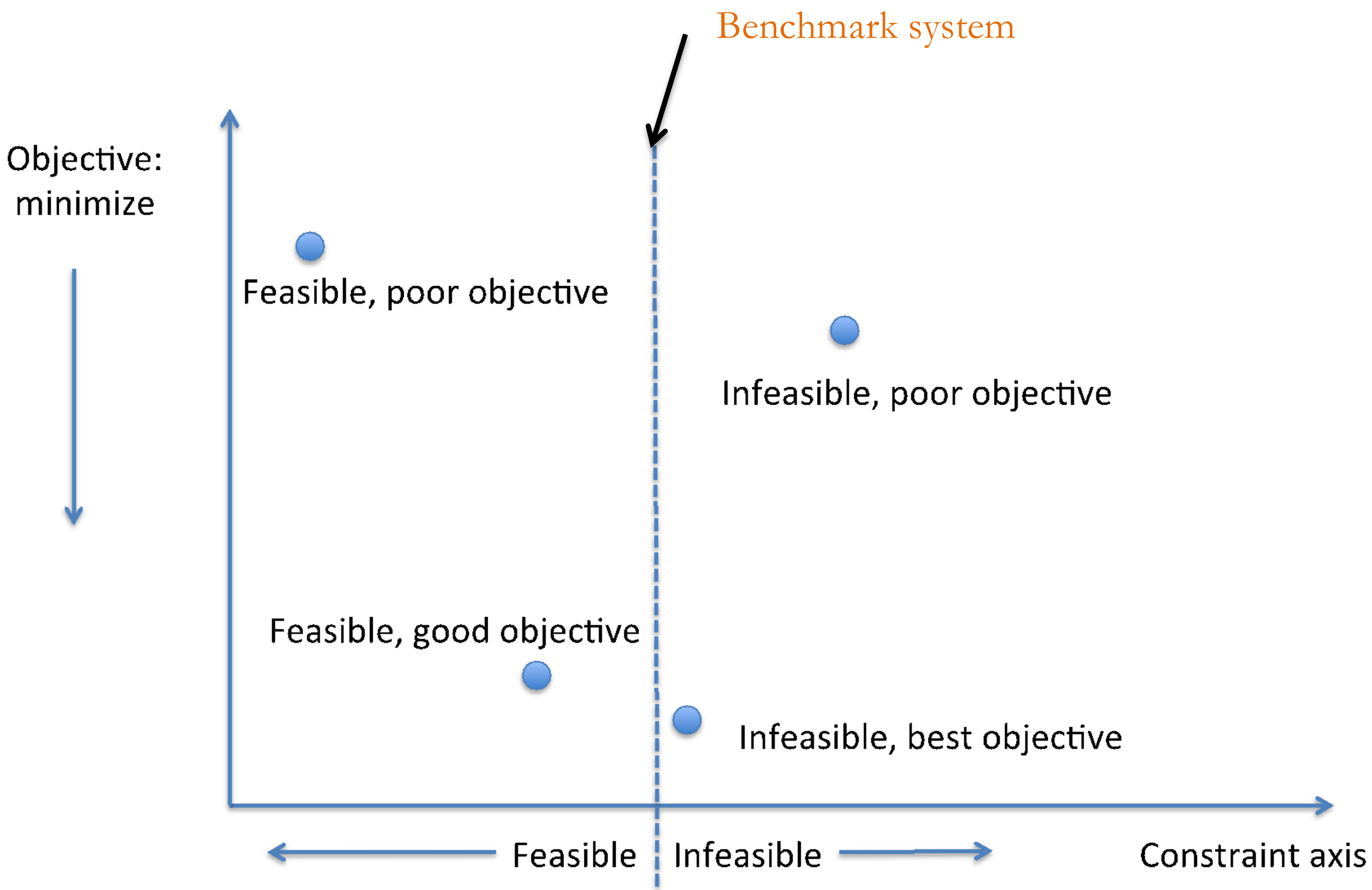
Using Developmental T&E to Inform Operational T&E Decision-Based Analysis

Dashi I. Singham, Ph.D.
Research Associate Professor
Operations Research Department
Naval Postgraduate School
May 2018

Research Ideas

- The decision to implement a new system is often based on a comparison to a benchmark
- When multiple options are available, we can use DT&E to weed out infeasible options
- Use two-stage statistical methods to decide how to allocate effort in OT&E
  - First stage: represents DT&E
  - Second stage: represents OT&E

Choosing the best system configuration



Confidence Intervals

Estimated mean

Estimated standard deviation

$$\left[\bar{x}_n \pm t_{\alpha,n-1}\frac{s_n}{\sqrt{n}}\right]$$

Half-width

- Confidence intervals represent the uncertainty in the mean performance of a system based on *n* samples
- Often assume normality in the data
- The half-width should be small enough to ensure that the variation in the mean estimate is acceptable – choose δ as this acceptable half-width.

Fixed Sampling Rules – Choosing the sample size

Can use standard deviation estimate

$$n \geq \left(\frac{z_\alpha \sigma}{\delta}\right)^2$$

Desired half-width of confidence interval

- If a variance estimate is available, can calculate ahead of time how many samples should be taken to obtain a confidence interval with a half-width smaller than δ.
- Challenges:
  - hard to choose δ
  - n might be large
  - Variance estimate might not be available, but can be estimated if samples available

First stage screening process

- Let benchmark system that defines minimum performance be μ.
- Determine the probability of a system having performance better than the benchmark using p-values from first stage (DT&E)

$$p_i = F_{t_{n-1}}\left(\frac{\overline{X}_i - \mu}{\hat{\sigma}_i/\sqrt{n}}\right)$$

- Eliminate the systems with small p-values

$$p_i \leq \alpha$$

- Test remaining systems in the second stage.

Benchmarked sample size calculation

Instead of using the fixed-sample rule:

$$n \geq \left(\frac{z_\alpha \sigma}{\delta} \right)^2$$

Calculate sample sizes needed to distinguish from the benchmark using first stage information:

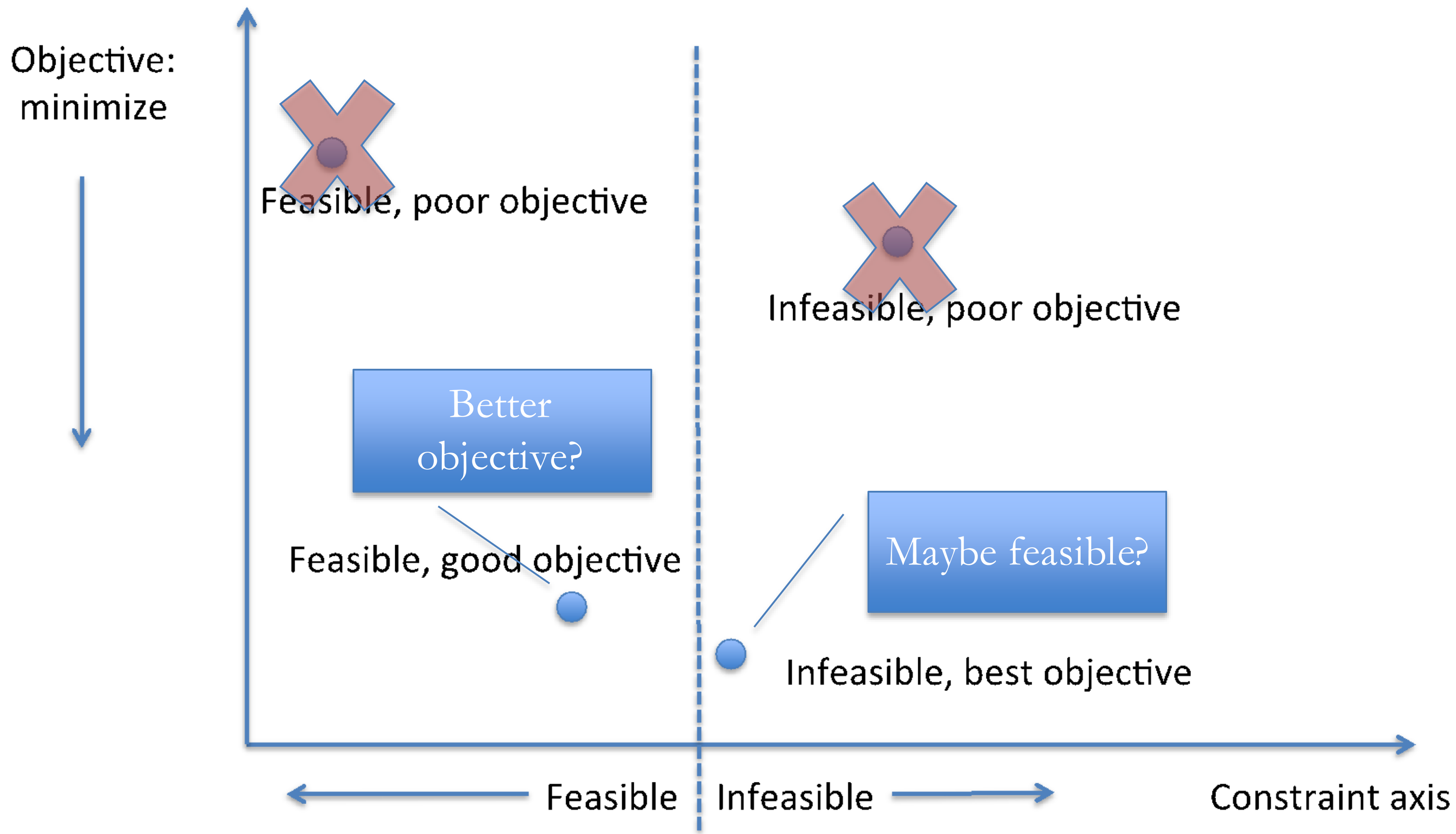
Variance estimate

$$n_i \geq \frac{t^2_{n-1,\alpha} \tilde{\sigma}^2_i}{|\overline{X}_i - \mu|^2}$$

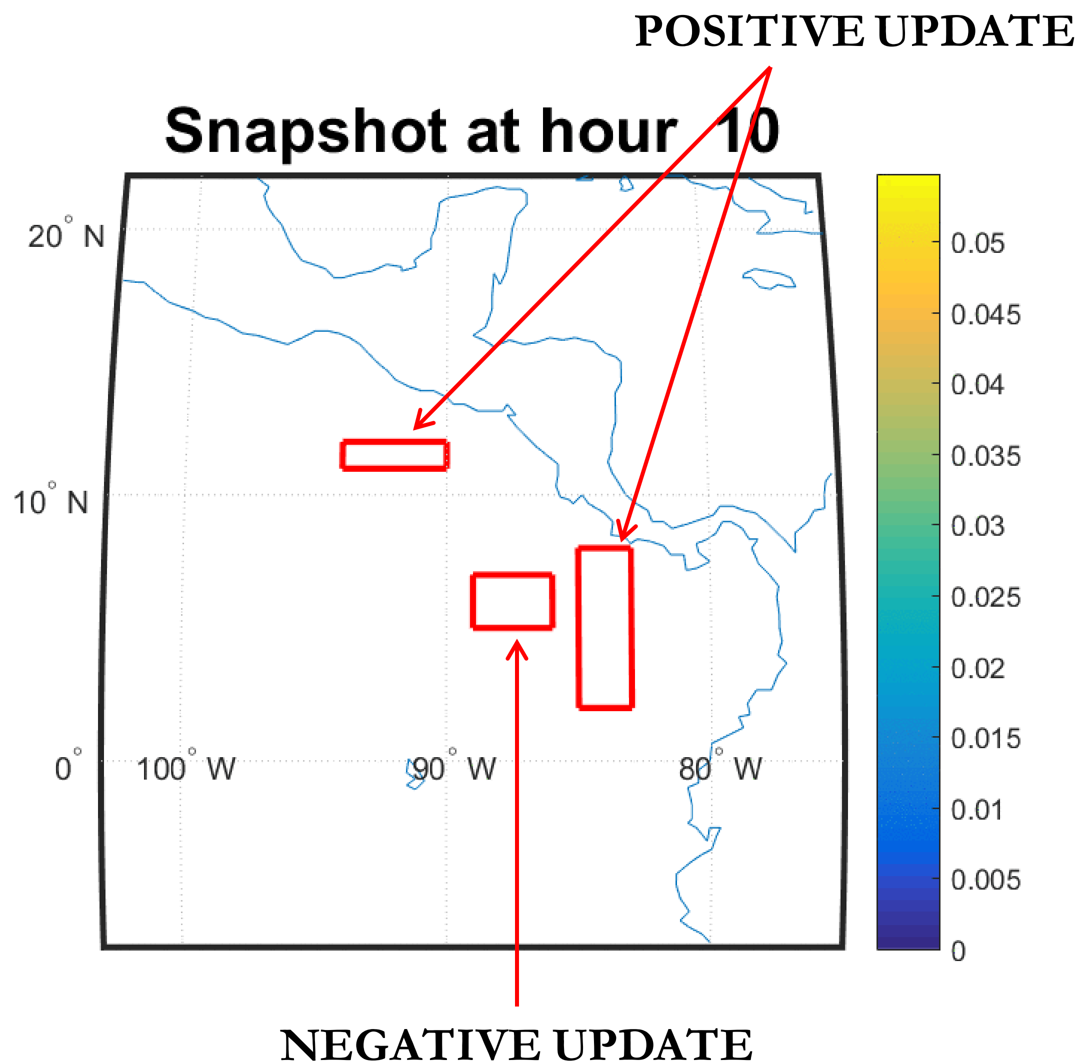
Difference in first stage system *i* from benchmark system μ

Systems with performance close to the benchmark will require more samples.

Choosing the best system



Sensors tracking moving target



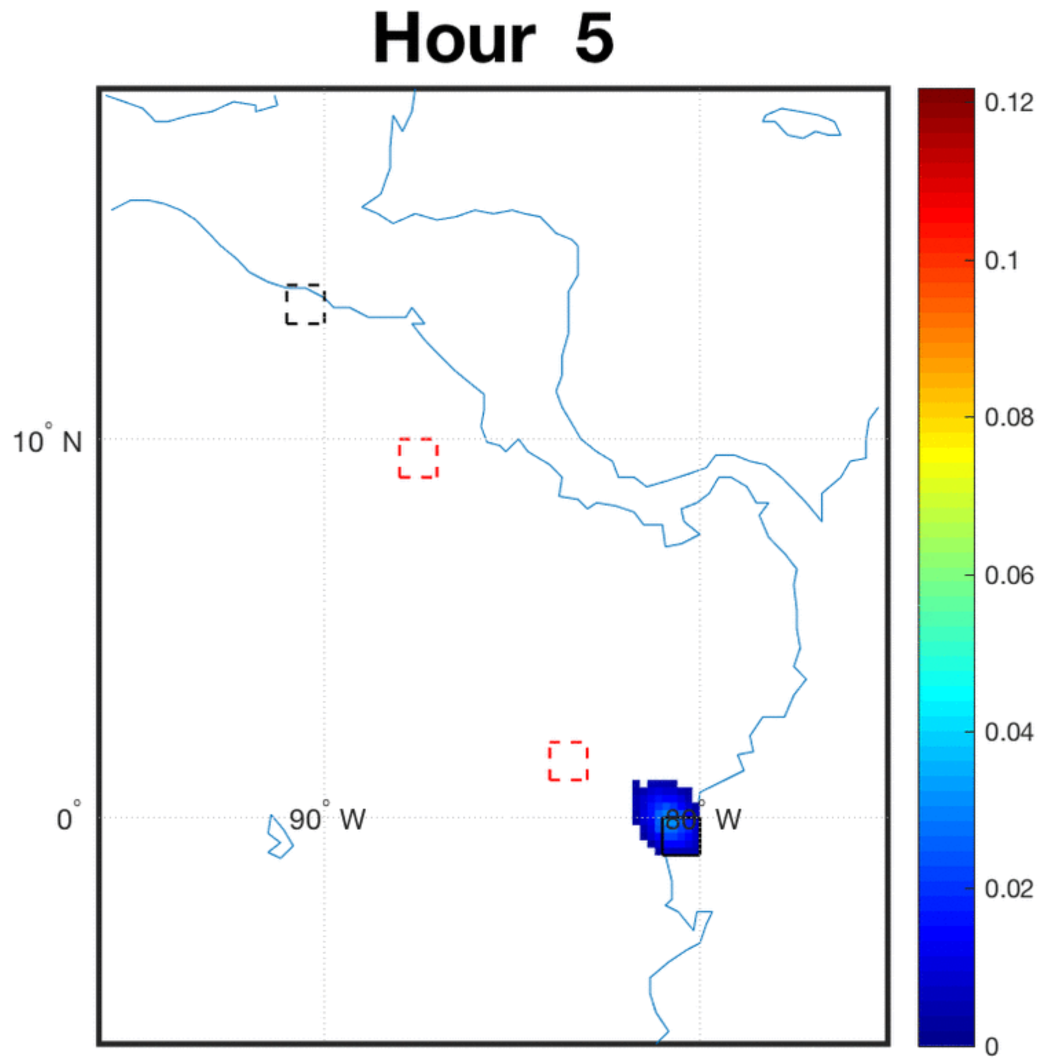
System configurations

System Configuration	Coverage Width (each sensor, degrees)
Lynx (single) benchmark	0.72
Dual20	0.20
Dual30	0.30
Dual35	0.35
Dual37	0.37
Dual38	0.38
Dual39	0.39
Dual40	0.40
Dual50	0.50

Dual sensor system is **feasible** if it has a higher probability of detection than the single Lynx.

A dual sensor system is **optimal** if it has the smallest single sensor coverage area out of all feasible systems (assuming smaller coverage is lower cost).

Sensor configuration example



First stage results – 30 replications for each config

System Configuration	First Stage Mean	p-value	Number of Samples	Proportion of Samples for Second Stage
Lynx (single) benchmark	0.1457	---	35	9%
Dual20	0.0408	0	---	---
Dual30	0.0892	0	---	---
Dual35	0.1155	0	---	---
Dual37	0.1273	0	---	---
Dual38	0.1407	0.17	220	58%
Dual39	0.1543	0.96	65	17%
Dual40	0.1542	0.97	57	15%
Dual50	0.2235	1	2	1%

First stage results – 30 replications for each config

System Configuration	First Stage Mean	p-value	Number of Samples	Proportion of Samples for Second Stage
Lynx (single) benchmark	0.1457	---	35	9%
Dual20	0.0408	0	---	---
Dual30	0.0892	0	---	---
Dual35	0.1155	0	---	---
Dual37	0.1273	0	---	---
Dual38	0.1407	0.17	220	58%
Dual39	0.1543	0.96	65	17%
Dual40	0.1542	0.97	57	15%
Dual50	0.2235	1	2	1%

Second stage results

System Configuration	First Stage Mean	p-value
Lynx (single) benchmark	0.1437	---
Dual38	0.1404	0
Dual39	0.1473	1
Dual40	0.1542	1
Dual50	0.2279	1

Conclusions and Future Work

- A planned two-stage experiment can potentially save costly OT&E tests by
  - eliminating configurations from DT&E that are likely infeasible
  - re-allocating effort for DT&E
- More testing should be allocated
  - “close to the boundary” of feasibility,
  - to systems with higher variability
  - systems likely to be the optimal/best
- Could directly incorporate cost into second stage allocations