middle|\, AggIn\left(\overrightarrow{xhr}\right) = \overrightarrow{xlr}\right\}$$

High resolution model →

Ensemble of high-resolution outputs

$$E_{out}\left(\overrightarrow{xlr}\right) = \bigcup_{\overrightarrow{xhr} \in E_{in}\left(\overrightarrow{xlr}\right)} \left\{\overrightarrow{yhr} \,\middle|\, \overrightarrow{yhr} = HR(\overrightarrow{xhr})\right\}$$

↓ **Aggregation function** ↓

Ensemble of aggregated high-resolution outputs

$$E_{agg}\left(\overrightarrow{xlr}\right) = \bigcup_{\overrightarrow{yhr} \in E_{out}\left(\overrightarrow{xlr}\right)} \left\{\overrightarrow{yagg} \,\middle|\, \overrightarrow{yagg} = AggOut\left(\overrightarrow{yhr}\right)\right\}$$

Low-resolution inputs

Start

$$\overrightarrow{xlr}$$

↑ **Disaggregation function** ↑

ow resolution model →

Low-resolution outputs

$$LR\left(\overrightarrow{xlr}\right) \cong WhatAttributes\left(E_{agg}\left(\overrightarrow{xlr}\right)\right)$$

Validity is Consistency Between a Model and the Real World

Ensemble of real-world observations

$$E_{obs}\left(\overrightarrow{xlr}\right) = \left\{\overrightarrow{vreal} \,\middle|\, AggIn\left(\overrightarrow{vreal}\right) = \overrightarrow{xlr}\right\}$$

Aggregation function

Ensemble of aggregated real-world observations

$$E_{agg}\left(\overrightarrow{xlr}\right) = \bigcup_{\overrightarrow{vreal} \in E_{obs}\left(\overrightarrow{xlr}\right)} \left\{\overrightarrow{yagg} \,\middle|\, \overrightarrow{yagg} = AggOut\left(\overrightarrow{vreal}\right)\right\}$$

Disaggregation function

Model inputs

Start

$$\overrightarrow{xlr}$$

Model

Model outputs

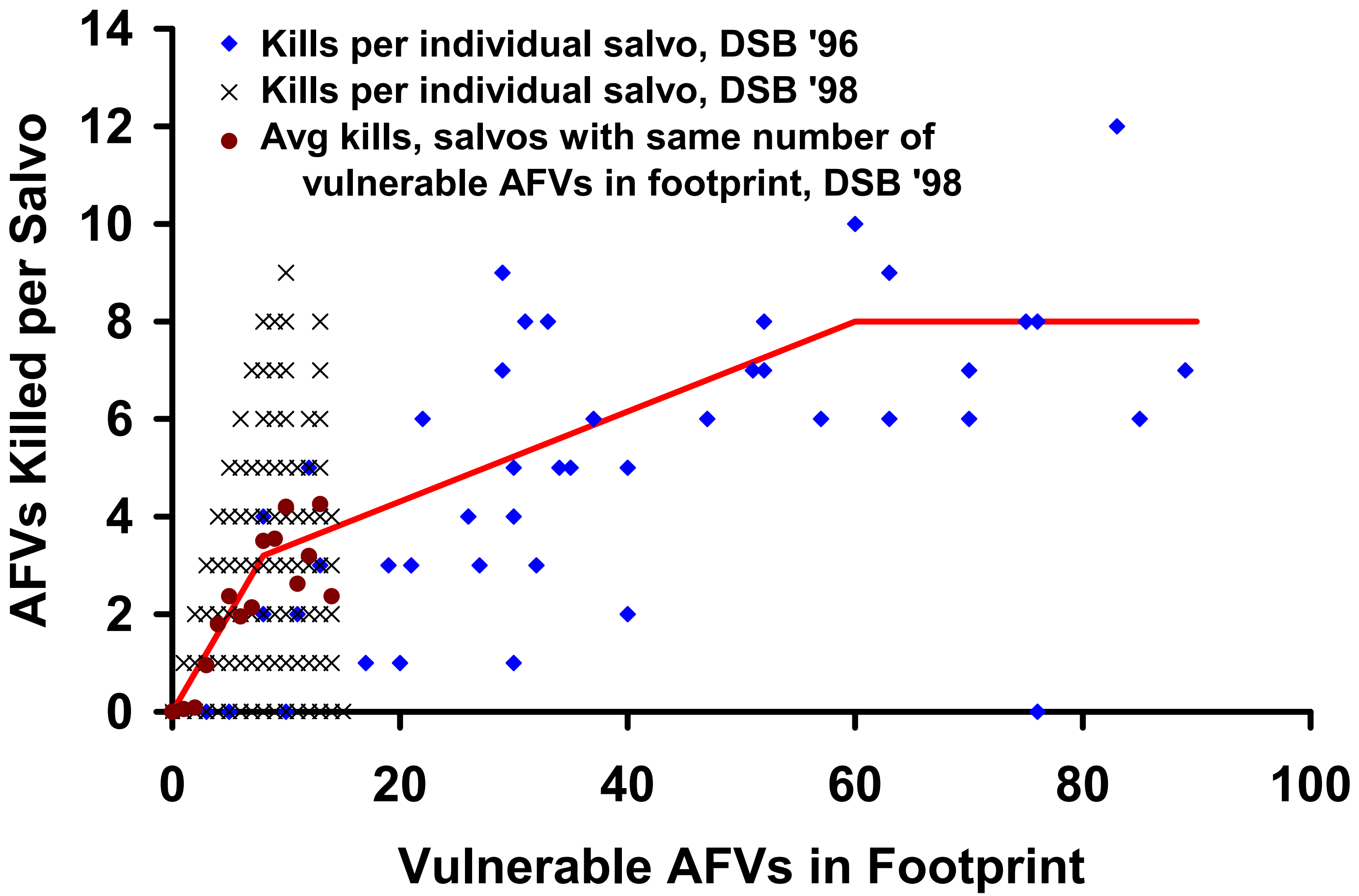
$$LR\left(\overrightarrow{xlr}\right) \cong WhatAttributes\left(E_{agg}\left(\overrightarrow{xlr}\right)\right)$$

What Attributes of $E_{agg}(\overrightarrow{xlr})$ Should the Low-Resolution Model Estimate?

- Assume $\overrightarrow{yagg}$ is a scalar
- If the high-resolution model is well-behaved, then $E_{agg}(\overrightarrow{xlr})$ is an interval
  - If it is a small interval, estimate the midpoint
  - If you intend to use an *a fortiori* argument, estimate the MAX or MIN
  - Otherwise, you must estimate something else
    - Range (MAX and MIN)
    - Confidence interval
    - Mean and standard deviation

  } Require a probability distribution on $\overrightarrow{xhr}$ (confidence interval; mean and standard deviation)
- If the high-resolution model is ill-behaved, e.g., chaotic, then $E_{agg}(xlr)$ need not be an interval

Kills per Salvo Versus Vulnerable AFVs in Footprint



Sources of Variation in Kills per Shot for Fixed # Vulnerable AFVs

- Random variation
  - JANUS suite has Monte Carlo elements
- Hidden variables – phenomena in the high-resolution model that aren't retained in PEM
  - Kinds of AFVs (tank, BMP, …)
  - Pattern (linear, "cross", …)
  - Interference (e.g., noise from vehicles not in footprint)
  - ???

Possible Bias from Using Average Kills per Salvo from This Sample

- Manipulation of hidden variables by Blue or Red
  - e.g., Red noisemakers as decoys
- Quantities that were held constant in the high-resolution cases we had, but could be varied in future cases

These issues make experimental design problematic.

Outline

- Example of working at multiple levels of resolution
- Consistency and validation
- Motivated metamodels, or the importance of a good story

Why a Good Story is Important
(or why a model shouldn't be merely a black box)

- Establish face validity
- Present, explain, and generalize results to client
- Extrapolate results
- Build a better metamodel

Extreme Approaches to Building a Metamodel

- Statistical
  - Build a data set of many cases from object model
  - Use statistical methods (e.g., regression) to analyze data
  - Disregard knowledge of the model's "innards"
- Phenomenological
  - Start with the most exact theory available
  - Simplify by aggregating, rearranging, combining terms
  - Use physical insights (e.g., conservation laws)
  - Disregard data generated by the object model

This experiment explored how to constructively combine the two approaches

Desiderata for a Good Metamodel

- Goodness of fit
- A plausible storyline connecting outcome to inputs
  - For cognitive purposes
  - To support extrapolation
- Parsimony
  - To permit exploratory analysis
- Identification of "critical components"
  - Sets of variables all of which must work to make the system as a whole perform adequately
  - Enter the metamodel as nonlinearities, e.g.,
    - Products
    - Thresholds

Illustration of Critical Components (using a "Halt Phase" example)

D = Distance at which Red column is halted (Outcome)
A = Number of Blue shooters
δ = Number of Red vehicles killed per shooter-day
T = Time Blue starts shooting
ξ = Number of Red vehicles to kill
V = Velocity of Red column

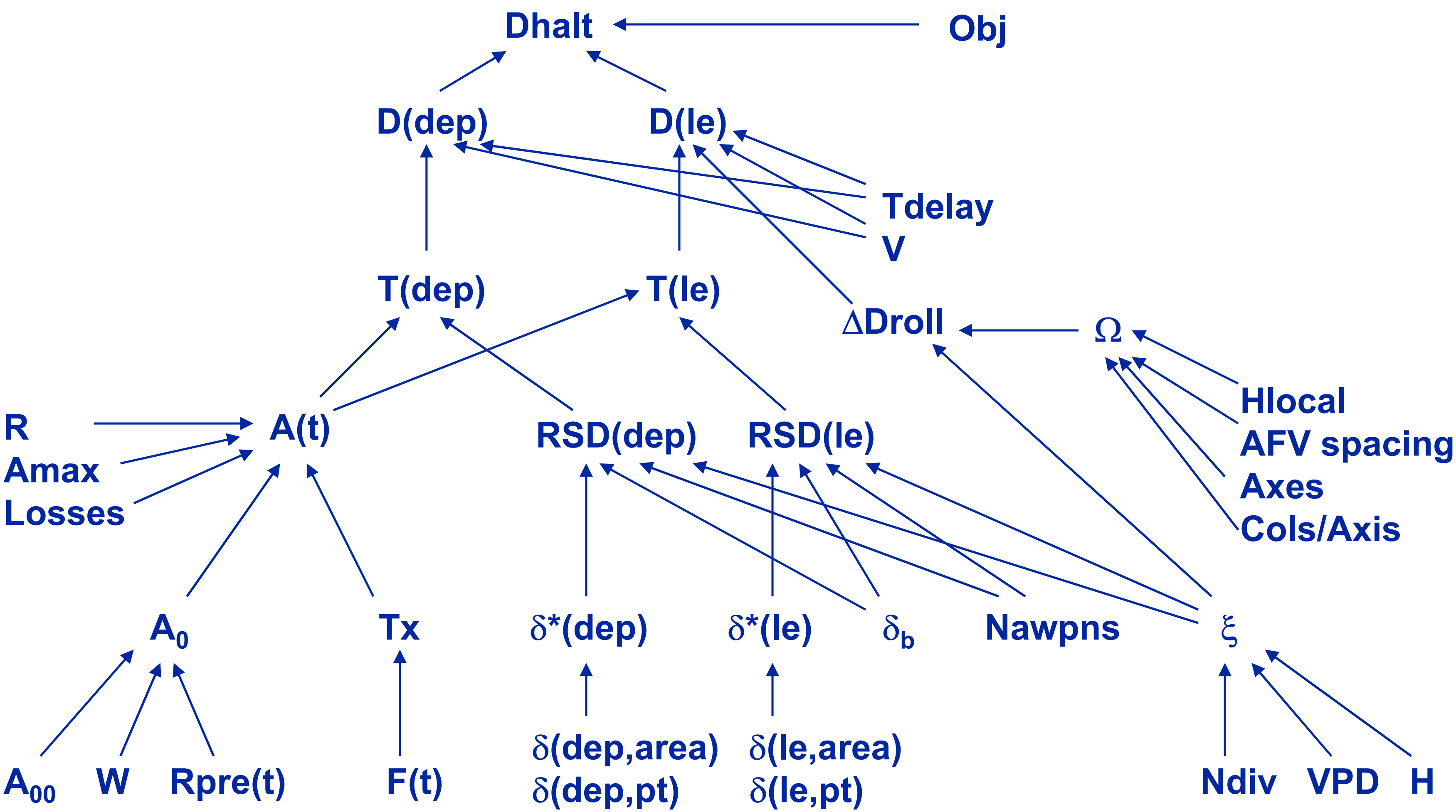
In a linear model, tradeoffs are fixed, universal

$$D = a_0 + a_1 A + a_2 \delta + a_3 T + a_4 \xi + a_5 V$$

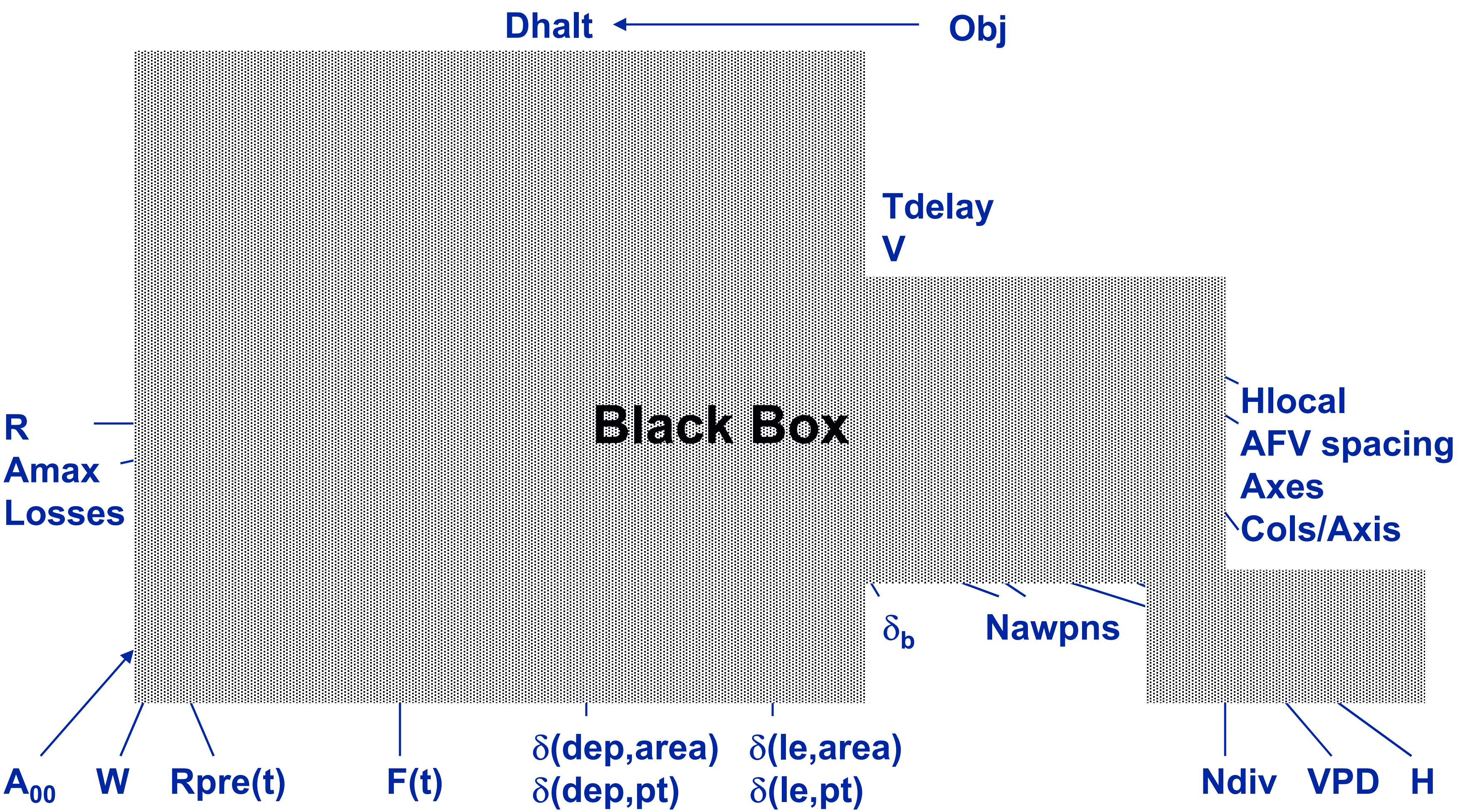
But not in a model based on (nonlinear) phenomenology

$$D = V \times \left(T + \frac{\xi}{A\delta} \right)$$

EXHALT-CF:
Use Long-Range Fires to Halt an Armored Column



Metamodel 1: Regress Dhalt on Lowest Level Inputs



Scorecard for Metamodel 1

	Model 1
RMSE (km)	140
Storyline	No
# Calib Const	15
# Aggreg Vars	14
Crit Comps	No

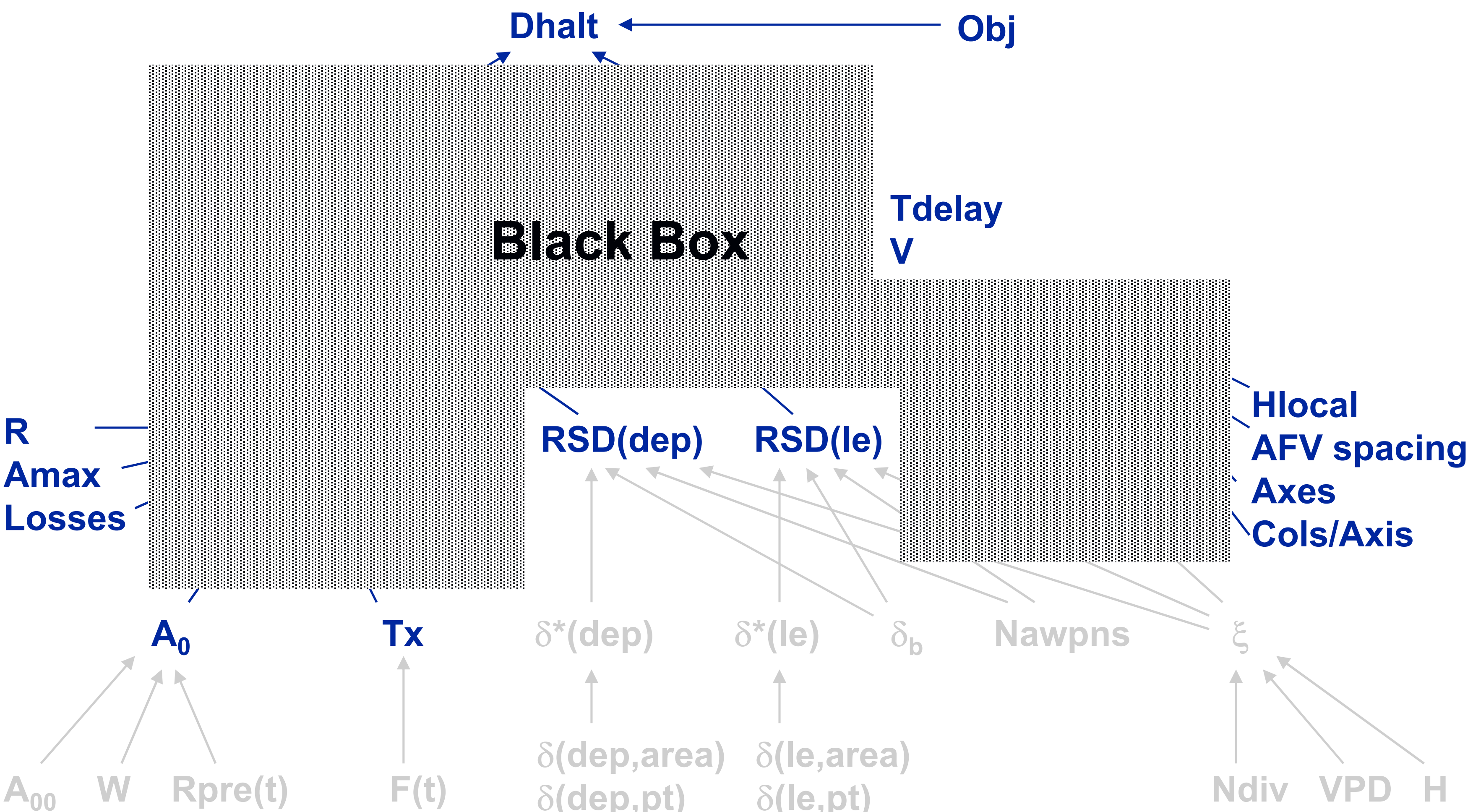
Fit linear (in undetermined coefficients) model *L* to cases where:

$$0 < Dhalt < Obj$$

Calculate metamodel *M* as:

$$M = Max\{\ 0,\ Min\{\ L,\ Obj\ \}\}$$

Metamodel #2: Regress Dhalt on Intermediate Variables



Scorecard for Metamodels 1 & 2

	Model 1	Model 2
RMSE (km)	140	84
Storyline	No	Fragments
# Calib Const	15	11
# Aggreg Vars	14	10
Crit Comps	No	No

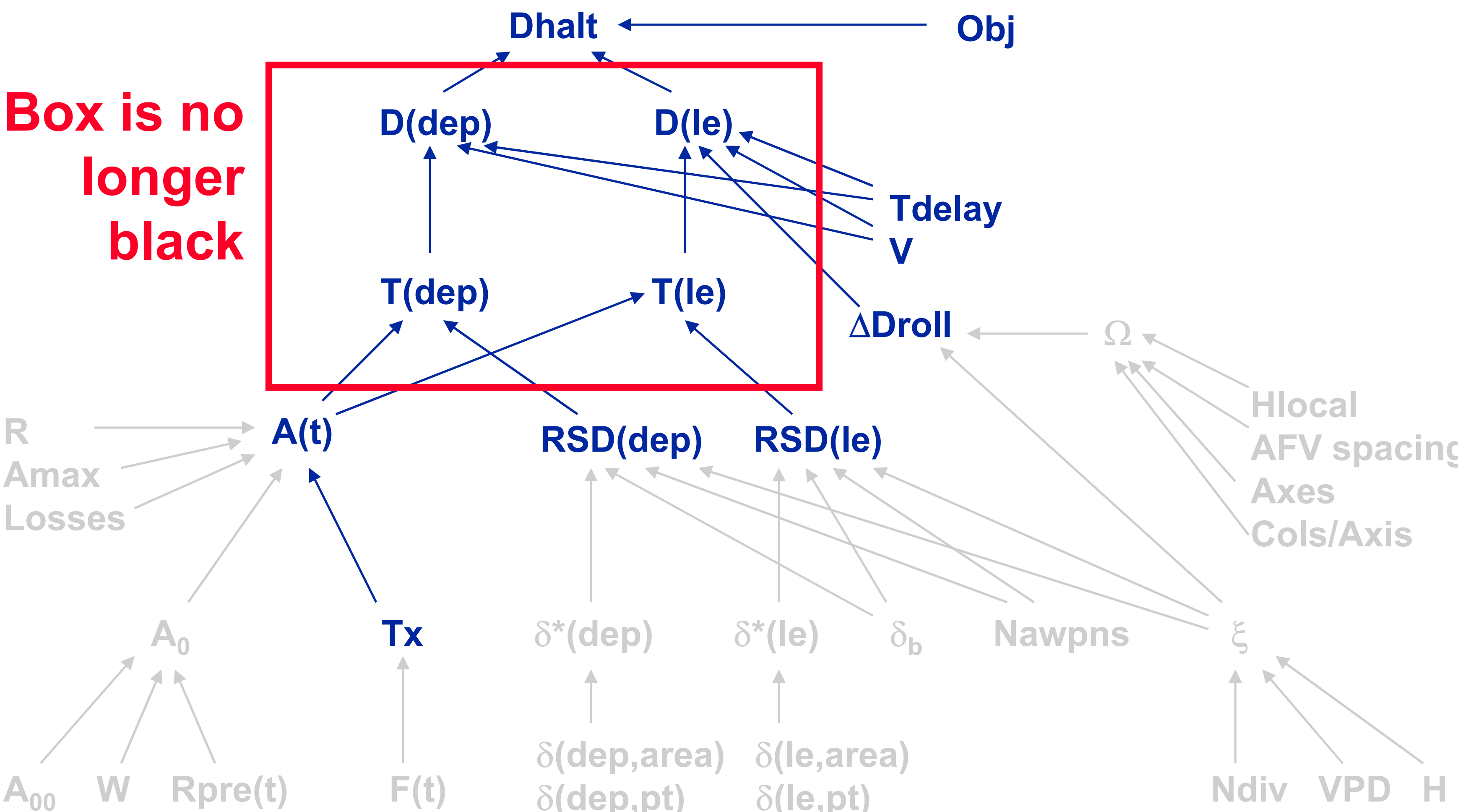
Fit linear model *L* to cases where:

$$0 < Dhalt < Obj$$

Calculate metamodel *M* as:

$$M = Max\{\, 0, Min\{\, L, Obj \,\}\}$$

Metamodel 3 & 4: Derive Form from a Story, Calibrate with Statistics



The Storyline for Metamodels 3 and 4

Blue chooses the strategy with the smaller halt distance:

$$Dhalt = Max\{0, Min\{ D(dep), D(le), Obj \}\}$$

Against "In Depth" strategy, Red has constant velocity once he starts:

$$D(dep) = V \times (T(dep) - Tdelay)$$

"Leading Edge" strategy rolls Red back:

$$D(le) = V \times (T(le) - Tdelay) - \Delta Droll$$

Estimating Halt Times for Metamodel 3

T(*dep*) and *T*(*le*) should satisfy

$$\overline{A} \times \left(T(dep) - Tx\right) = RSD\left(dep\right), \; \overline{A} \times \left(T(le) - Tx\right) = RSD\left(le\right)$$

(Average number of shooters multiplied by time spent shooting equals shooter-days required to halt Red.)

Use some convenient approximation to $\overline{A}$

The Form of Metamodel 3

Use regression to determine coefficients in:

$$D(dep) = c_0 - c_1 VTdelay + c_2 VTx + c_3 V \frac{RSD(dep)}{\overline{A}}$$

$$D(le) = c_4 - c_5 VTdelay + c_6 VTx + c_7 V \frac{RSD(le)}{\overline{A}} - c_8 \Delta Droll$$

Note "critical component." To halt a fast Red (large *V*) in a short distance (small *D*), Blue must:

- **Reduce *Tx* (response time); *and***
- **Decrease *RSD* (e.g., better weapons); *and***
- **Increase $\overline{A}$ (number of shooters)**

Scorecard for Metamodels 1-3

	Model 1	Model 2	Model 3
RMSE (km)	140	84	30
Storyline	No	Fragments	Yes
# Calib Coeffs	15	11	9
# Aggreg Vars	14	10	5
Crit Comps	No	No	Yes

Fit linear models $L(dep)$ and $L(le)$ to cases where:

$$0 < D(dep) < Obj, \quad 0 < D(le) < Obj$$

Calculate metamodel M as:

$$M = Max\{\ 0,\ Min\{\ L(dep),\ L(le),\ Obj\ \}\}$$

The Form of Metamodel 4

Solve these integral equation exactly:

$$\int_{Tx}^{T(dep)} A(t)dt = RSD(dep), \quad \int_{Tx}^{T(le)} A(t)dt = RSD(le)$$

Then use regression to determine coefficients for:

$$D(dep) = c_0 - c_1 VTdelay + c_2 VTx + c_3 V[T(dep) - Tx]$$

$$D(le) = c_4 - c_5 VTdelay + c_6 VTx + c_7 V[T(le) - Tx] - c_8 \Delta Droll$$

Scorecard for Metamodels 1-4

	Model 1	Model 2	Model 3	Model 4
RMSE (km)	140	84	30	8
Storyline	No	Fragments	Yes	Yes, but
# Calib Coeffs	15	11	9	9
# Aggreg Vars	14	10	5	5
Crit Comps	No	No	Yes	Yes

Fit linear models *L*(*dep*) and *L*(*le*) to cases where:

$$0 < D(dep) < Obj,\ 0 < D(le) < Obj$$

Calculate metamodel *M* as:

$$M = Max\{\ 0,\ Min\{\ L(dep),\ L(le),\ Obj\ \}\}$$

Summary of Motivated Metamodeling Experiment

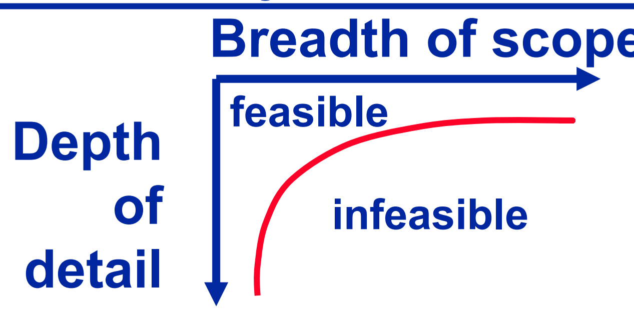
- Statistics uninformed by phenomenology yielded a poor metamodel
  - EXHALT-CF is not linear, and it is not smooth
  - Strategy choice (“In Depth” vs. “Leading Edge”) introduces a “kink”
- Adding phenomenology from bottom up yielded a better but still mediocre metamodel
  - Built in meaningful combinations of low-level variables
  - Inputs still related to outcome through a “black box”
- Adding phenomenology from top down yielded good to very good metamodels
  - Story (theory, if you prefer) now connects inputs to outcome -- no more “black box”

Wrap-up:
MRM is a Fruitful Response to:

- **Cognitive needs**
  - Understanding
  - Explaining to colleagues
  - Explaining to clients

A story (theory) made formal becomes a metamodel, validated by tests for consistency

- **Need for breadth AND depth**
  - MRM ties models at different levels into a consistent package



- **Massive uncertainty**
  - Unbridled extrapolation
  - Exploratory analysis
  - Need models with few input variables

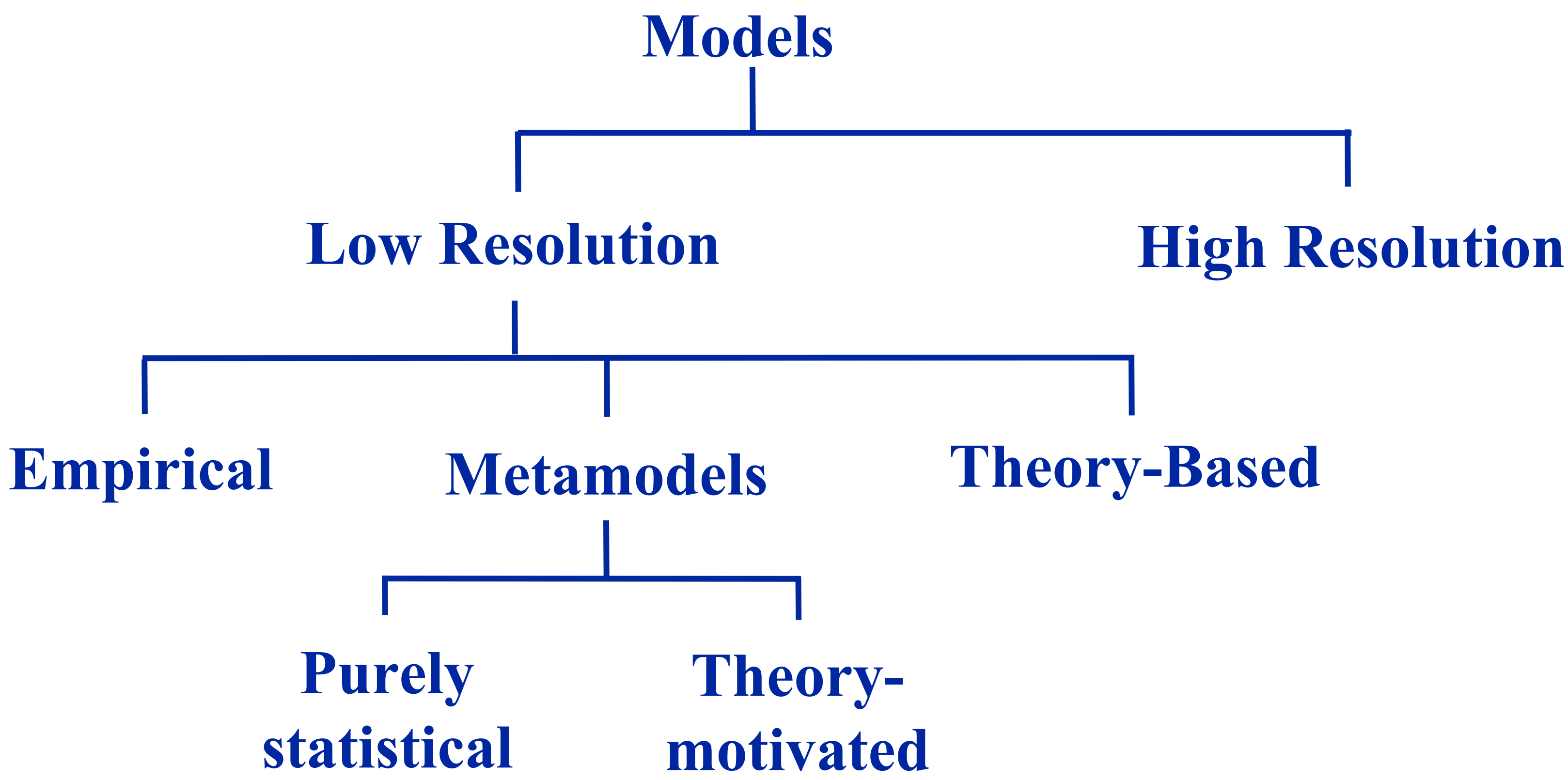
Need a different view of validation
- Model structure correct
- Domain includes cases of interest

Backups

The What and Why of Metamodels

- What is a metamodel?
  - A relatively small, simple model that mimics selected behaviors of a large, complex model
- Some reasons to build a metamodel
  - Cognitive: You need a plausible story to explain why the large model behaves as it does
  - Exploratory analysis: A small model has the analytic agility for exploring behavior over a vast input domain
  - Parsimony: Having a small number of independent variables is so important that we call it out separately

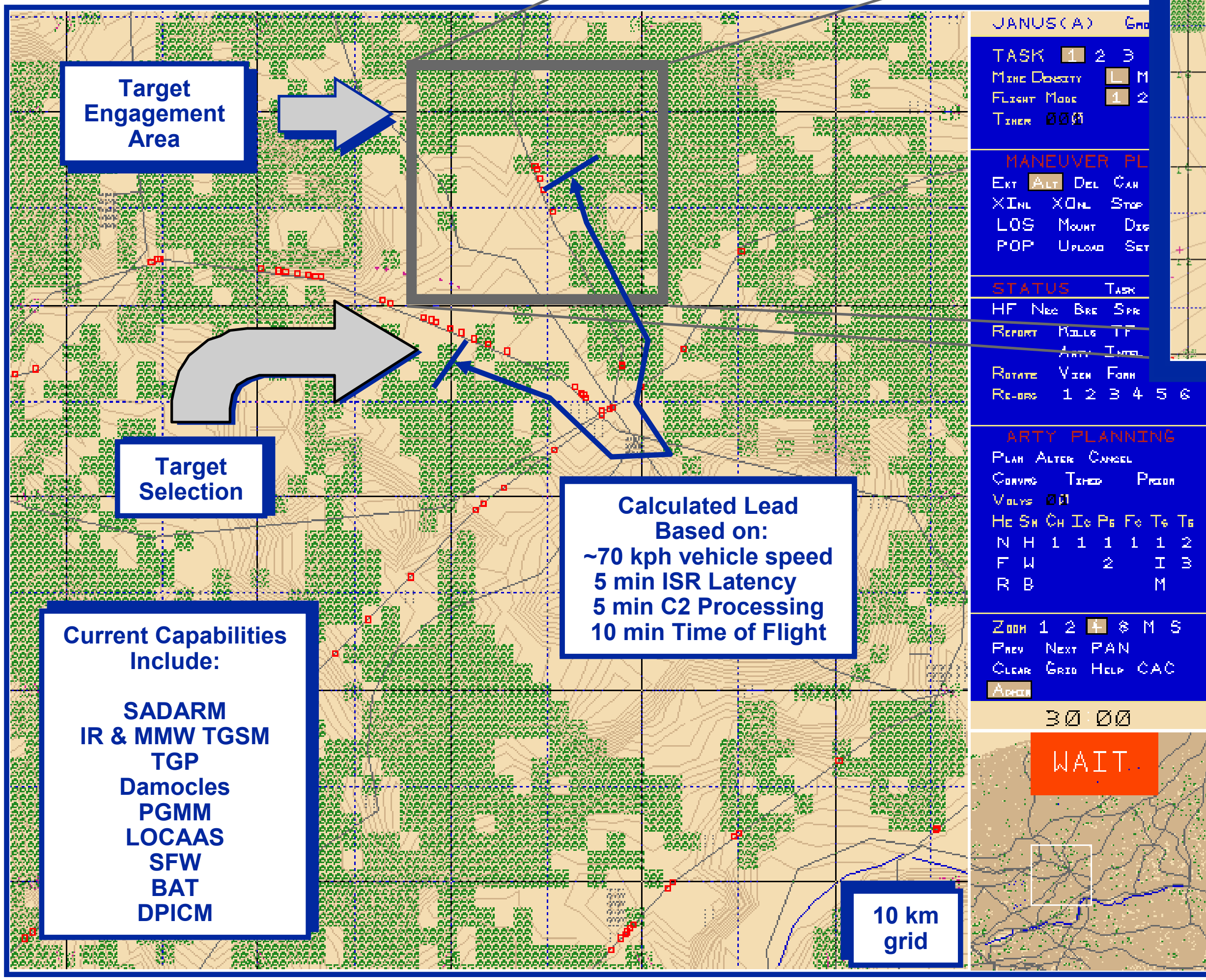
Taxonomy of Models

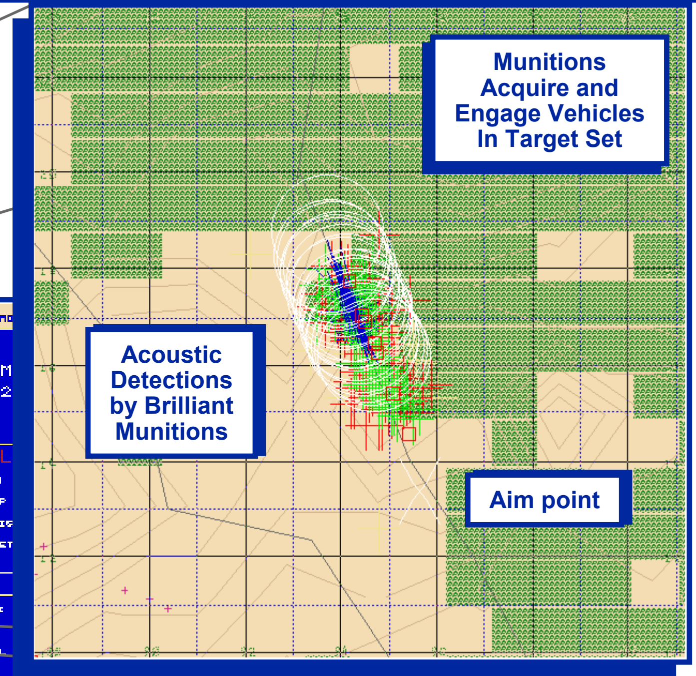


Importance of MRM

- Cognitive needs
  - Use high-resolution models to investigate underlying phenomena
  - Use low-resolution models for analysis of choice, and for forest-vs.-trees reasons
- Explanatory power
- Uncertainty, ignorance, and chaos
  - Detail is not the same as knowledge
- Exploratory analysis, analytical agility
  - Economy

Man-in-the-loop selects target and leads it; MADAM Simulates Kills





Lessons About Long Range Precision Fires
(from Exploratory Analysis with PEM)

- Effectiveness of long-range precision fires varies by a factor of 50-100, depending on:
  - Time_of_last_update (TOA_error)
  - Weapon characteristics (Footprint)
  - Terrain (Clearing widths)
  - Density of Red column
- Small clearings nullify value of a large footprint, but enhance the value of a small TOA_error
- High Red density nullifies importance of TOA_error
- A large footprint, high TOA_error weapon (ATACMS/BAT) is more effective than a small footprint, low TOA_error weapon (aircraft-delivered SFW) <u>only if</u> both Red density and clearing width are sufficiently large
  - I.e., if on average ATACMS/BAT finds enough Red vehicles in killing zone

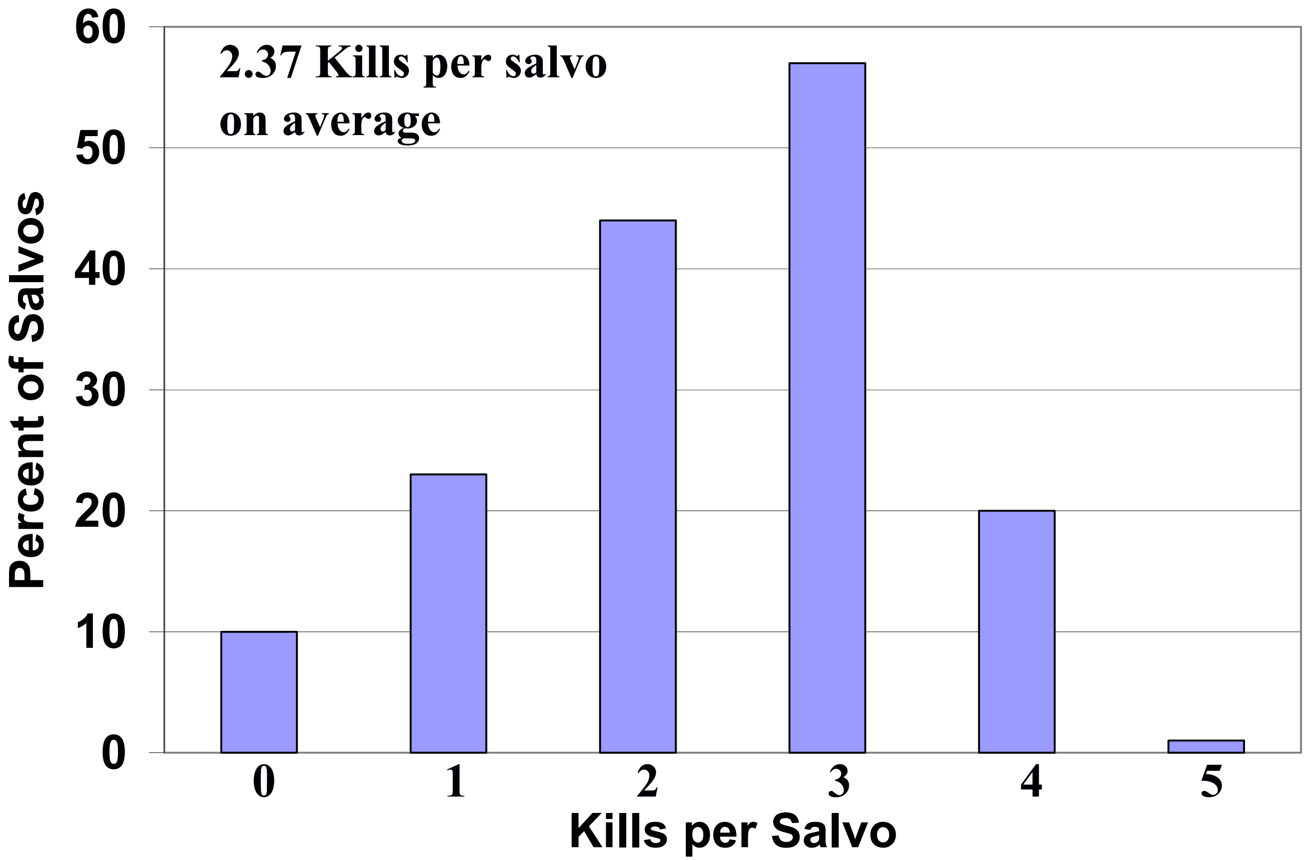
Conclusions are quantitative, not just qualitative

Estimating Shots per Day: A Loose End

- Objective was to estimate kills per day by ATACMS/BAT
- We have:
  - Kills/salvo = f(# vulnerable AFVs)
  - # vulnerable AFVs = g(TOA_error, footprint, terrain, Red_formation)
- We still need:
  - # shots per day by TOA_error, footprint, terrain, and Red_formation
- Don't know how to get it from high-resolution simulation, so assume a number
  - Based on shot opportunities?
  - Or on capacity to shoot?

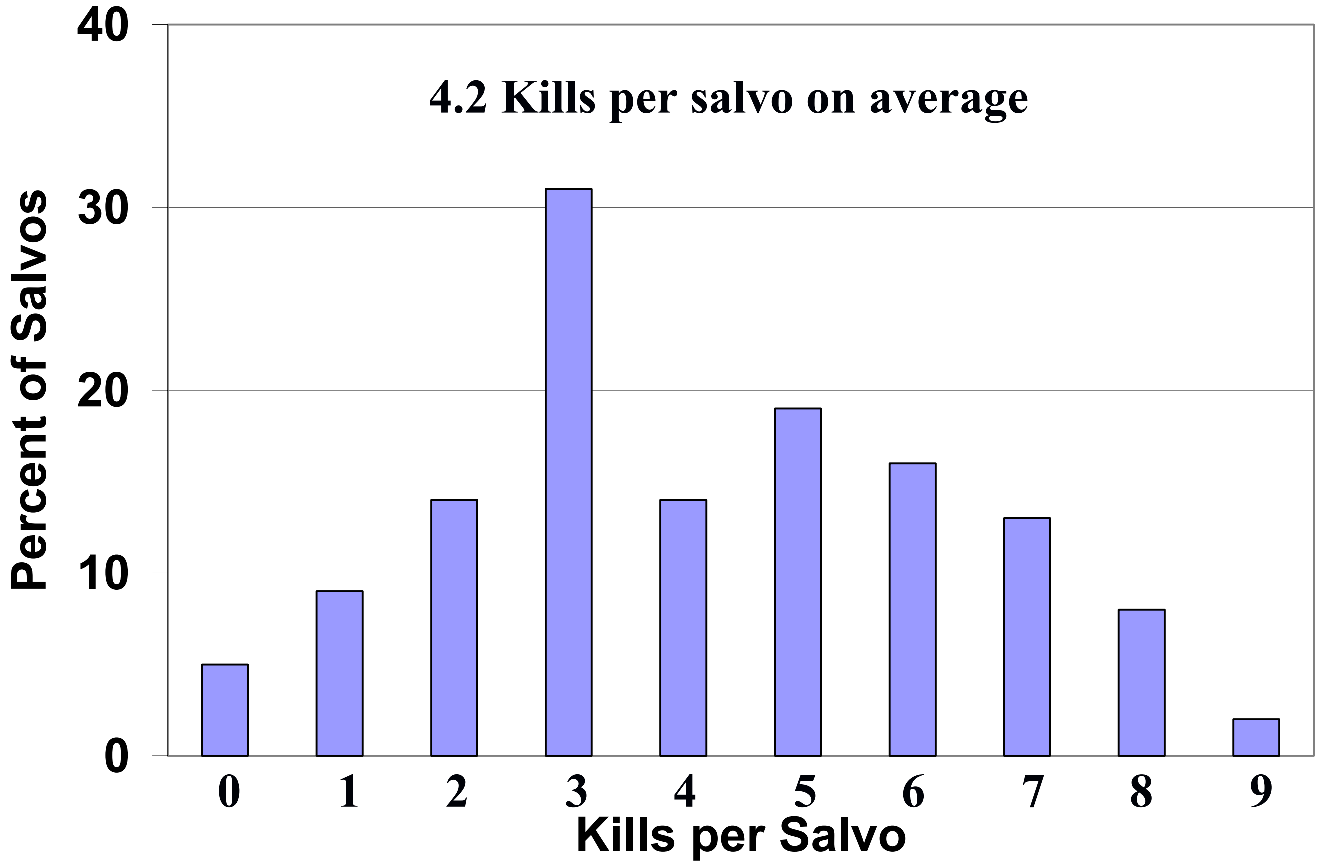
Kills per Salvo is Highly Variable

(DSB '98 salvos with 5 vulnerable AFVs in footprint)

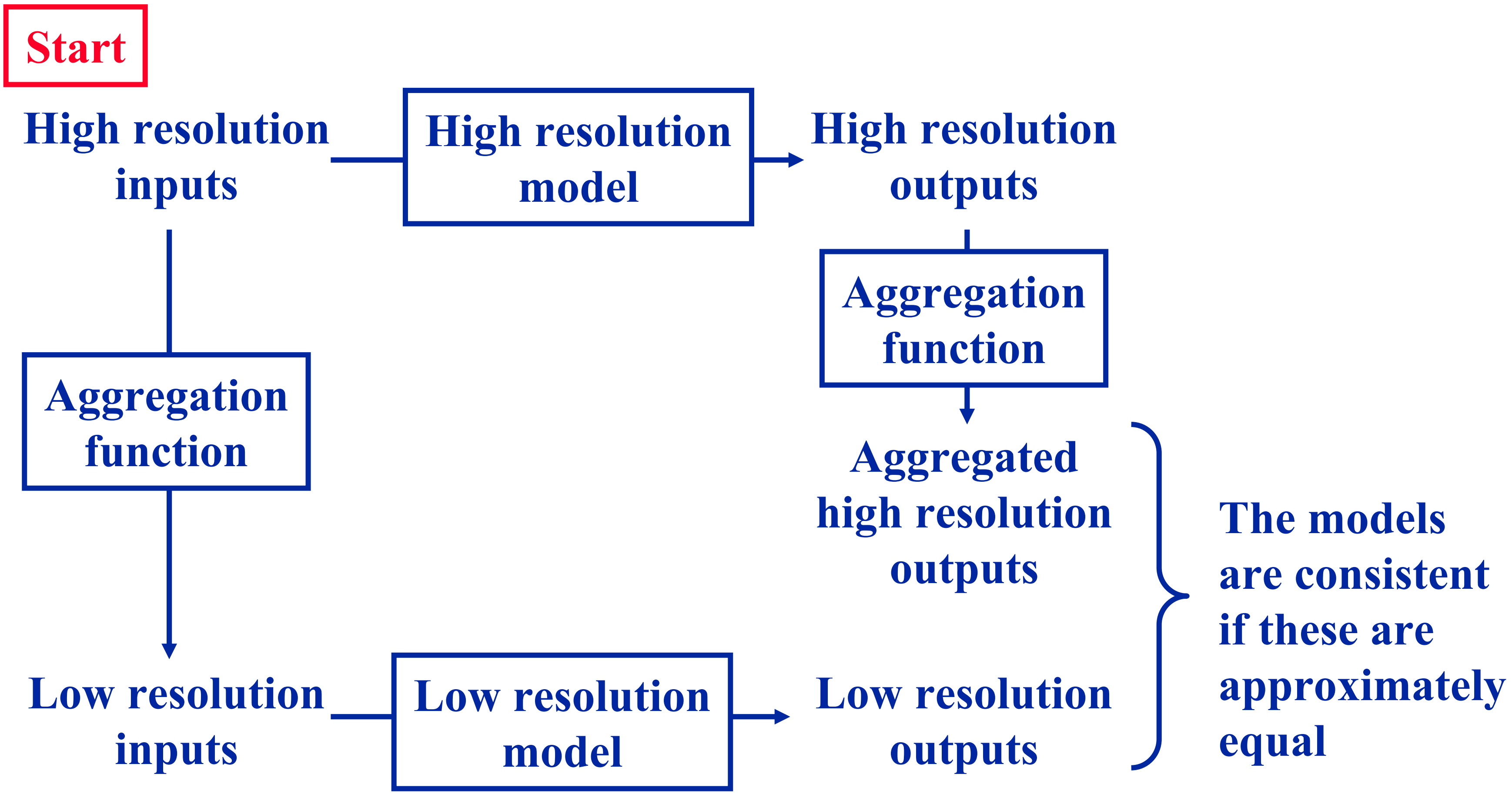


Kills per Salvo is Highly Variable - II

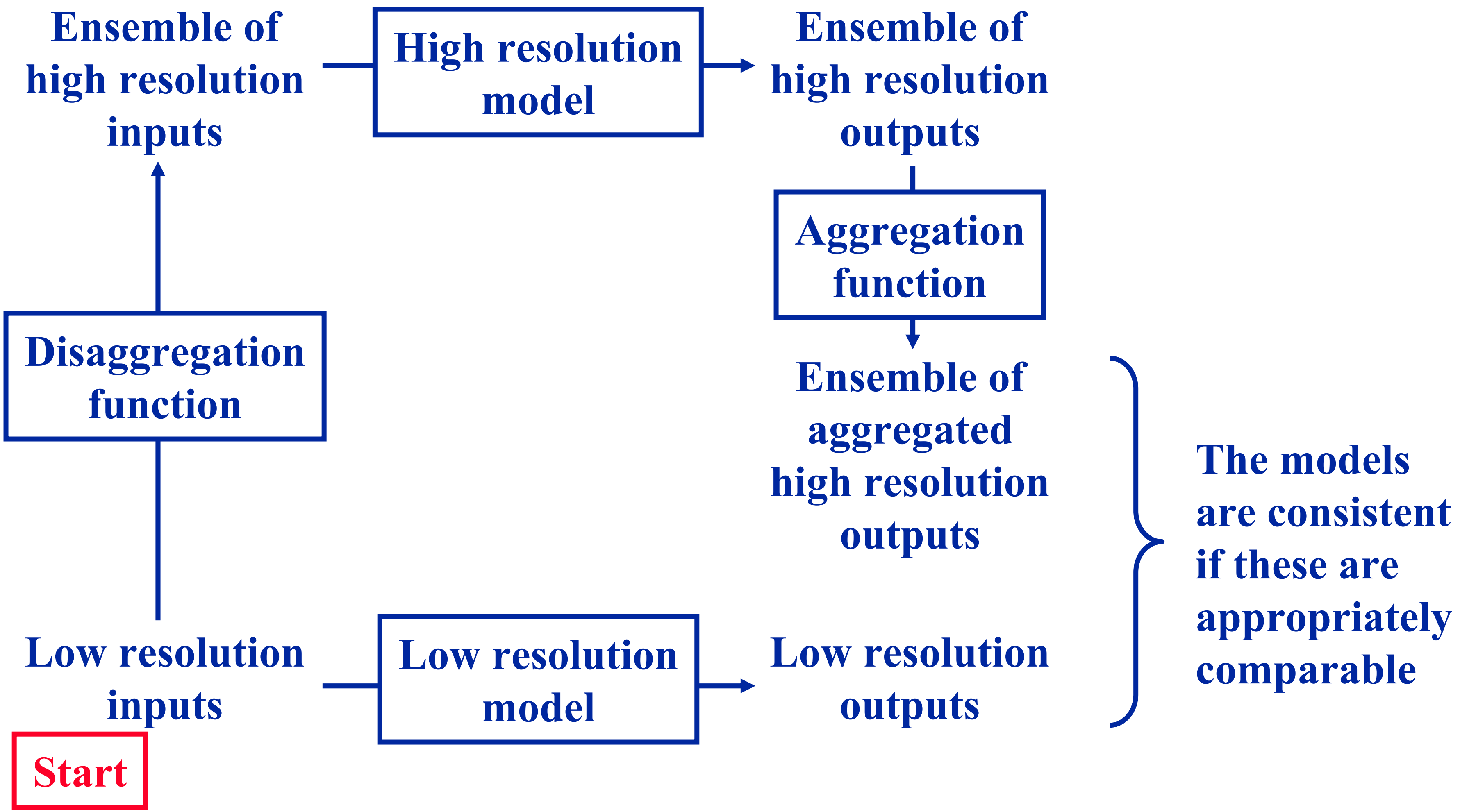
(DSB '98 salvos with 10 vulnerable AFVs in footprint)



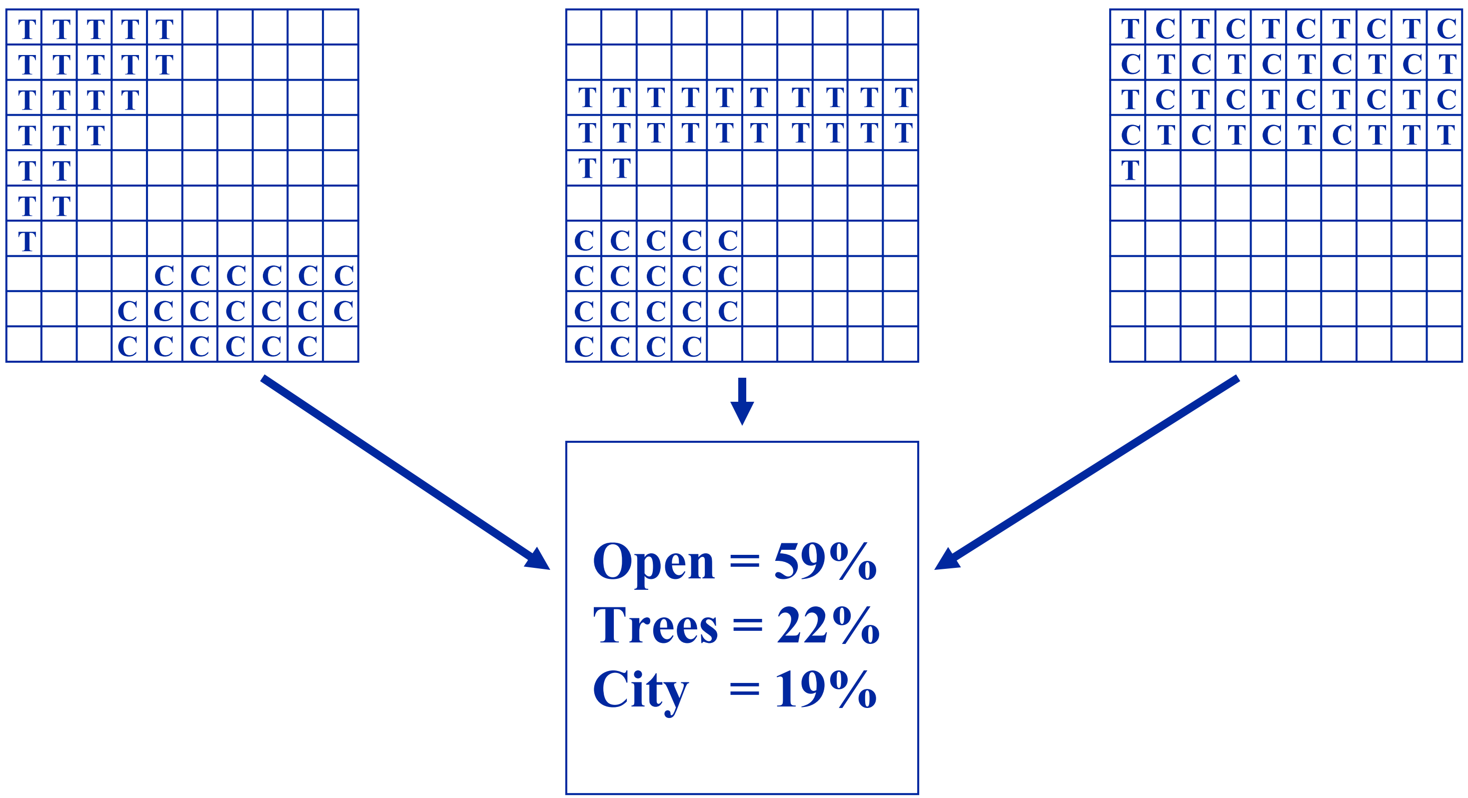
The Usual Definition of Consistency Between High- and Low-Resolution Models



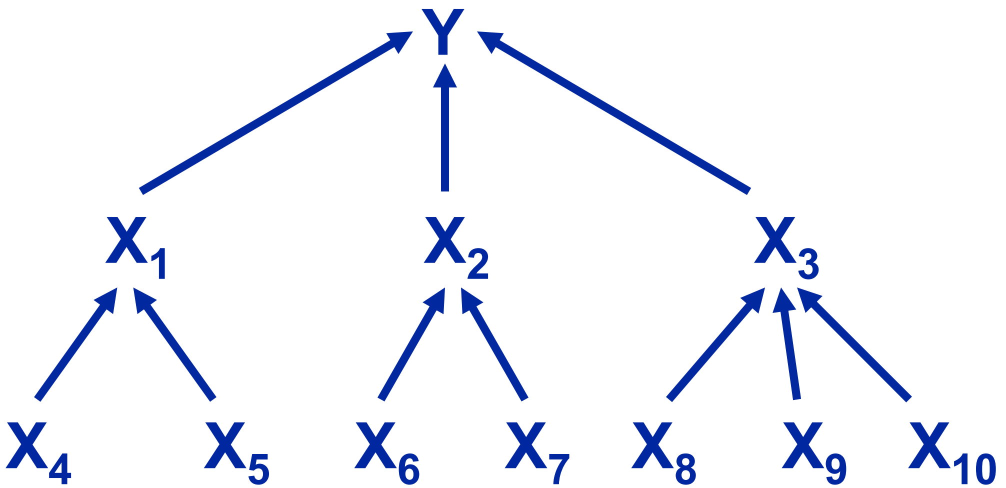
A More General Definition of Consistency Between High- and Low-Resolution Models



Ways to Change Resolution – I
Change Grid Size



Ways to Change Resolution – II
Use Intermediate Variables



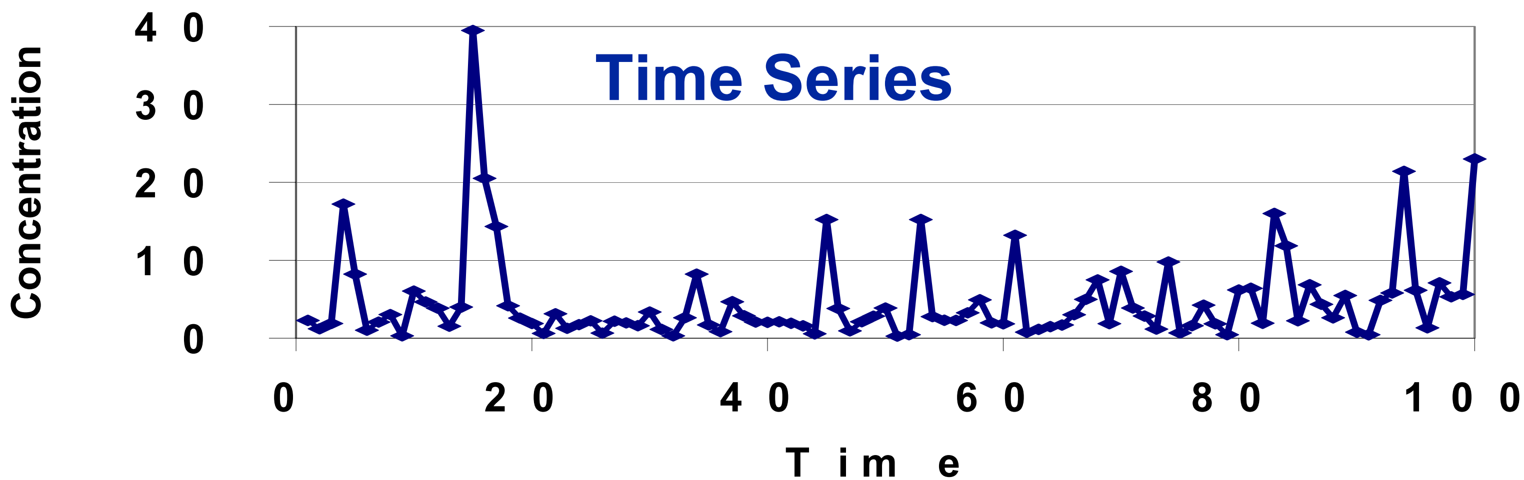
$$Y = F(X_1, X_2, X_3)$$

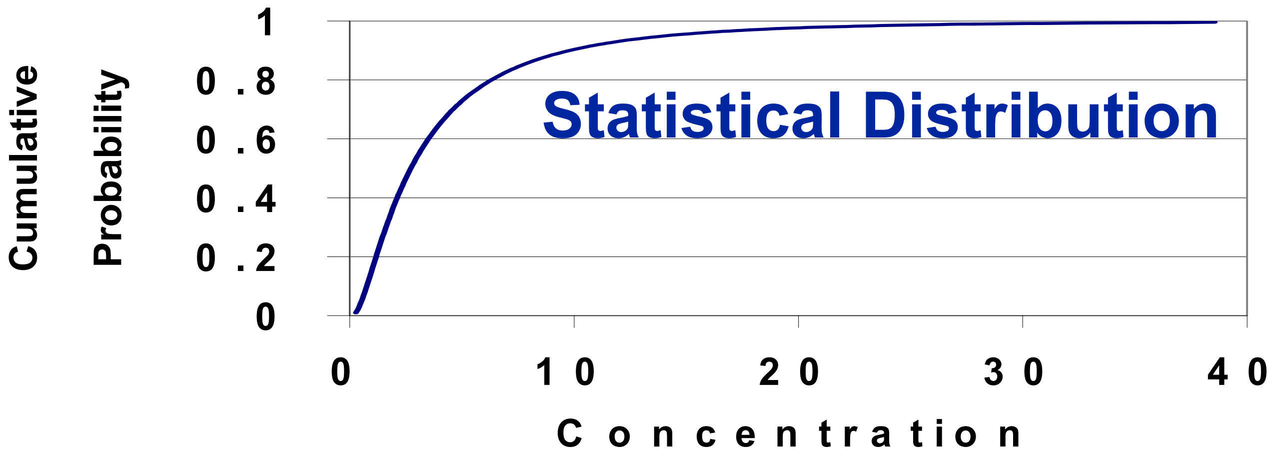
$$X_1 = G_1(X_4, X_5)$$

$$X_2 = G_2(X_6, X_7)$$

$$X_3 = G_3(X_8, X_9, X_{10})$$

Ways to Change Resolution – III Physical Versus Statistical Description

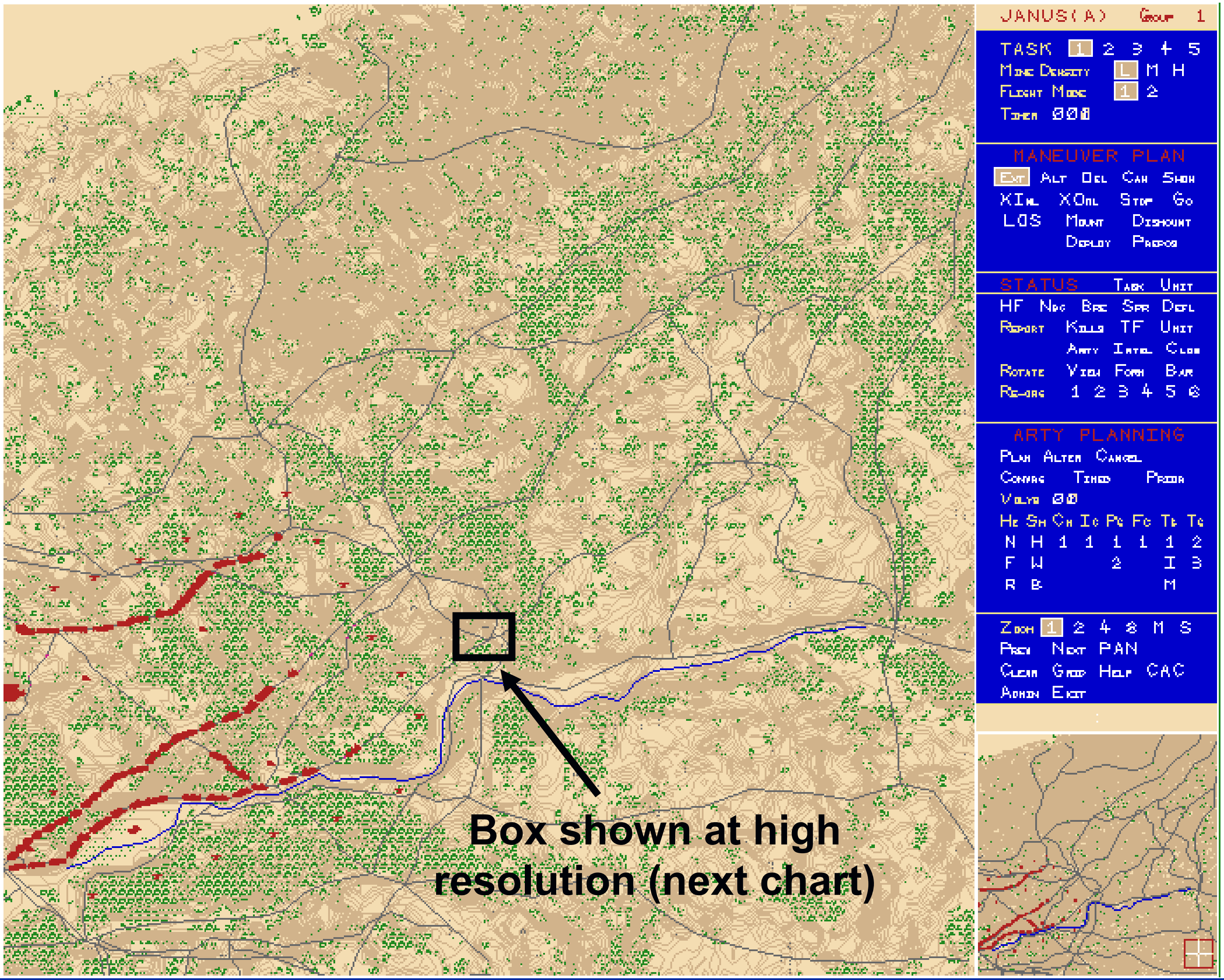




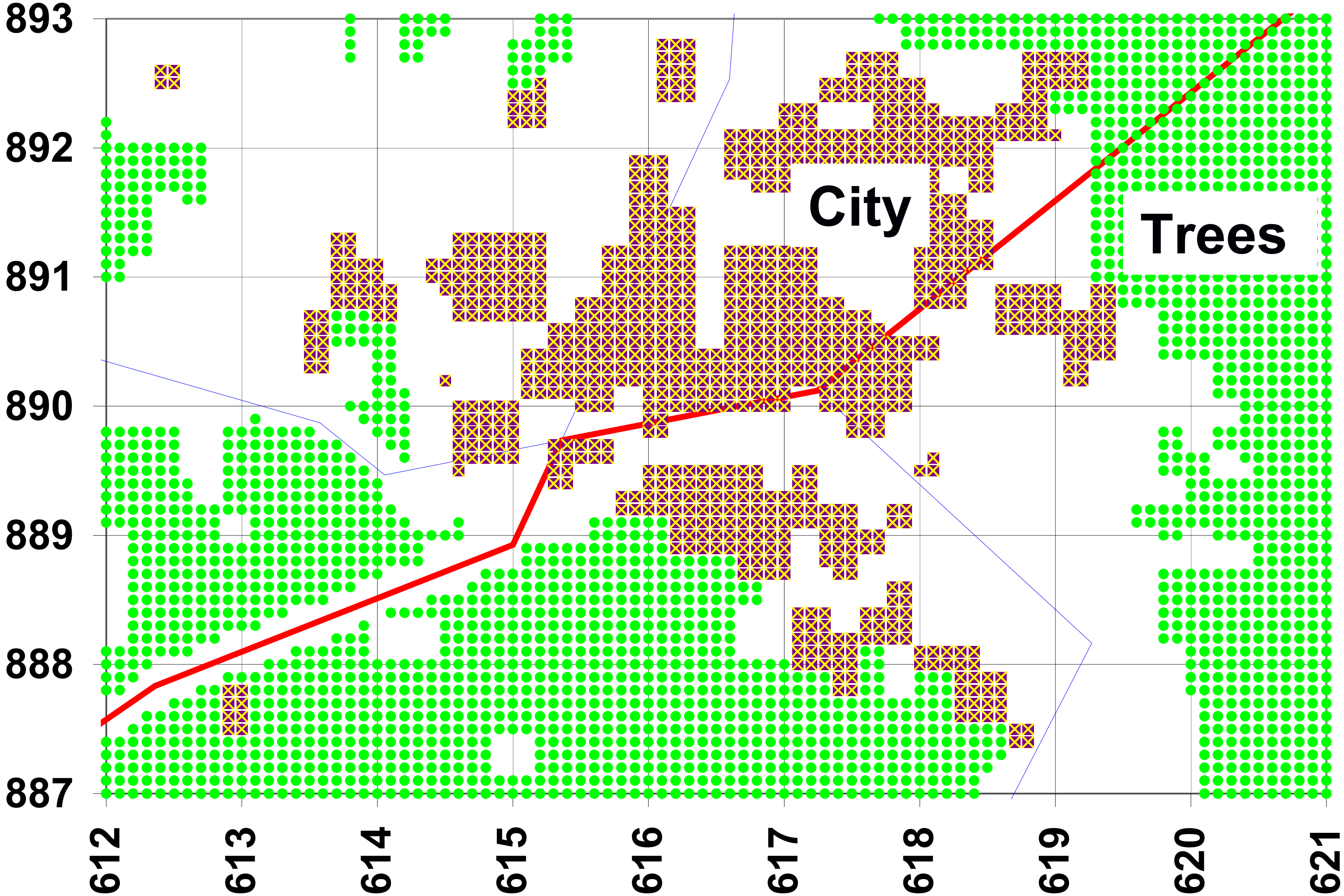
Ways to Change Resolution - IV
Abstraction

- In PEM, the concept of “clearing” is elementary
- In high-resolution simulation, the concept must be derived from more elementary constructs (terrain, etc.)
- Definition of a clearing depends on resolution
  - Man-in-the-loop must see a clearing to target it (low resolution)
  - Weapon detects vehicles in the open from close up (high resolution)
- Definition may also depend on weapon characteristics
  - Orientation effects (e.g., LOCAAS)

Terrain Seen at Low Resolution
(DSB ‘98)



A Clearing Seen At High Resolution
(DSB '98)



Use a Story to Guide Extrapolation

"Early in the process the evolution of the system should be dominated by the expansion of the cloud, the rate of which is proportional to t^3. Later the evolution the system should be dominated by a decay process, which occurs like $e^{-\alpha t}$. So the quantity of interest should rise to a maximum and then fall asymptotically to zero."

Our Modus Operandi is Experimental

- In a real-world problem, the large model could:
  - Have thousands of inputs
  - Be costly to run
  - Provide only a few cases for statistical analysis
- For this exercise our large model is EXHALT-CF:
  - Approximately 60 inputs
  - Small enough to allow us to generate unlimited cases
  - Complex enough to provide a fair test
- We explore the spectrum between the extreme approaches
  - Start with "pure" statistics
  - Progressively add more phenomenology

Generate an Analysis Data Set

- Specify distributions for 25 EXHALT-CF input variables
  - Mostly uniform distributions
  - Sample all possible regions without regard for likelihood
- Set remaining input variables to constants
- Generate a Monte Carlo sample of 1000 cases
- Collect data for each case

Poor Identification of Nonlinearities in Metamodel 1

EXHALT	METAMODEL 1
(Vehicles to kill)	
=(# divisions)	Coeff significant at P=.05
×(vehicles/division)	Coeff not significant!
×(break point)	Coeff not significant!
(Halt distance) =Smaller of distances for "In Depth" strategy, "Leading Edge" strategy	Variables for both strategies affect halt distances in all cases

Regression frequently generates flatly wrong conclusions about the importance of variables.