INSTITUTE FOR DEFENSE ANALYSES

Tutorial on Sensitivity Testing in Live Fire Test and Evaluation

Thomas Johnson
Laura Freeman
Ray Chen

June 2016
Approved for public release.
IDA Document NS D-5829
Log: H 16-000717

INSTITUTE FOR DEFENSE ANALYSES
4850 Mark Center Drive
Alexandria, Virginia 22311-1882

About This Publication

A sensitivity experiment is a special type of experimental design that is used when the response variable is binary and the covariate is continuous. Armor protection and projectile lethality tests often use sensitivity experiments to characterize a projectile's probability of penetrating the armor. In this mini-tutorial we illustrate the challenge of modeling a binary response with a limited sample size, and show how sensitivity experiments can mitigate this problem. We review eight different single covariate sensitivity experiments and present a comparison of these designs using simulation. Additionally, we cover sensitivity experiments for cases that include more than one covariate, and highlight recent research in this area.

INSTITUTE FOR DEFENSE ANALYSES

IDA Document NS D-5829

Tutorial on Sensitivity Testing in Live Fire Test and Evaluation

Thomas Johnson
Laura Freeman
Ray Chen

Executive Summary

A sensitivity experiment is a special type of sequential experimental design that is used for binary outcomes. In this tutorial we look at a common live fire test outcome – whether armor is penetrated or not by a projectile. Armor protection and projectile lethality tests often use sensitivity experiments to characterize a projectile's probability of penetrating armor as a function of the projectile's velocity. These tests are referred to as "sequential" because the experimental design is sequentially updated after each shot is recorded. Simply put, after every shot the velocity of the next projectile shot is updated based on previous test outcomes. Sensitivity experiments are often used in armor characterization testing when the objective is to estimate the velocity at which the projectile has a 50 percent probability of penetration. In past work, the authors compared numerous single factor sequential designs and concluded that 3Pod was best in terms of robustness to model misspecification, and accuracy.

Multi-factor sequential design, as the name suggests, deals with more than one continuous factor. Velocity is typically a primary factor for armor tests, but secondary factors include obliquity angle, yaw angle, armor temperature, and other physics-based continuous parameters that affect projectile penetration. Sequential design, and multi-factor sequential design in particular, are well-suited for Live Fire Test and Evaluation because such tests are often conducted in a controlled laboratory environments where precise control of multiple continuous factors is possible.

In this mini-tutorial we illustrate the challenge of modeling a binary response with a limited sample size, and show how sensitivity experiments can mitigate this problem. We review eight different single factor sensitivity experiments, and present a comparison of these designs using simulation. Additionally, we present sensitivity experiments for cases that include more than one factor, and highlight recent research in this area.

Tutorial on Sensitivity Experiments in Live Fire Test and Evaluation

Tom Johnson
Laura Freeman
Ray Chen

Sensitivity Experiments Best Practices

Outline

1. Introduction to Binary Response Experiments
2. Binary Response Test Design Challenges
3. 1-D Sensitivity Test Designs
4. 2-D Sensitivity Test Designs
5. Case Study: Greg Hutto

Introduction to Binary Response Experiments

Types of Binary Response Experiments

Pharmaceutical Industry

Lethal dose

Effective dose

Defense Industry

Lethality of munitions

Survivability of systems

Armor Characterization

Defense Industry Requirements

“Munition shall have a V50 less than 2,000 ft/s”

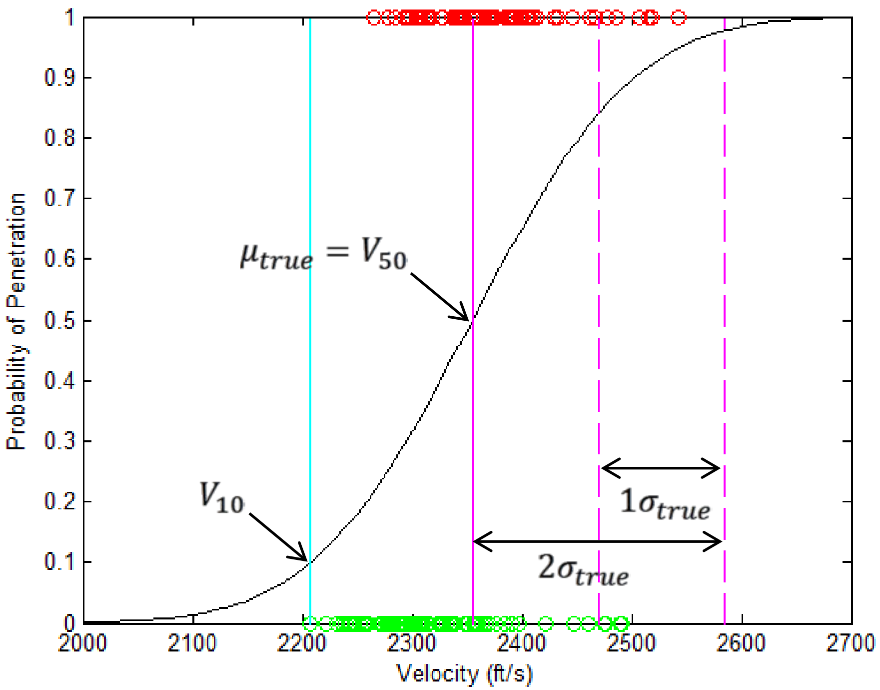
“Armor shall have a v50 greater than 2,300 ft/s”

Historically, an arithmetic mean estimator is used to calculated V50

IDA

Regression Models

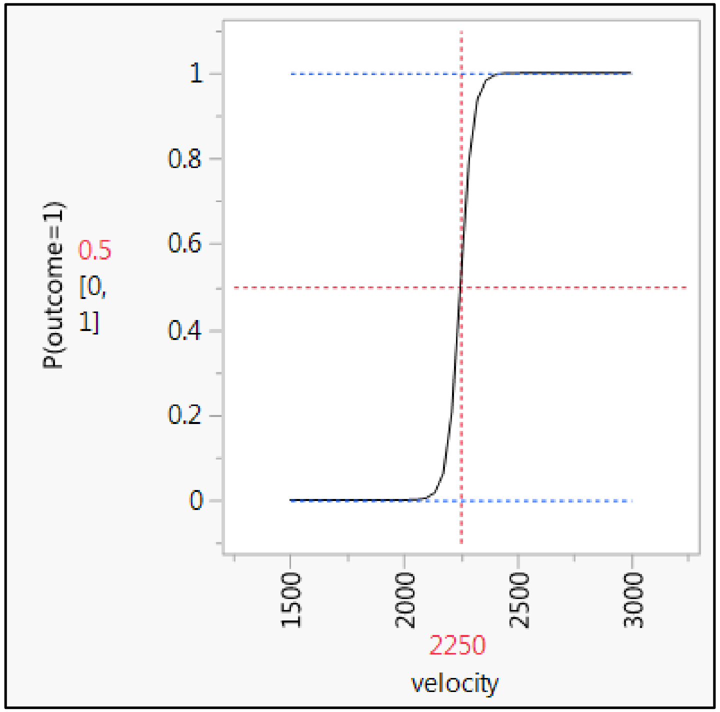
Link Name (distribution)	Probability of perforation $\pi(v)$	Velocity where probability of perforation is π, $\hat{V}_\pi$	Estimated V_{50}
Logit (Logistic)	$\dfrac{e^{\hat{\beta}_0+\hat{\beta}_1 v}}{1+e^{\hat{\beta}_0+\hat{\beta}_1 v}}$	$\dfrac{\ln\left(\dfrac{\pi}{1-\pi}\right)-\hat{\beta}_0}{\hat{\beta}_1}$	$\dfrac{-\hat{\beta}_0}{\hat{\beta}_1}$
Probit (Normal)	$\Phi\left(\hat{\beta}_0+\hat{\beta}_1 v\right)$	$\dfrac{\Phi^{-1}(\pi)-\hat{\beta}_0}{\hat{\beta}_1}$	$\dfrac{-\hat{\beta}_0}{\hat{\beta}_1}$



Binary Response Test Design Challenges

Binary Response Designs Need Special Consideration

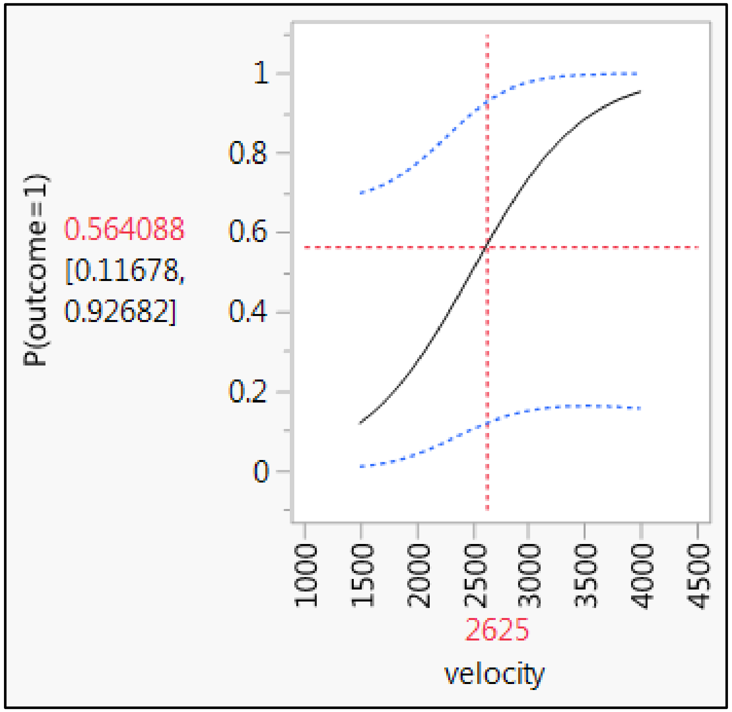
Run #	Velocity	Response
1	1500	0
2	1500	0
3	1500	0
4	1500	0
5	3000	1
6	3000	1
7	3000	1
8	3000	1



"Evidence of perfect fit" yields bad logistic model fit

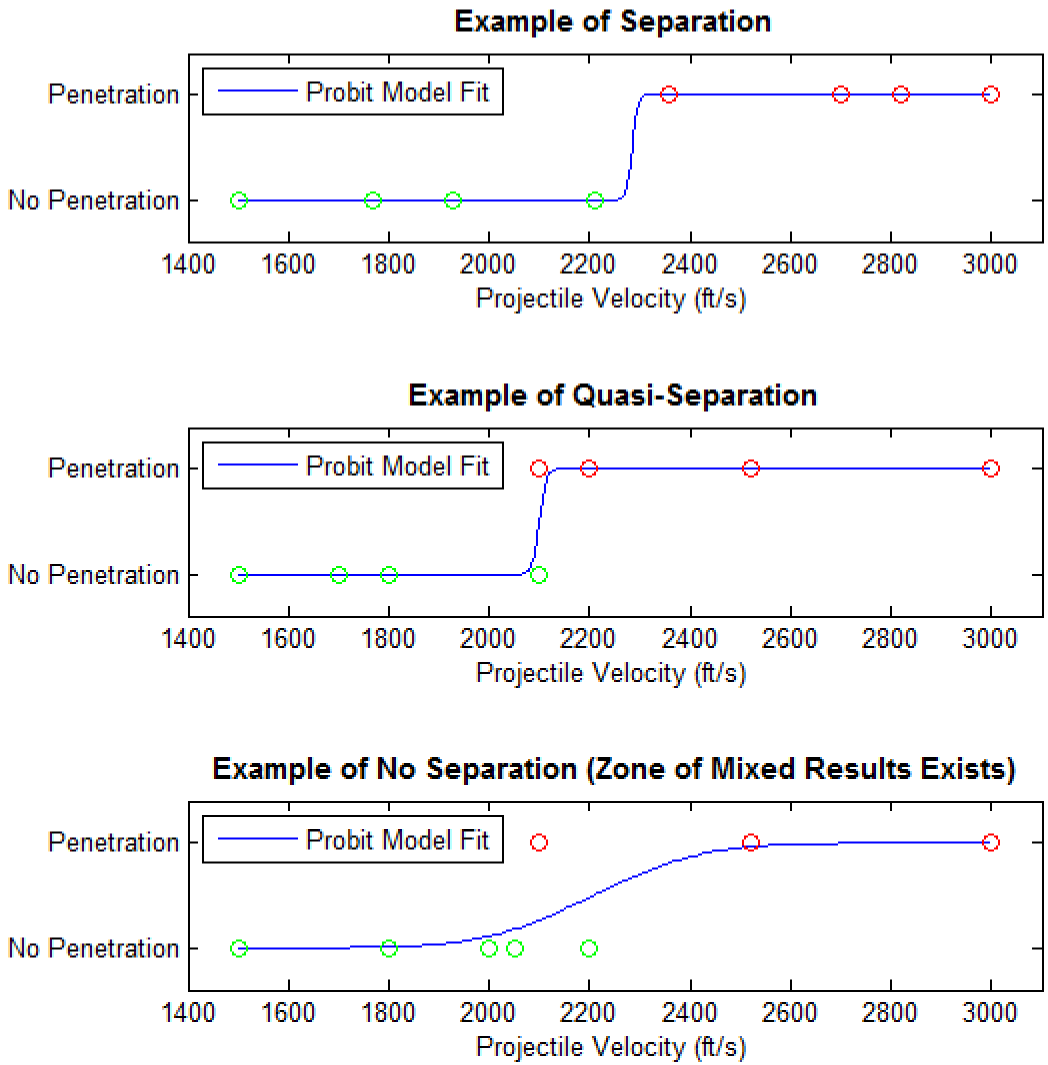
Binary Response Designs Need Special Consideration

Run #	Velocity	Response
1	1500	0
2	1500	0
3	1500	0
4	**1875**	**1**
5	**2625**	**0**
6	3000	1
7	3000	1
8	3000	1



A zone of mixed results provides a good rough estimate of the logistic model curve

Zone of Mixed Results



Test Designs to Achieve a Zone of Mixed Results

Sequential Methods with Initials Designs

Bayesian Methods

1-D Sensitivity Test Designs

Up and Down

Details of Implementation

Rules
- If projectile <u>does</u> penetrates armor, <u>decrease</u> velocity.
- If projectile <u>does not</u> penetrate armor, <u>increase</u> velocity.

Inputs
- Step size
- Velocity of projectile for trial number one

Other details
- fixed step size
- step size calculated from anticipated standard deviation
- Initial shot typically taken at predicted V50

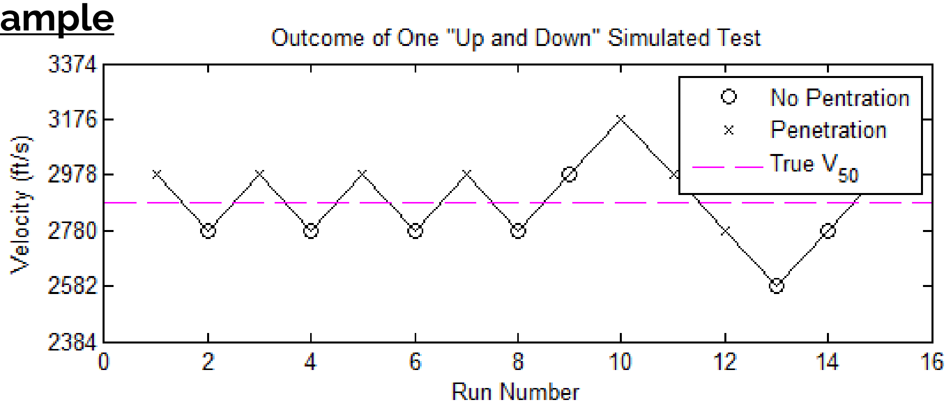
Background

- Most well-known sequential experimentation procedure, primarily due to its ease of implementation
- Developed by Dixon in 1948

Advantages

- Useful for estimating V50
- The rules are simple and practical to implement

Example

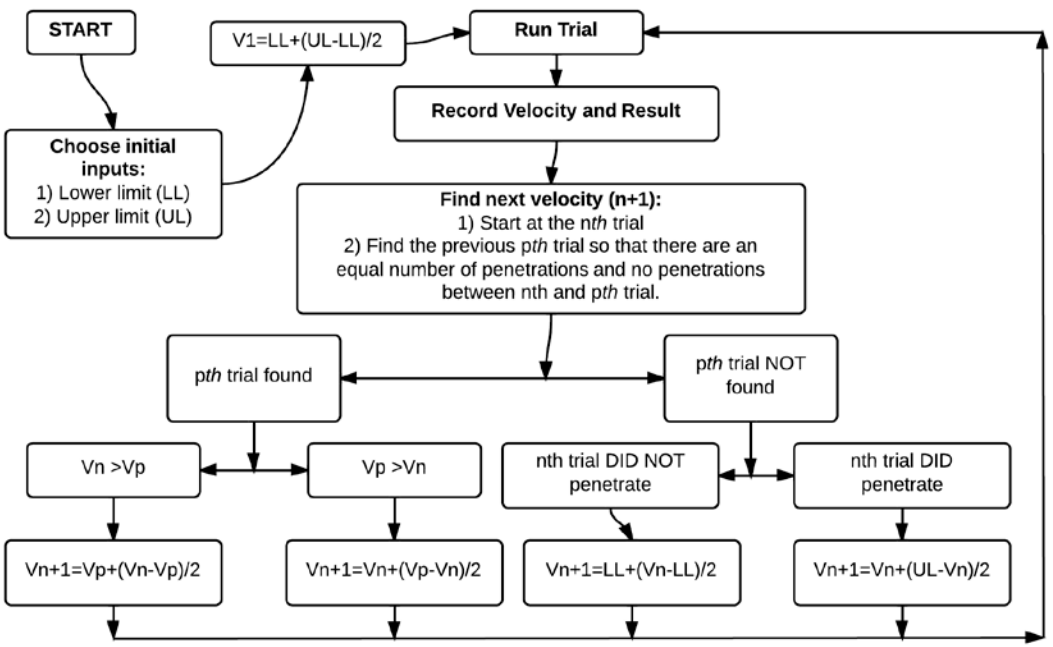


Disadvantages

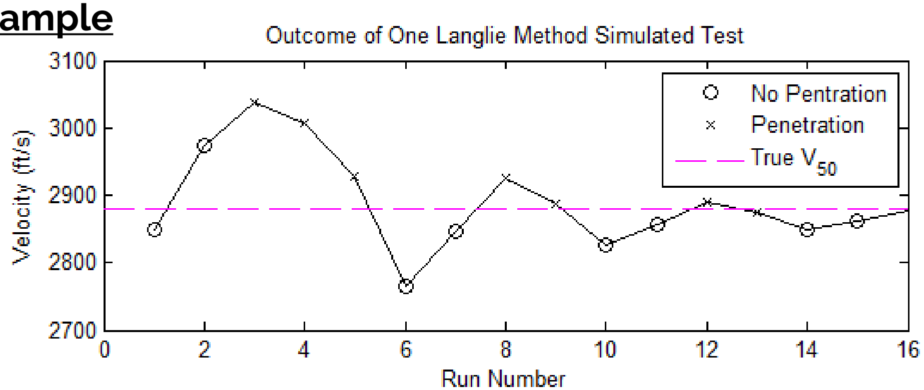
- Not good for V10
- Constant step size can lead to problems (especially for large steps)

Langlie Method

Details of Implementation



Example



Background

- Numerous modified versions exist
- Developed in early 60s

Advantages

- Useful for estimating V50
- Has an adaptive step size

Disadvantages

- Not designed for d-optimal curve fitting
- Not as easy to implement as up and down method

K-in-a-row

Details of Implementation

- If projectile does penetrates armor, decrease velocity.
- If projectile does not penetrate armor k times in a row, increase velocity.
- The step size is chosen based on the standard deviation of the predicted response curve.
- Targets Pth quantile of interest where

$$P = 1 - \left(\frac{1}{2}\right)^{\left(\frac{1}{k}\right)}$$

- Typically, k=2 (P≈0.3) or k=3 (P≈0.2)

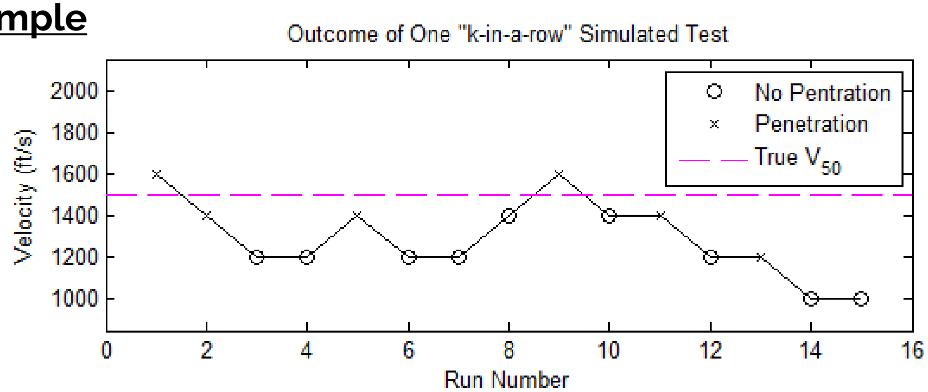
Background

- Similar to Up and Down Method
- Not typically used in armor testing

Advantages

- Useful for estimating percentiles away from the median
- Easy to implement (similar to Up and Down method)

Example



Disadvantages

- Less accurate for estimating V50
- A constant step size is susceptible to problems

Robbins Monroe

Details of Implementation

- Start the test at predicted V50.
- Determine the velocity of the next shot using

$$x_{n+1} = x_n - c(y_n - P)/n$$

where c is an arbitrary constant , yn is the outcome of the nth trial (0,1), P is the desired percentile of interest and n is the number of trials. C is optimal when:

$$c_{opt} = [F'(V_p)]^{-1}$$

where F is the response curve and Vp is the velocity at the pth percentile

- Step size decreases as n increases

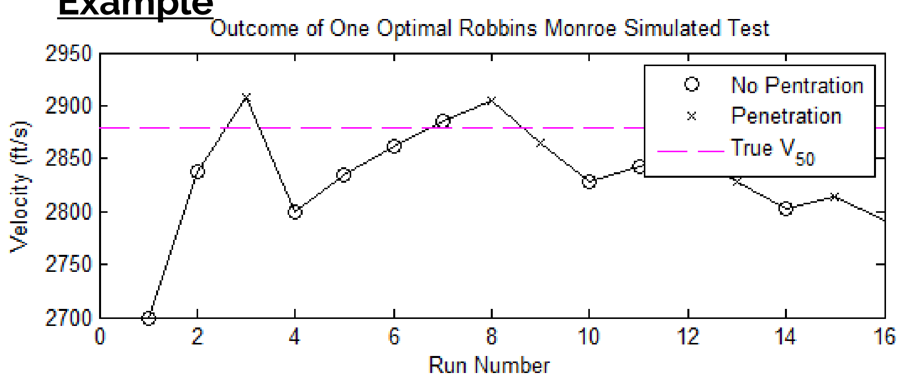
Background

- Developed in 1951
- Numerous variants of this method exist
- Used in armor testing by ARL
- Joseph (2004) improved upon method

Advantages

- Useful for estimating all quantiles
- A dynamic step size has advantages

Example



Disadvantages

- Justification for values of c may seem arbitrary, poor choices of c can lead to inaccurate results
- Poor guess of the velocity of the first shot can lead to slow convergence and/or convergence to an inaccurate result

Neyer's Method

Details of Implementation

- **Phase 1**: **Generate penetrations and non-penetrations.** Bounds the problem. Determines if initial gate is too far left, right or narrow.
- **Phase 2**: **Break separation.** Provides unique MLE coefficient estimates and an indication that velocity is in the ballpark of V50.

Initial Design

- **Phase 3**: **Refine model coefficients.** Use D-optimality criterion to dictate ensuing shots.

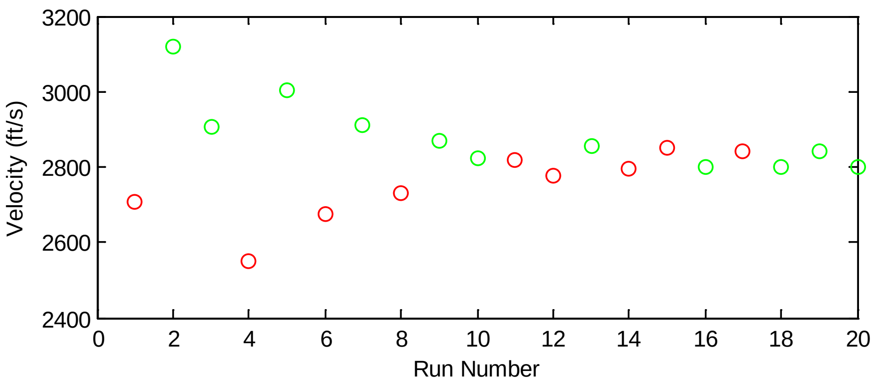
Background

- Developed by Neyer in 1989
- First to propose a systemic method for generating a good **initial design**

Advantages

- Initial design is useful for quickly estimating model coefficients
- Robust to misspecification of input parameters

Example



Disadvantages

- Requires coding and capability to do maximum likelihood estimation

3Pod

Details of Implementation

- **Phase 1: Generate penetrations and non-penetrations.** Similar to rules to Neyer's method. Uses slightly different logic and different step sizes.

- **Phase 2: Break separation.** Relies more heavily on conditional logic then Neyer's method.

Initial Design

- **Phase 3: Refine model coefficients (and estimate of Vp).** A portion of resources is devoted to D-optimal algorithm and the other portion in used for placing shots near Vp (velocity percentile value of interest) using Robbins Monroe Joseph method.

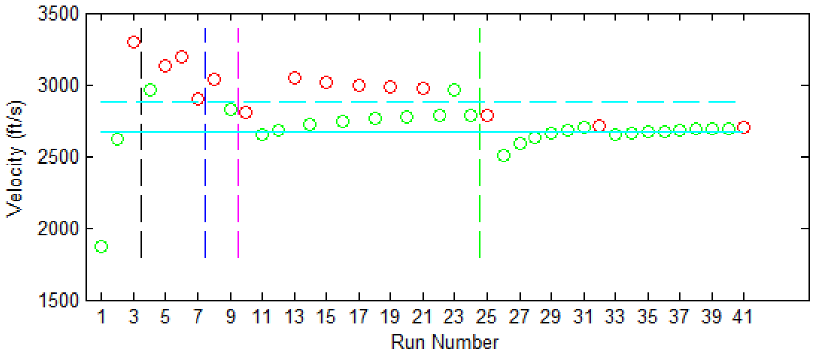
Background

- Developed by Wu in 2013
- Similar to Neyer's Method

Advantages

- Similar to Neyer's Method, good initial design

Example

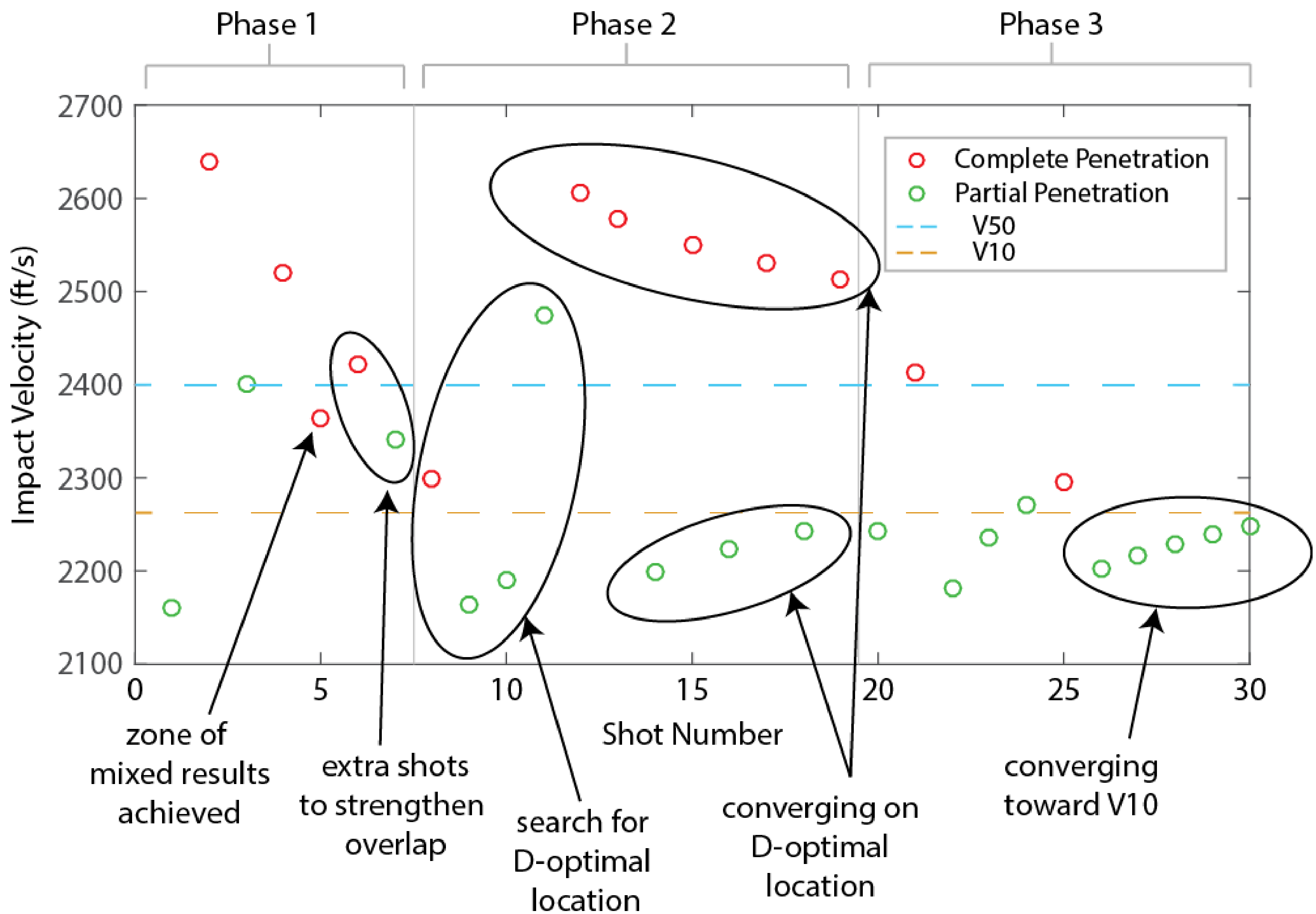


Disadvantages

- Requires maximum likelihood estimation
- More complex than Neyer's method

Example of 3Pod Results

- **Example of 30 Shots for 3-Phase Approach (3Pod)**



Simulation Comparison

Simulation Factors and Responses

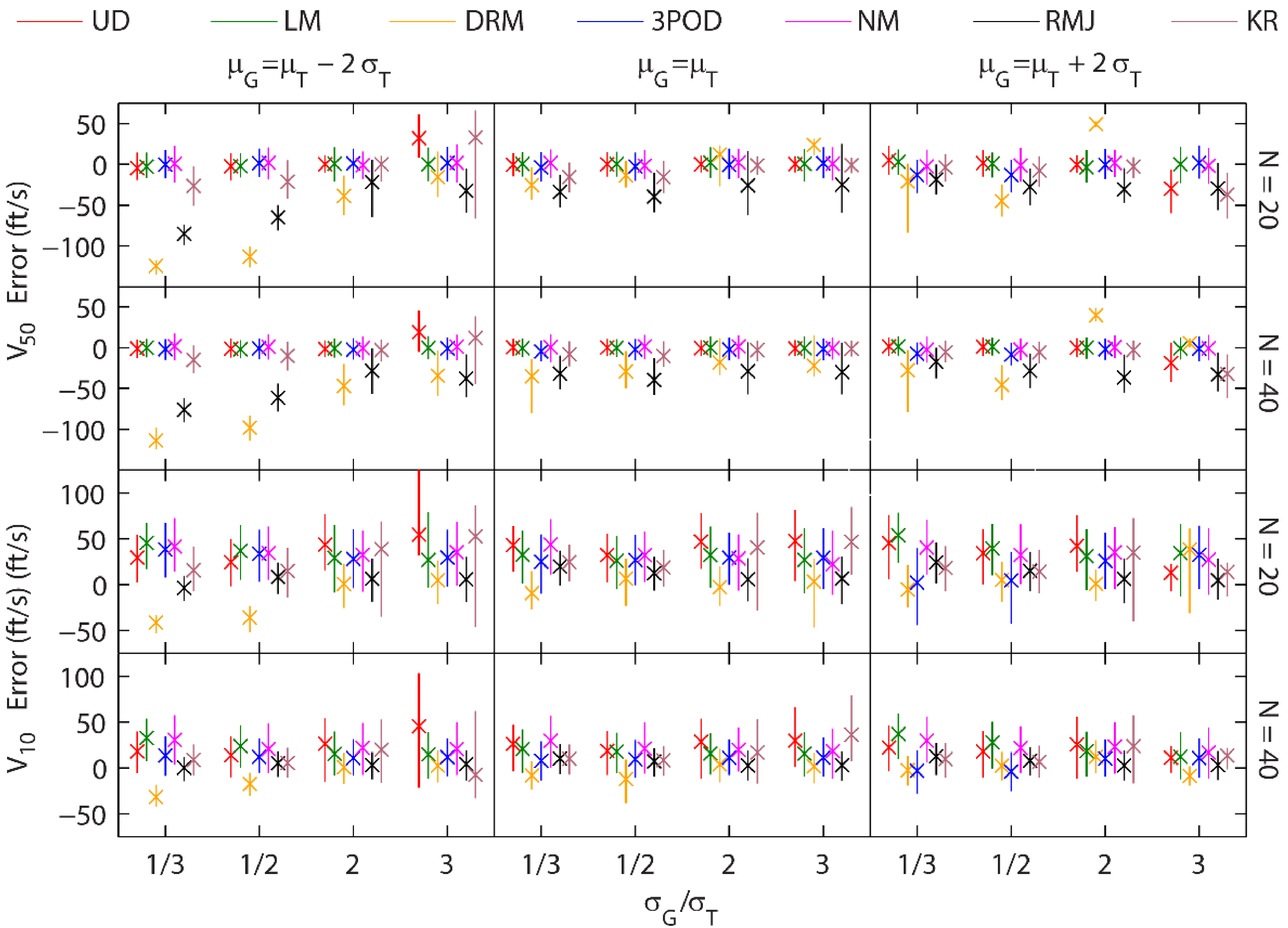
Response

1. V50 Error
2. V10 Error

Calculated as the difference between the "true" V50 (or V10) and the V50 (or V10) estimated with the simulated runs

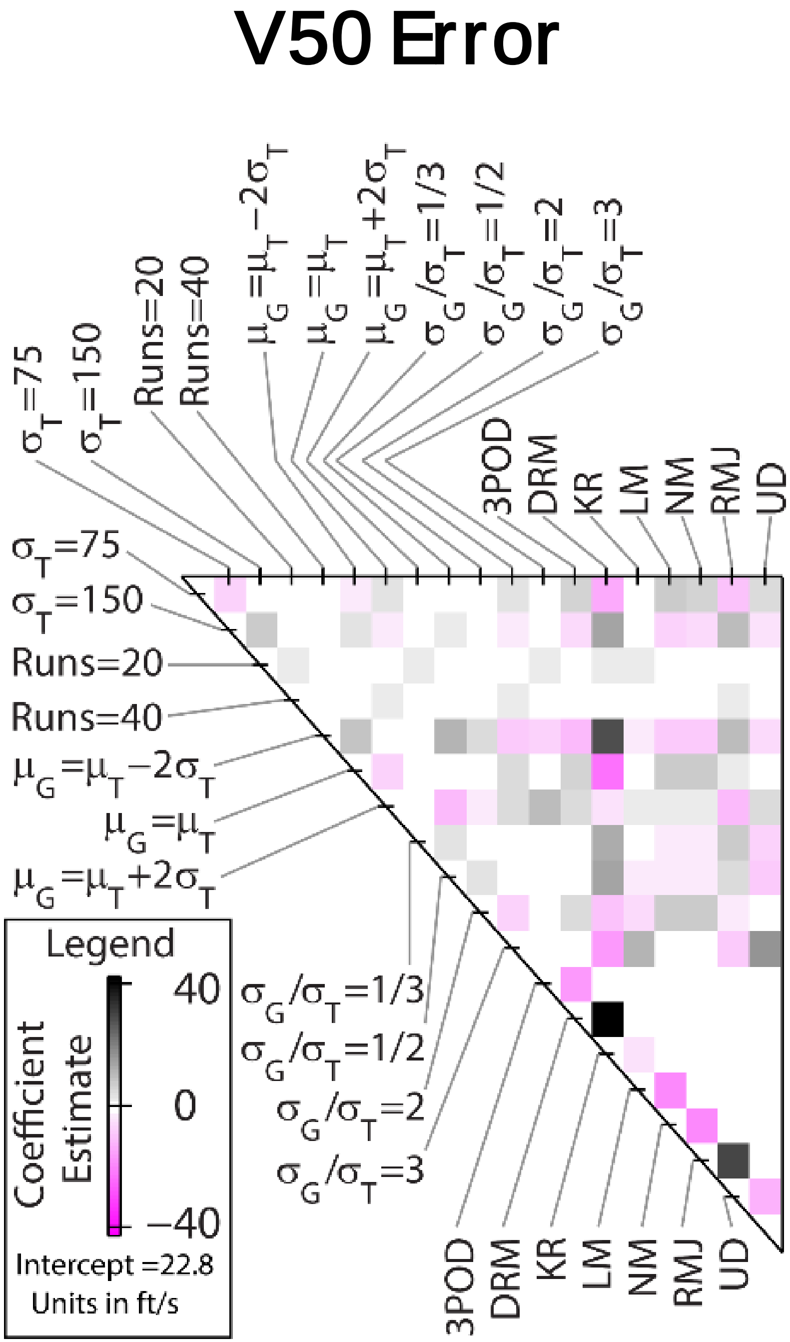
Factors

1. Estimator (Probit-MLE, Arithmetic Mean)
2. Method (Up Down Method, 3Pod, Langlie, etc…)
3. Stopping criteria ("3&3", break separation)
4. μ_{guess} $(\mu_{true} - 2\sigma_{true}, \mu_{true}, \mu_{true} + 2\sigma_{true})$
5. σ_{guess} $(1/3\sigma_{true}, 1/2\sigma_{true}, 2\sigma_{true}, 3\sigma_{true})$

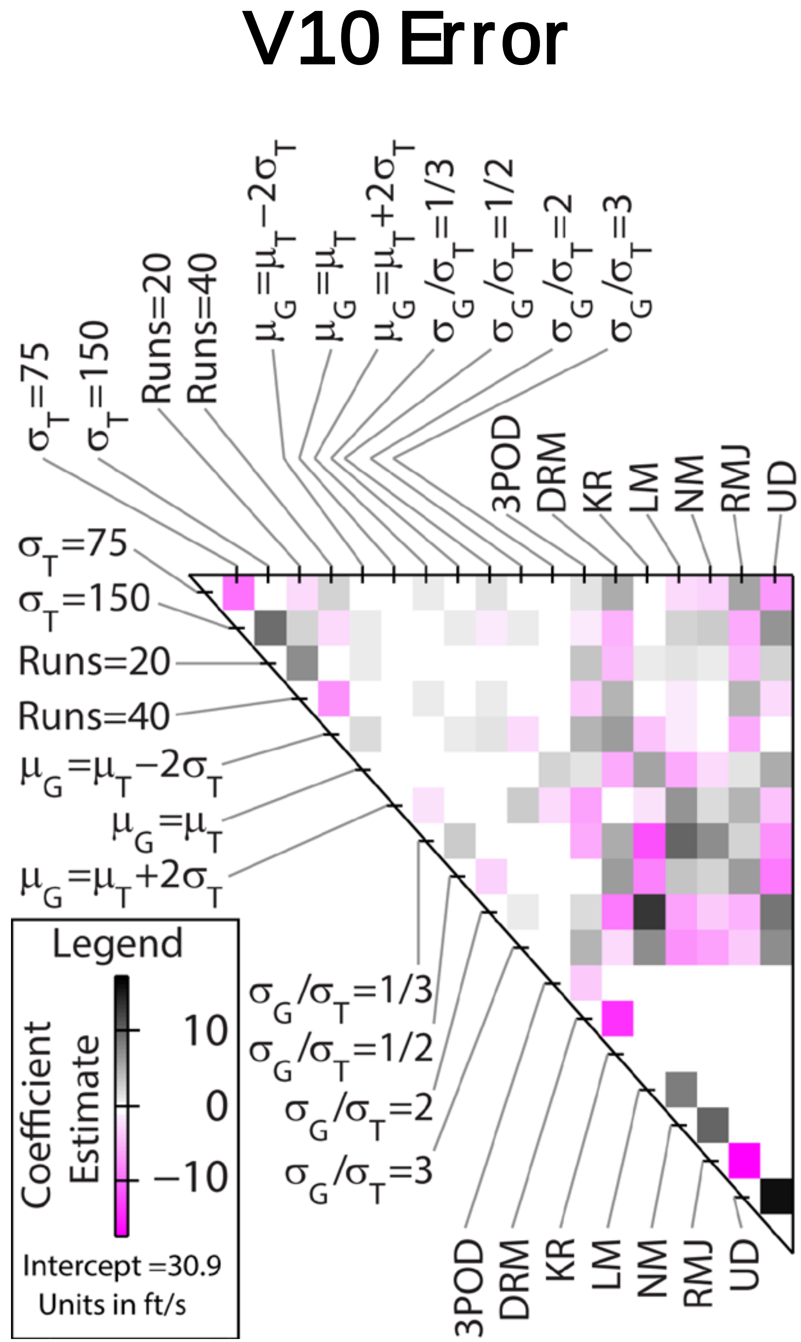


UD
LM
DRM
3POD
NM
RMJ
KR
$\mu_G = \mu_T - 2\sigma_T$
$\mu_G = \mu_T$
$\mu_G = \mu_T + 2\sigma_T$
V_{50} Error (ft/s)
V_{10} Error (ft/s) (ft/s)
N = 20
N = 40
N = 20
N = 40
50
0
−50
−100
100
1/3
1/2
2
3
σ_G/σ_T

IDA



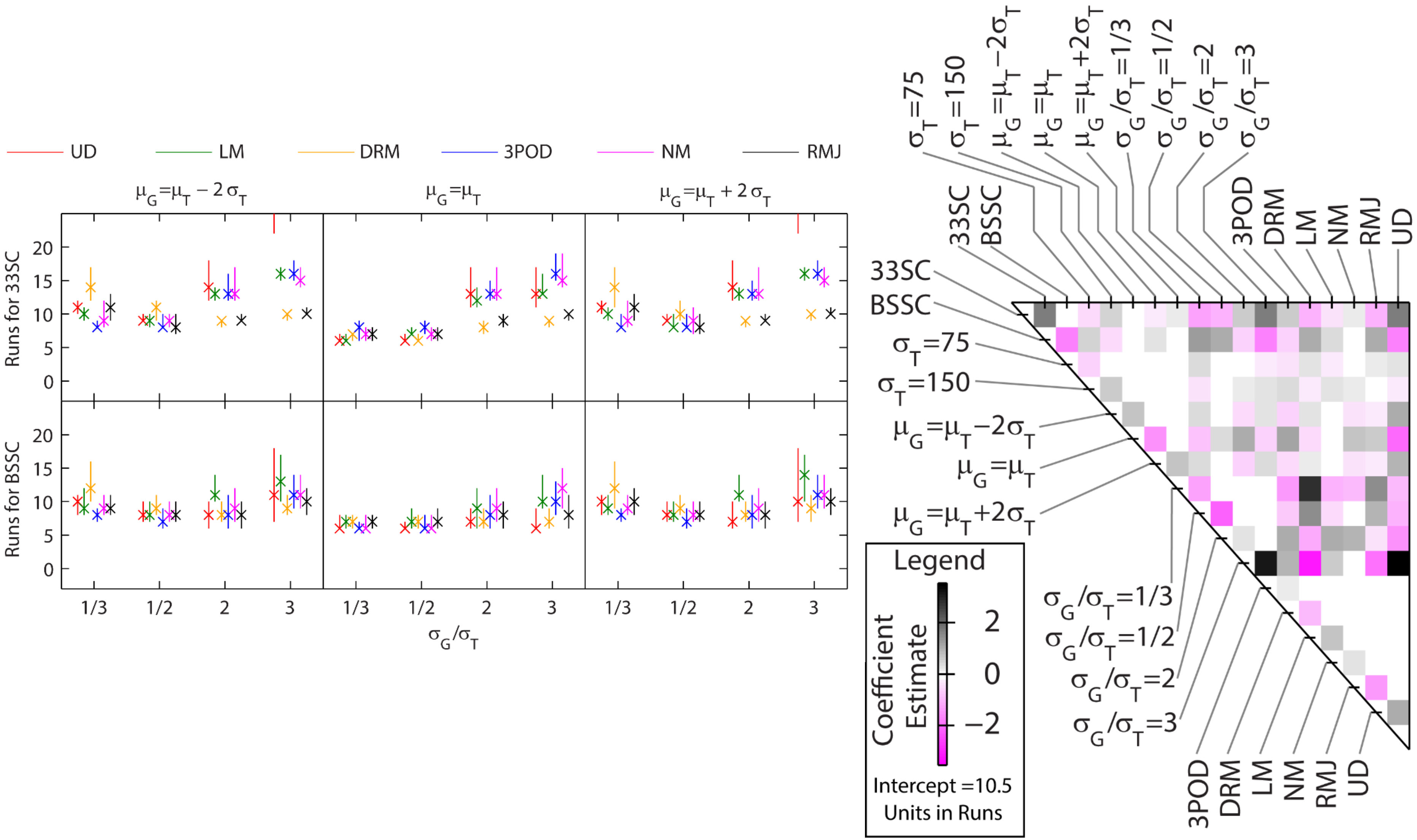
V50 Error
σT=75
σT=150
Runs=20
Runs=40
μG=μT−2σT
μG=μT
μG=μT+2σT
σG/σT=1/3
σG/σT=1/2
σG/σT=2
σG/σT=3
3POD
DRM
KR
LM
NM
RMJ
UD
Legend
Coefficient Estimate
40
0
−40
Intercept =22.8
Units in ft/s



V10 Error
σT=75
σT=150
Runs=20
Runs=40
μG=μT−2σT
μG=μT
μG=μT+2σT
σG/σT=1/3
σG/σT=1/2
σG/σT=2
σG/σT=3
3POD
DRM
KR
LM
NM
RMJ
UD
Legend
Coefficient Estimate
10
0
−10
Intercept =30.9
Units in ft/s

IDA

Runs for Stopping Criteria



Recommend 3Pod or Neyer Method

Provides entire logistic model curve fit

Robust estimate for V50 and V10

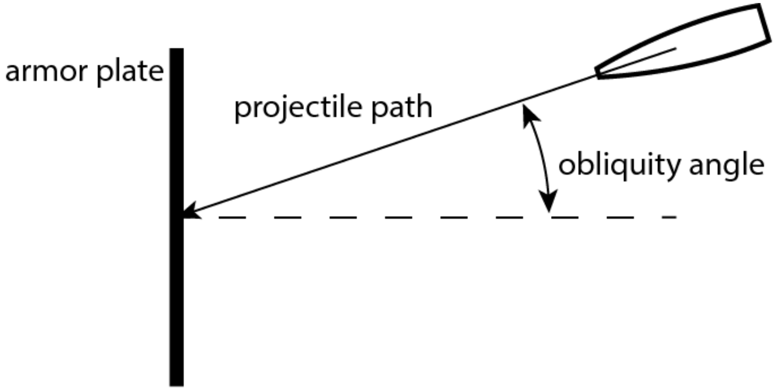
D-optimal approach

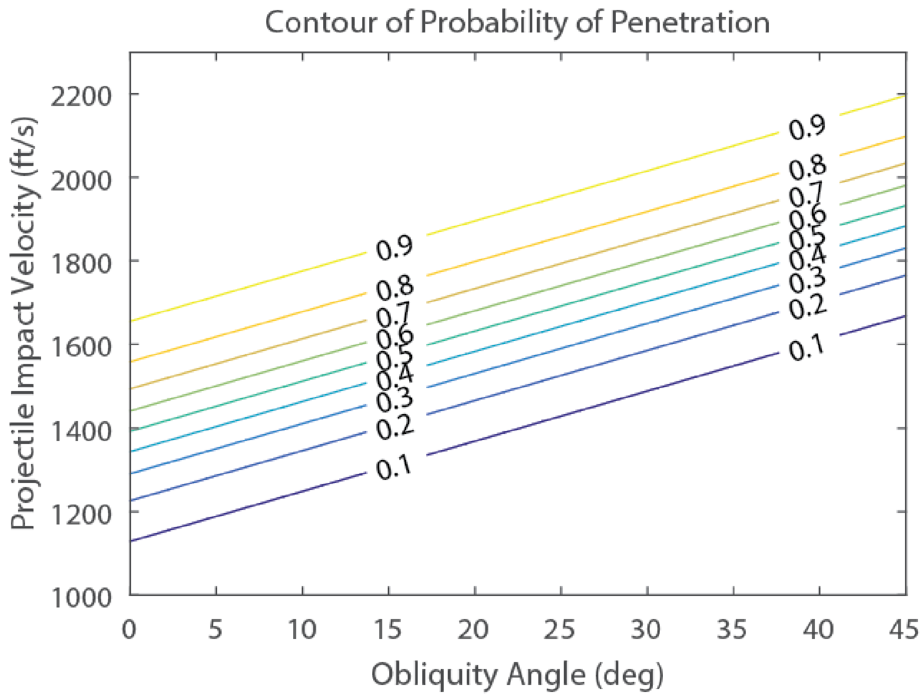
2-D Sensitivity Test Designs

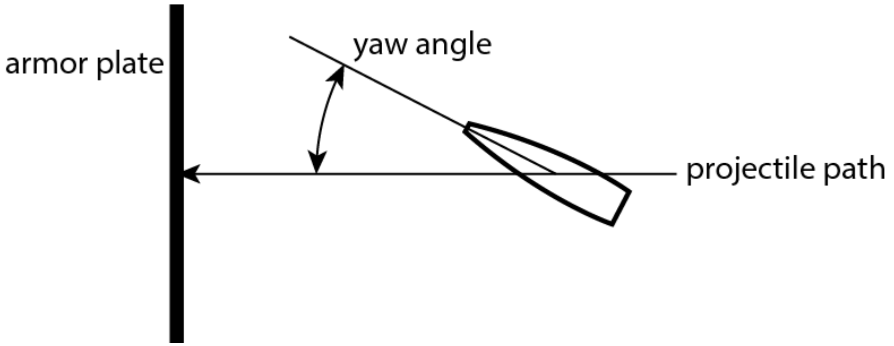
IDA

Sensitivity Test Designs with Two Factors

- Response is binary
- no interaction terms
- Two continuous factors
- Primary factor is velocity







Practical Multi-Factor Sequential Design

Practical multi-factor sequential designs:

1. Brute force use of single factor sequential designs in multi-dimensional design space
   - Intuitive design and easy to implement

		Armor Plate Size		
		S	M	L
Obliquity Angle (deg)	0	3Pod	3Pod	3Pod
	20	3Pod	3Pod	3Pod
	40	3Pod	3Pod	3Pod

Each 3Pod uses velocity as factor

2. Propose a modified sequential design to search D-optimal points across multiple factors

3. Bayesian Sequential Design by Dror and Steinberg (2008)
   - Established, practical sequential design for multiple factors
   - Uses prior information about armor performance to search for D optimal points

Role of D-Optimality in Sequential Designs

1. 3Pod, Neyer, and DS focus on D-optimality
   - D-optimality is a widely accepted design criteria
   - D-optimality is a widely accepted design criteria
   - minimizes the confidence ellipsoid on coefficients

Calculation of D-optimality

The D-optimality designs criterion for fitting a logistic model maximizes the determinant of the information matrix among all competing designs (Ω).

$$Max_{\Omega}\ |I(\beta)|$$

The fisher information matrix is $I(\beta) = |X'\Sigma X|$

X is the m x p model matrix.

Σ is the variance-covariance matrix for the m x 1 vector of binomial variables, each being $\sum_j y_{ij}$, the sum of events at the i^{th} design point.

Σ is an m x m diagonal matrix with the i^{th} diagonal element being $n_i P_i (1 - P_i)$.

2. Multi-factor sequential designs are compared in terms of D-efficiency
   - The D-efficiency of a candidate design is calculate as

$$\text{D-efficiency} = \frac{|X'\Sigma X|_{Candidate\ Design}}{|X'\Sigma X|_{D-optimal\ Design}}$$

D-Optimal Design with 1 Factor

- **The single factor logistic regression model,** $ln\left\{\frac{p}{1-p}\right\} = \beta_0 + \beta_1 x_1$**, can be reparametrized in terms of location-scale parameters as** $ln\left\{\frac{p}{1-p}\right\} = \frac{x_1-\mu}{\sigma}$ **, where** $\mu = -\frac{\beta_0}{\beta_1}$ **and** $\sigma = \frac{1}{\beta_1}$
  - μ is V_{50} and σ is the amount of slope in the curve
  - Figure 1 illustrates various logistic model curve fits
- **Abdelbasit and Plackett derived the determinant of the fisher information matrix:** $|I| = \frac{n^2 w_1 w_2}{\sigma^2}(x_1 - x_2)^2$**, where** $w_i = p_i(1-p_i)$ **and** $x_i = ln\left\{\frac{p_i}{1-p_i}\right\}\sigma + \mu$ **, for** $i = 1, 2$**.**
  - Assumes a 2 point design where where p_1 is symmetrical to p_2, and n is the number of runs at each point.
- **Abdelbasit and Plackett showed the solution is the** δ **that maximizes** $|I|$**, where** $p_1 = \delta$ **and** $p_2 = 1 - \delta$
- **The D-optimal solution (Figure 2) is** $p_1 = 0.176$ **and** $p_2 = 0.824$
  - Meaning that half of the shots are fired at $V_{17.6}$ and the other half are fired at $V_{82.4}$

D-Optimal 1-Factor Design Specifies Shots at $V_{17.6}$ and $V_{82.4}$

Figure 1 – Example Model Fits

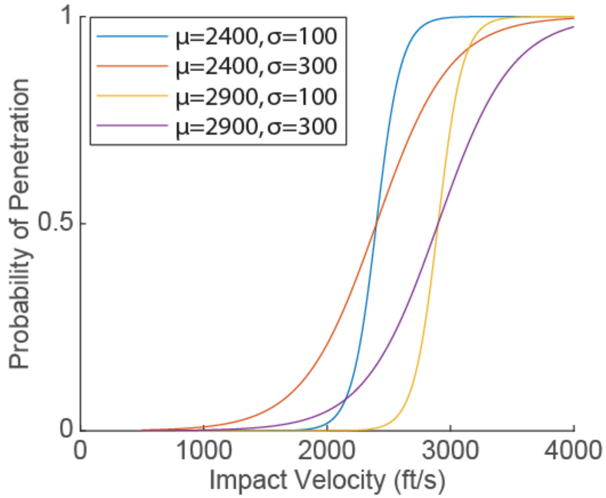
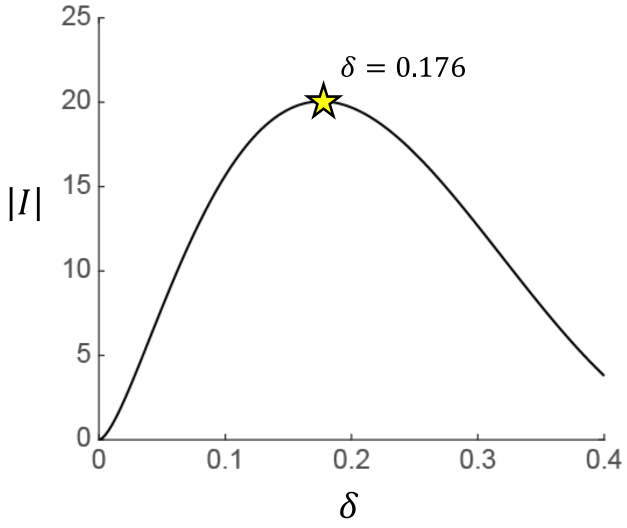


Figure 2 – Numerical Solution



Abdelbasit and Plackett, Journal for the American Statistical Association, Vol. 78, No. 381, pp. 90-98, March 1983.

D-Optimal Design with 2 Factors

- **The dual factor logistic regression model can be expressed as**

$$ln\left\{\frac{p}{1-p}\right\} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \quad \text{or} \quad ln\left\{\frac{p}{1-p}\right\} = u$$

obliquity angle Impact velocity

- **Sitter and Torsney (1995), and Jia and Meyers (2001) developed a 4 point D-optimal design**
  - 2 points are placed at the lower obliquity angle setting (θ_L) and 2 points are placed at the upper setting (θ_U)
  - Results in a location-scale parametrization:

$$\mu_L = -\beta_0/\beta_2 - \beta_1\theta_L/\beta_2\,, \mu_U = -\beta_0/\beta_2 - \beta_1\theta_U/\beta_2\,, \sigma = 1/\beta_2$$

  - **4 point D-optimal design**:

	Point 1	Point 2	Point 3	Point 4
Location	$(-u-\beta_0, 0)$	$(0, -u-\beta_0)$	$(u-\beta_0, 0)$	$(0, u-\beta_0)$
Weight	w	w	$\frac{1}{2} - w$	$\frac{1}{2} - w$

  - where u and w are numerically solved for using equations:

$$u^2(3 + 3e^u + 2u - 2ue^u) + \beta_0^2(1 + e^u + 2u - 2ue^u) + \sqrt{u^4 + 14\beta_0^2 u^2 + \beta_0^4}(1 + e^u + u - ue^u) = 0$$

$$w = \left(-u^2 + 6u\beta_0 - \beta_0^2 + \sqrt{u^2 + 14\beta_0 u + \beta_0^2}\right)/24\beta_0 u$$

Figure 3 – Example Model Fit

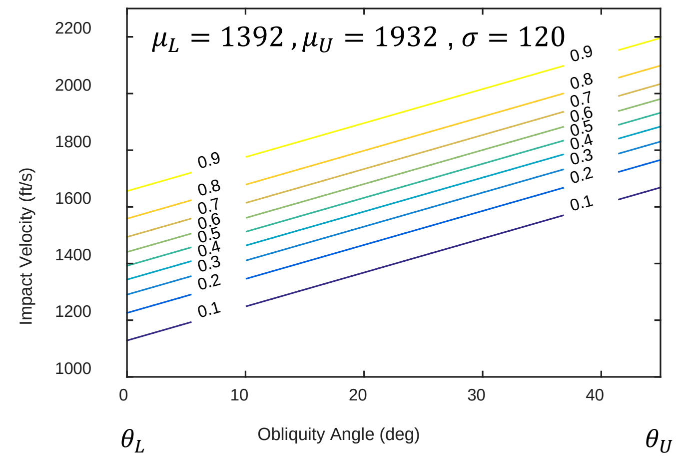
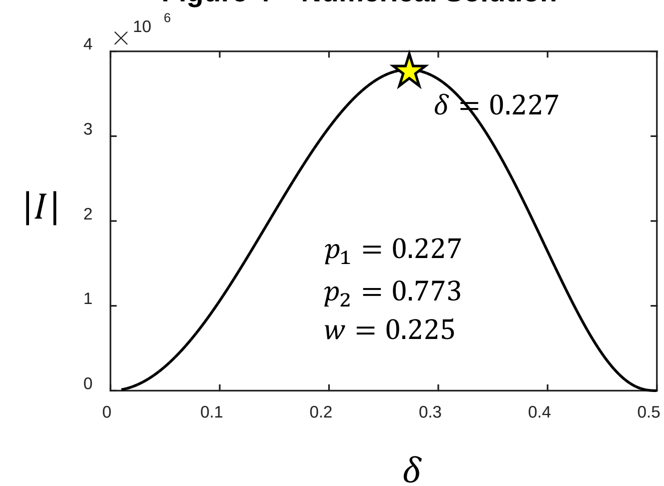


Figure 4 – Numerical Solution

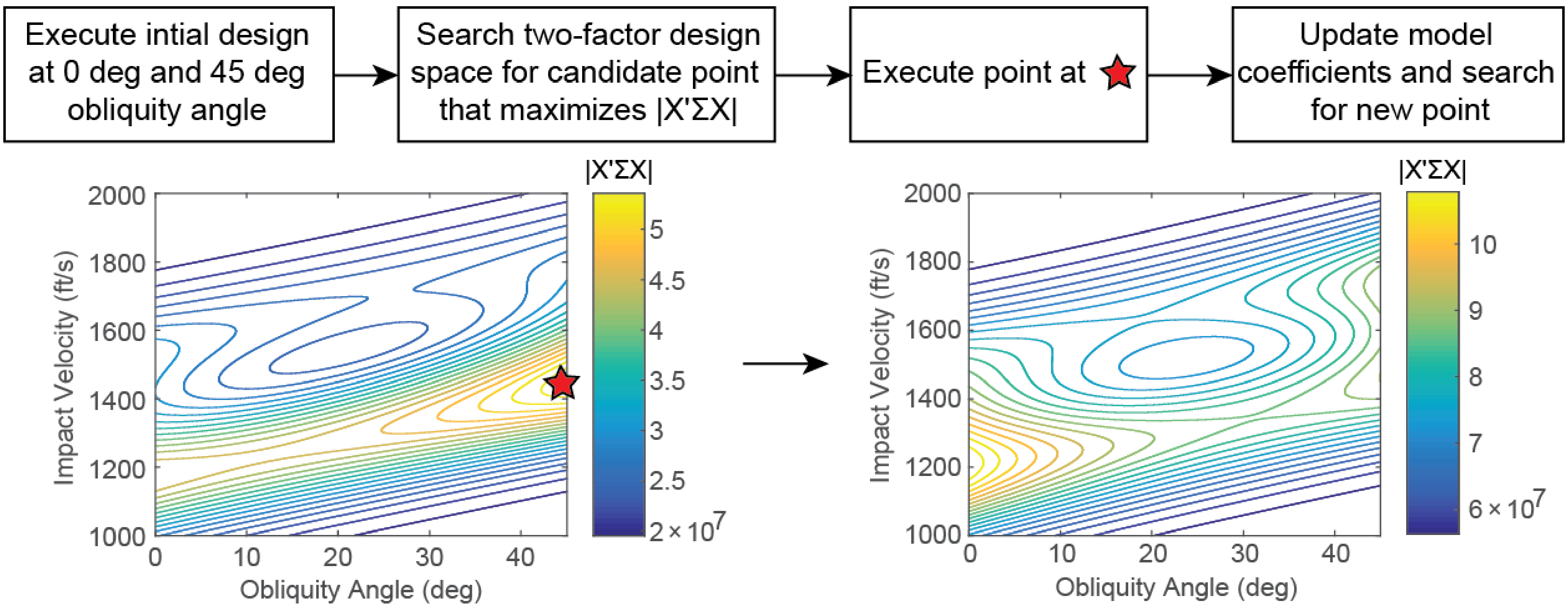


D-Optimal 2-Factor Design Specifies Shots at $V_{22.7}$ and $V_{77.3}$

Jia and Myers, Proceedings of the Annual Meeting of the American Statistical Association, August 2001.

Expanding 3Pod's D-Optimal Search to Two Factors

- **Proposed strategy to implement 3Pod in a two factor space**
  1. Conduct initial design with velocity as the factor at zero degree obliquity
  2. Conduct an additional initial design with velocity as the factor at 45 degree obliquity angle
  3. Select next point by searching velocity settings that maximize the determinant of the fisher information matrix.
     - » Constrain search to velocities at 0 and 45 degree obliquity since we know that is where the 4 point locally d-optimal points is



Theoretical Improvement

- **We can calculate the improvement gained by expanding the search to additional factors, since we can analytically solve for the D-optimal design**
- **Three 30 run designs considered:**

D-optimal Design

Obliquity Angle	
0 deg	45 deg
15 runs (7 runs @ V22.7, 8 runs @ V77.3)	15 runs (7 runs @ V22.7, 8 runs @ V77.3)

Design 1

Obliquity Angle	
0 deg	45 deg
15 runs (7 runs @ V17.6, 8 runs @ V82.4)	15 runs (7 runs @ V17.6, 8 runs @ V82.4)

Design 2

Obliquity Angle		
0 deg	22.5 deg	45 deg
10 runs (5 runs @ V17.6, 5 runs @ V82.4)	10 runs (5 runs @ V17.6, 5 runs @ V82.4)	10 runs (5 runs @ V17.6, 5 runs @ V82.4)

	D-optimal Design	Design 1	Design 2
$\lvert X'\Sigma X \rvert$:	1.5E9	1.4E9	1.0E9
D-efficiency:	1.0	.896	.600

- **These designs are infeasible in practice because we don’t have prior knowledge of coefficients**
  - We must run simulations that include an initial design to determine practical improvement

Simulation Setup

12 run factorial experiment

- Response: D-efficiency
- Factors:
  - Methods
    - 3Pod w/ 1-factor D-optimal search (3Pod-1D)
    - 3Pod w/ 2-factor D-optimal search (3Pod-2D)
    - Dror-Steinberg Method (D-S)
    - Langlie Method
  - Sample Sizes
    - 60, 120

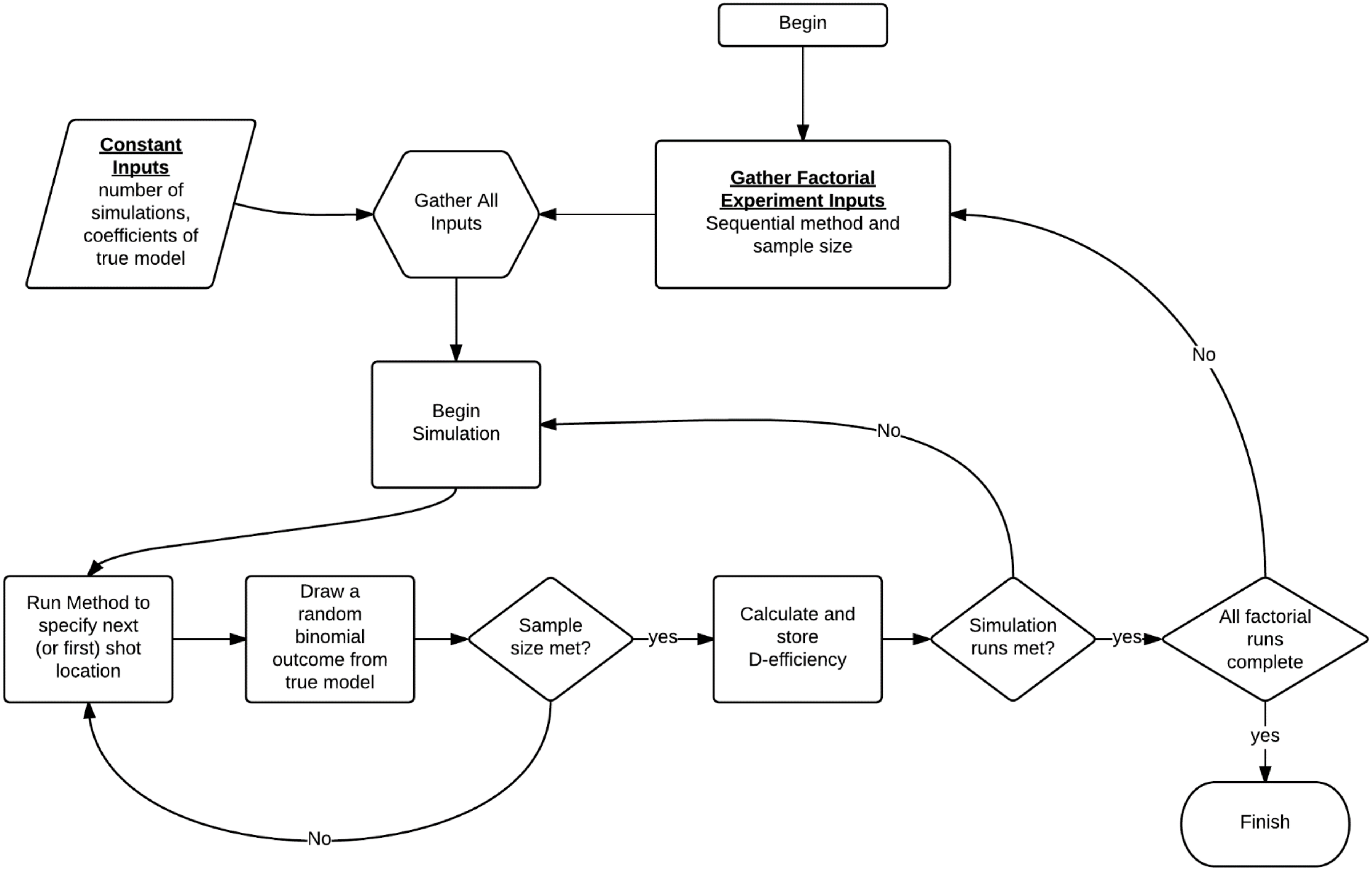
Method Input parameters

- D-S requires prior uniform distributions on model coefficients
- 3Pod requires specification of σ_G and μ_G at 0 and 45 degree obliquity angle
- To make a fair comparison, inputs for each method need to be equivalent

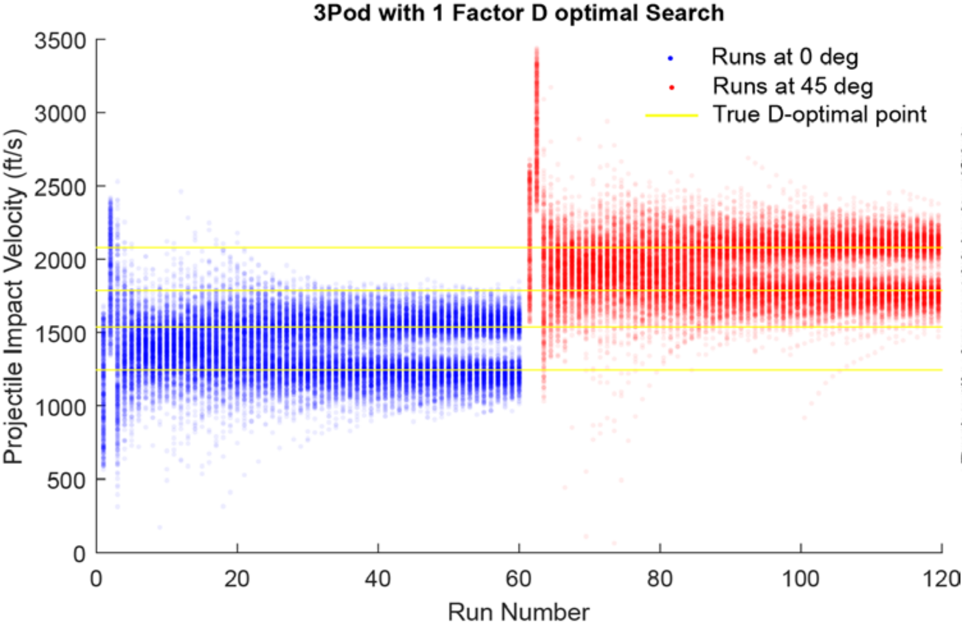
Constant inputs into simulation

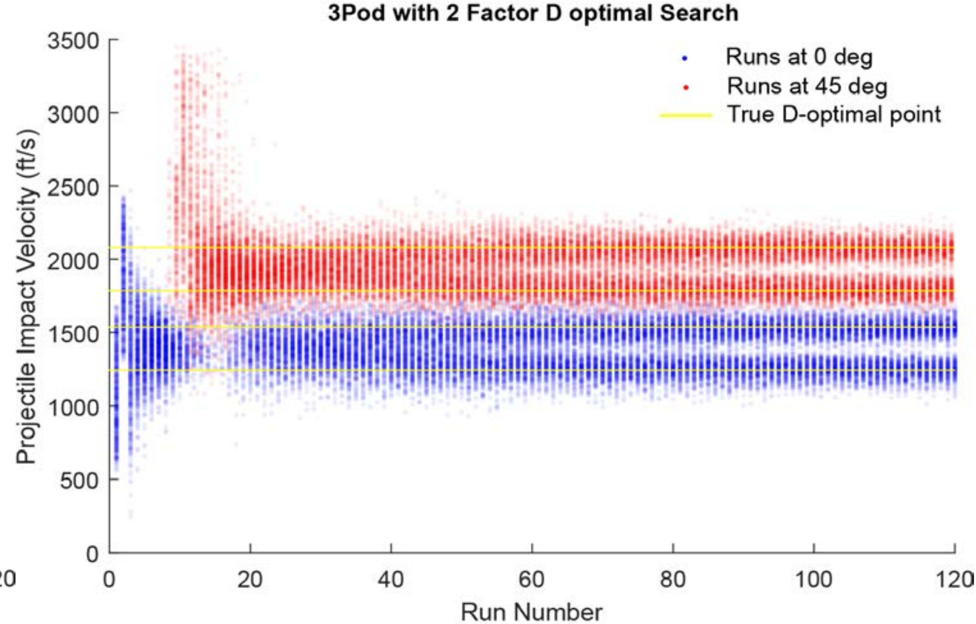
- Assumed true logit model: $b_T = [b_{0T} \quad b_{1T} \quad b_{2T}] = [-11.6 \quad -.1 \quad .0083]$
- Number of simulations per factorial trial: 1,000

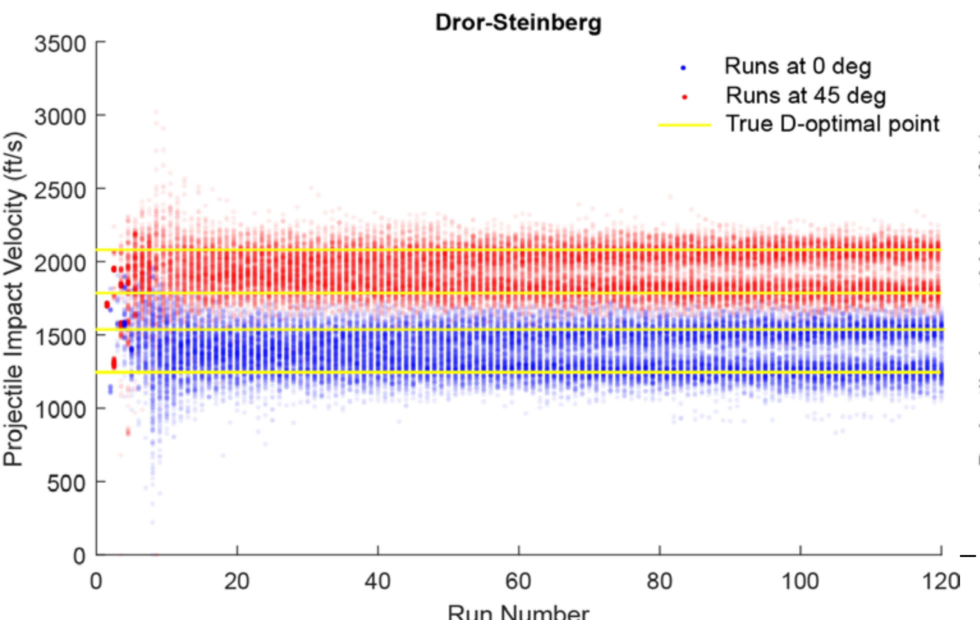
Simulation Setup

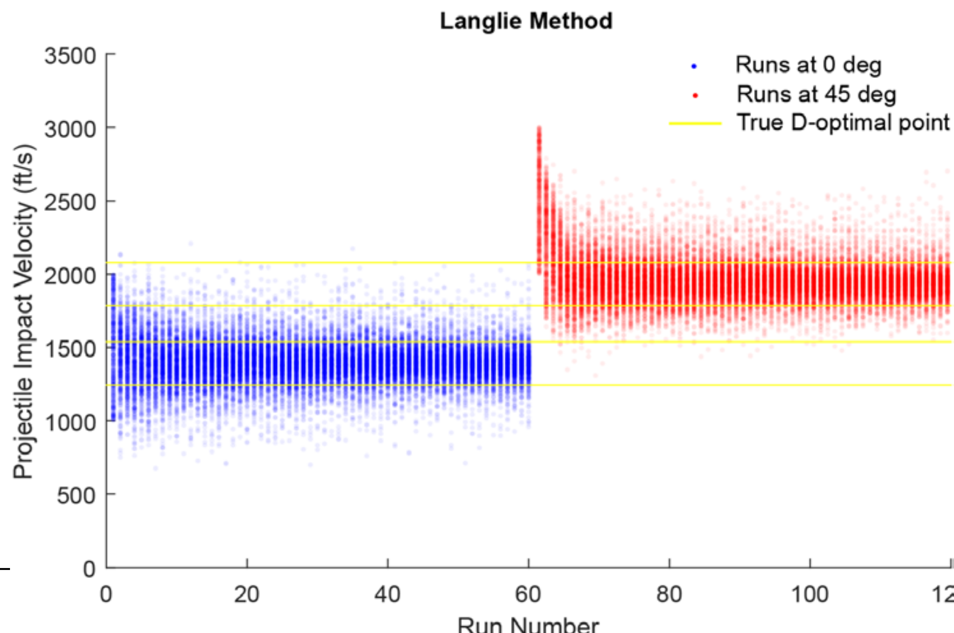


Simulation Results

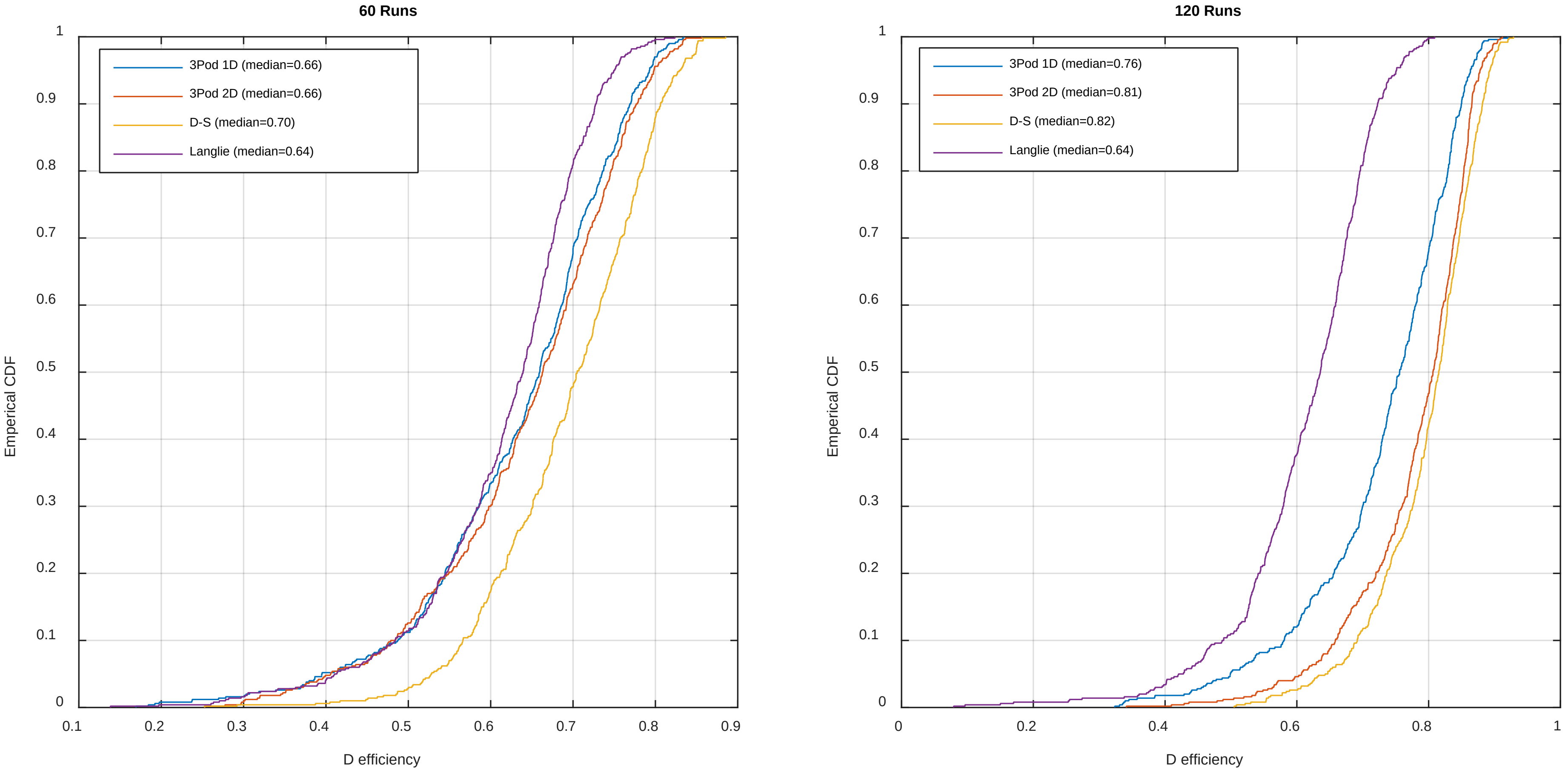








Simulation Results



Recommendations

D-S and 3Pod2D perform best

Further investigation into the practicality, and robustness of D-S is needed