Operator Aided Decision Processes for UAVs in a Stochastic Environment

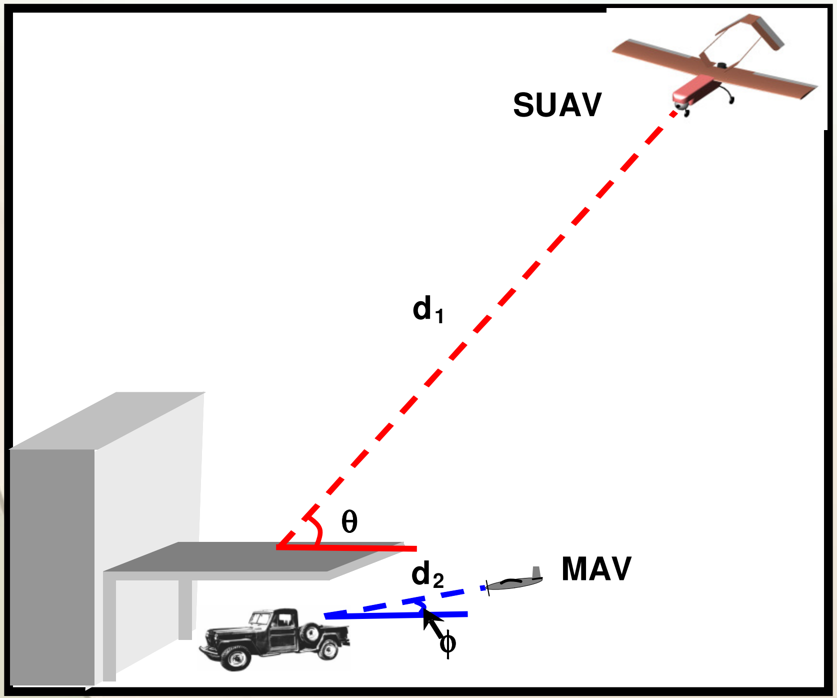
John J. Baker

Anouck Girard, Raymond W. Holsapple, Meir Pachter, Phil Chandler

Air Force Research Lab – AFRL/RBCA
The University of Michigan

COUNTER Scenario

- A small aerial vehicle (SAV) discovers objects of interest (OOI) from a high altitude (~ 1000 ft.)
- The SAV then assigns tours to multiple micro aerial vehicles (MAVs) to investigate the OOI at a lower altitude (~ 100 ft.)
- An operator at a remote ground station is tasked with identifying OOI that have a defining feature (e.g. a weapon or explosive)

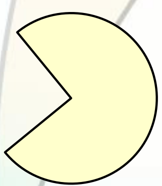


Stochastic Controller

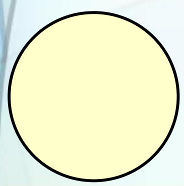
- MAVs investigate the various OOI
- Operator indicates feature or no feature
- Controller calculates an expected reward for a revisit
- MAV will revisit if the expected cost is within the threshold
- The threshold is a function of the following
  - Exisiting fuel (reserve)
  - Operator response
  - Operator delay
  - Expected reward
  - Remaining OOI on the tour

Notation and Terminology

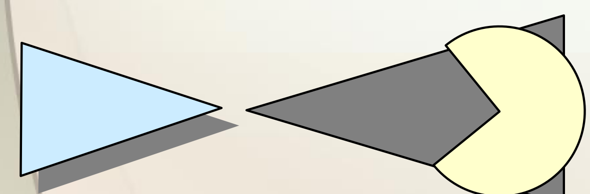
Feature



No Feature



Feature Visible



Feature Not Visible

t = target truth

v = feature visibility

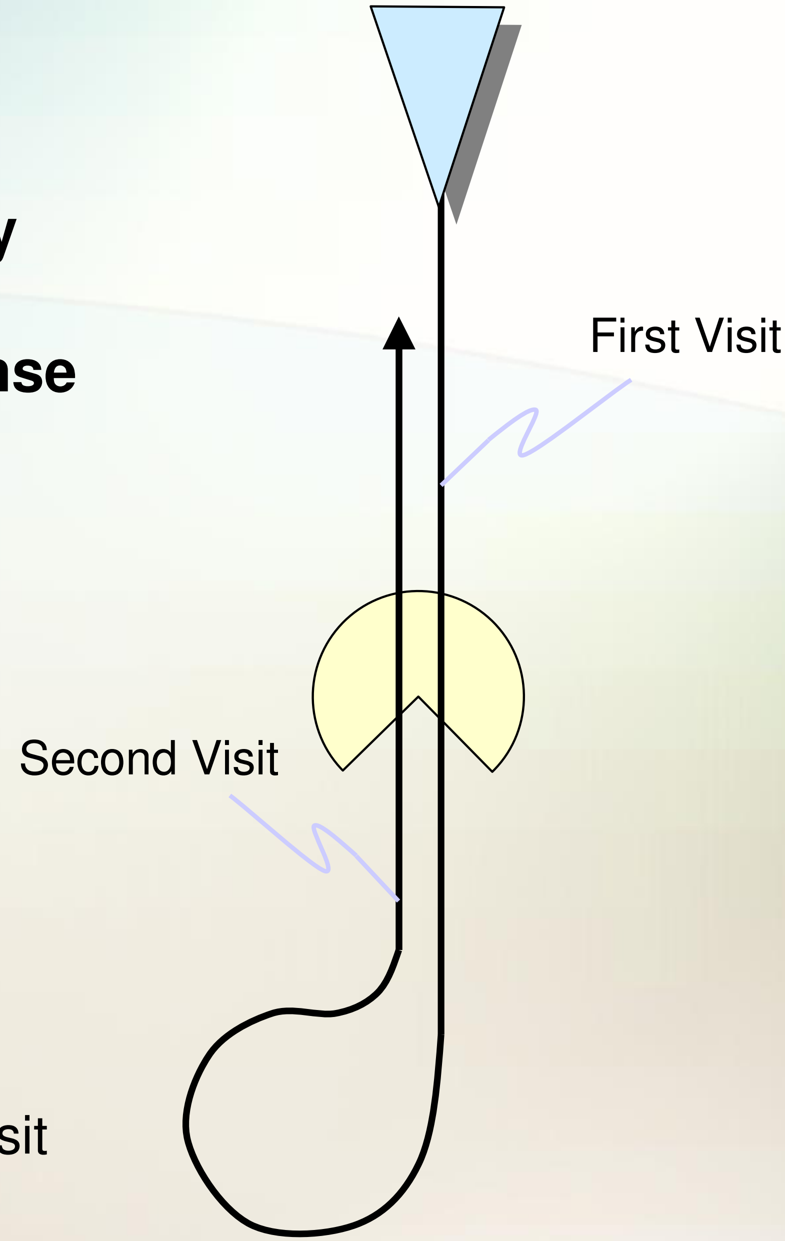
r = operator response

T = true

F = false

subscript 1 : first visit

subscript 2 : second visit



Motivation

The original expected reward function was setup for automatic target recognition† (ATR). This is not inline with the existing COUNTER scenario because the MAVs are not equipped with an ATR sensor.

The ATR overrides any response the operator may make:

$$\boxed{\begin{aligned} &P(\ t{=}T \mid r_1{=}T \cap r_2{=}ATR{:}T\ ) = 1 \\ &P(\ t{=}T \mid r_1{=}F \cap r_2{=}ATR{:}T\ ) = 1 \\ &P(\ t{=}F \mid r_1{=}T \cap r_2{=}ATR{:}T\ ) = 0 \\ &P(\ t{=}F \mid r_1{=}F \cap r_2{=}ATR{:}T\ ) = 0 \end{aligned}}$$

Reasons for not using an ATR sensor include:

- Lighting conditions (shadows, glare)
- Limited onboard computational power
- Noisy data

† Pachter, M., Chandler, P. and Darbha, S., \Optimal Control of an ATR Module Equipped MAV-Human Operator Team," *Proceedings of Cooperative Control and Optimization Conference,* Gainesville, FL, January 2006.

Objective

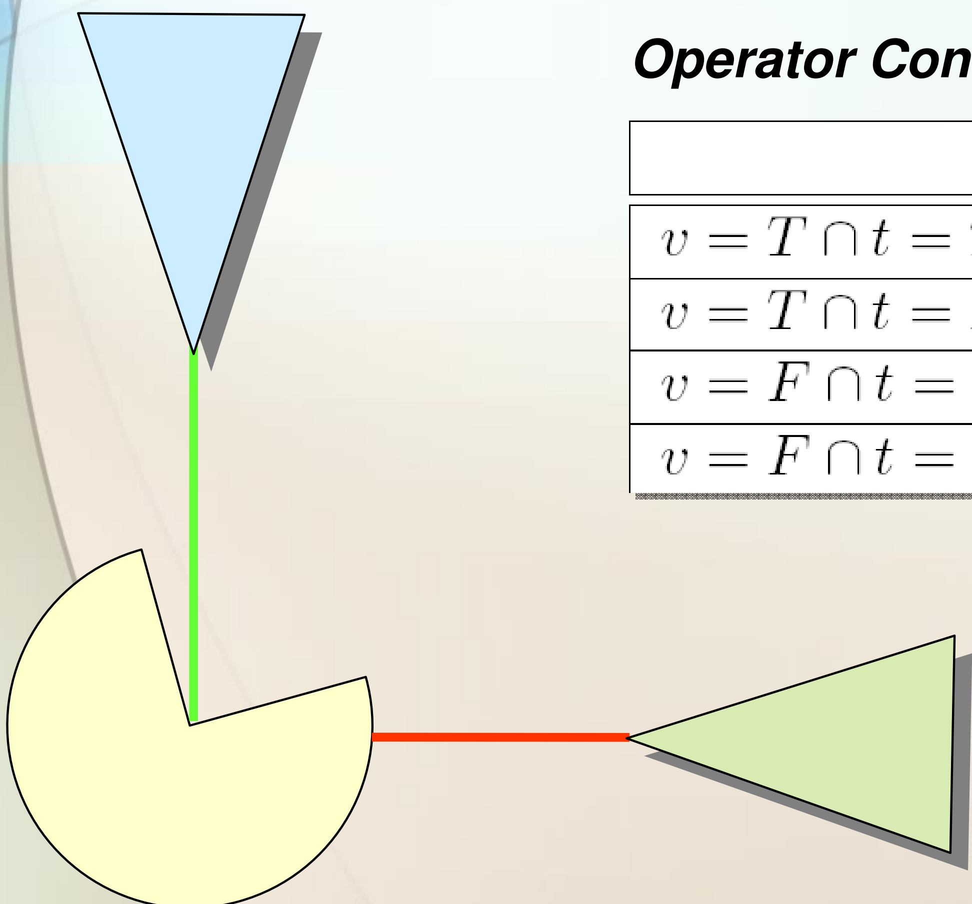
- Design a new reward function using visibility and response probabilities for two visits
- Evaluate various reward methods
- Benchmark these against an operator delay revisit threshold

A Priori Terms

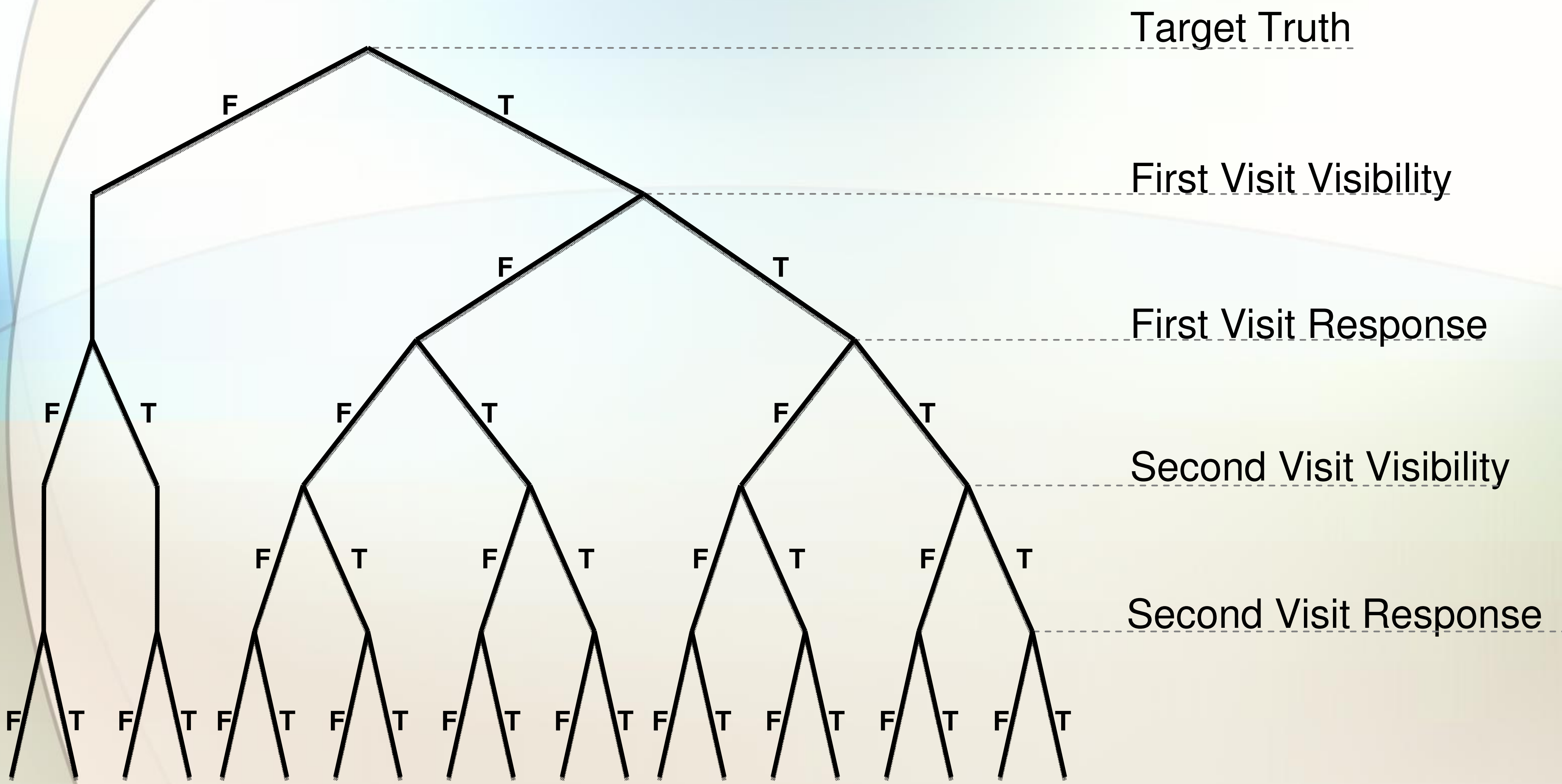
$P(t{=}T)$ .... Probability that an OOI has a feature
P_{to} .... Probability of Detection
$1\text{-}P_{fto}$ .... Probability of False Alarm.
$P_i(\ r \mid v \cap t\ )$ .... Probability of Response given Visibility & Target Truth

Operator Confusion Matrix

	$r = T$	$r = F$
$v = T \cap t = T$	P_{TO}	$1 - P_{TO}$
$v = T \cap t = F$	undef	undef
$v = F \cap t = T$	$1 - P_{FTO}$	P_{FTO}
$v = F \cap t = F$	$1 - P_{FTO}$	P_{FTO}



Probability Outcome Tree



Expected Reward Functions

Let $A = r_1 \cap v_1$ and $B = r_2 \cap v_2$

Method 1: Sum of Reciprocals†

$$Reward = \log\left(\frac{P(t=T|A\cap B)}{P(t=F|A\cap B)} + \frac{P(t=F|A\cap B)}{P(t=T|A\cap B)}\right) - \log\left(\frac{P(t=T|A)}{P(t=F|A)} + \frac{P(t=F|A)}{P(t=T|A)}\right) \quad 1$$

Method 2: Mutual Information†

$$\begin{aligned} Reward \quad = \quad & \left(P(t=T\cap A\cap B)\cdot log\left(\frac{P(t=T|A\cap B}{P(t=T)}\right) + P(t=F\cap A\cap B)\cdot log\left(\frac{P(t=F|A\cap B}{P(t=F)}\right)\right) \\ - \quad & \left(P(t=T\cap A)\cdot log\left(\frac{P(t=T|A}{P(t=T)}\right) + P(t=F\cap A)\cdot log\left(\frac{P(t=F|A}{P(t=F)}\right)\right) \end{aligned} \quad 2$$

Method 3: Discrete Values

Discrete sub-reward values chosen for each outcome

† Girard, A., Pachter, M. and Chandler, P., \Decision Making Under Uncertainty and Human Operator Model for Small UAV Operations," *Proceedings of AIAA Guidance, Navigation and Control Conference,* Keystone, CO, August 2006.

Probabilities

Used as gains applied against the individual rewards:

$$
\begin{aligned}
\underline{P(r_2 \cap v_2 | r_1 \cap v_1)} \quad &= \quad [P(r_1|v_1 \cap T) \cdot P(r_2|v_2 \cap t = T) \cdot P(v_1 \cap v_2|t = T) \cdot P(t = T) ... \\
&+ \quad P(r_1|v_1 \cap t = F) \cdot P(r_2|v_2 \cap t = F) \cdot P(v_1 \cap v_2|t = F) \cdot P(t = F)] ... \\
&\div \quad [P(r_1|v_1 \cap t = T) \cdot P(v_1|t = T) \cdot P(t = T) ... \\
&+ \quad P(r_1|v_1 \cap t = F) \cdot P(v_1|t = F) \cdot P(t = F)]
\end{aligned}
\qquad (3)
$$

Values used inside of the individual reward functions:

$$
\begin{aligned}
\underline{P(t|r_1 \cap v_1)} \quad &= \quad \frac{P(t \cap r_1 \cap v_1)}{P(r_1 \cap v_1)} \\
&= \quad \frac{P(r_1|t \cap v_1) \cdot P(v_1|t) \cdot P(t)}{P(t = T \cap r_1 \cap v_1) + P(t = F \cap r_1 \cap v_1)} \\
&= \quad \frac{P(r_1|t \cap v_1) \cdot P(v_1|t) \cdot P(t)}{P(r_1|t = T \cap v_1) \cdot P(v_1|t = T) \cdot P(t = T) + P(r_1|t = F \cap v_1) \cdot P(v_1|t = F) \cdot P(t = F))}
\end{aligned}
\qquad (4)
$$

$$
\begin{aligned}
\underline{P(t|r_1 \cap v_1 \cap r_2 \cap v_2)} \quad &= \quad \frac{P(t \cap r_1 \cap v_1 \cap r_2 \cap v_2)}{P(r_1 \cap v_1 \cap r_2 \cap v_2)} \\
&=^{\S} \quad P(r_1|v_1 \cap t) \cdot P(r_2|v_2 \cap t) \cdot P(v_1 \cap v_2|t) \cdot P(t) ... \\
&\div \quad [P(r_1|v_1 \cap t = T) \cdot P(r_2|v_2 \cap t = T) \cdot P(v_1 \cap v_2|t = T) \cdot P(t = T) ... \\
&+ \quad P(r_1|v_1 \cap t = F) \cdot P(r_2|v_2 \cap t = F) \cdot P(v_1 \cap v_2|t = F) \cdot P(t = F)]
\end{aligned}
\qquad (5)
$$

Value

$$Val_T = \sum P(r_2 \cap v_2 \mid r_1 \cap v_1)_m * Reward_m \quad \text{For } m= 1,\ldots,16$$

$$Val_F = \sum P(r_2 \cap v_2 \mid r_1 \cap v_1)_n * Reward_n \quad \text{For } n = 17,\ldots,20$$

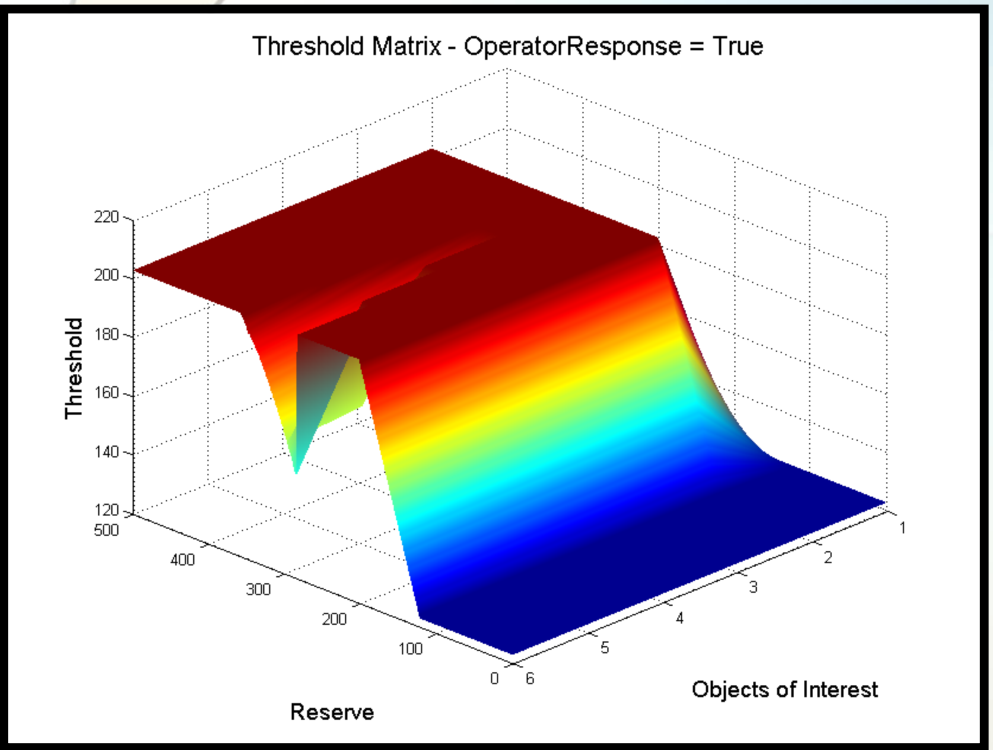
Threshold Matrix ($P_T Val_T$ or $P_F Val_F$, Targets, Reserve, Operator Delay)

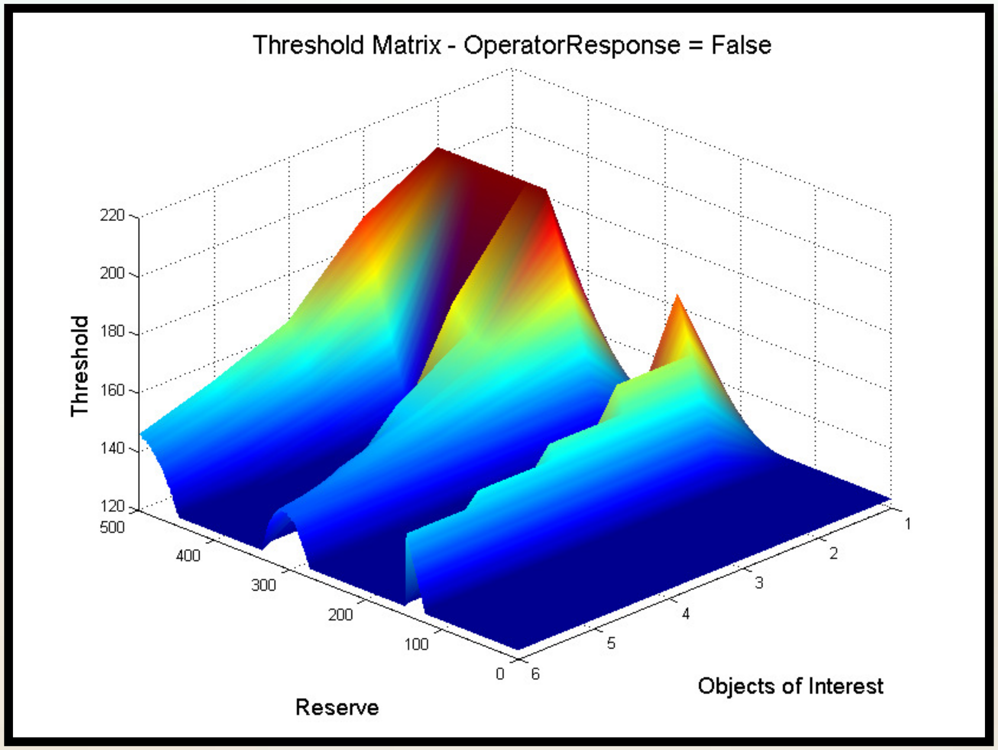
From the threshold matrices we can obtain two surfaces. One for each of the response that operator can make on the first visit.

The threshold value is determined by how much reserve fuel remains, the operator delay, and how many OOI are left to visit.

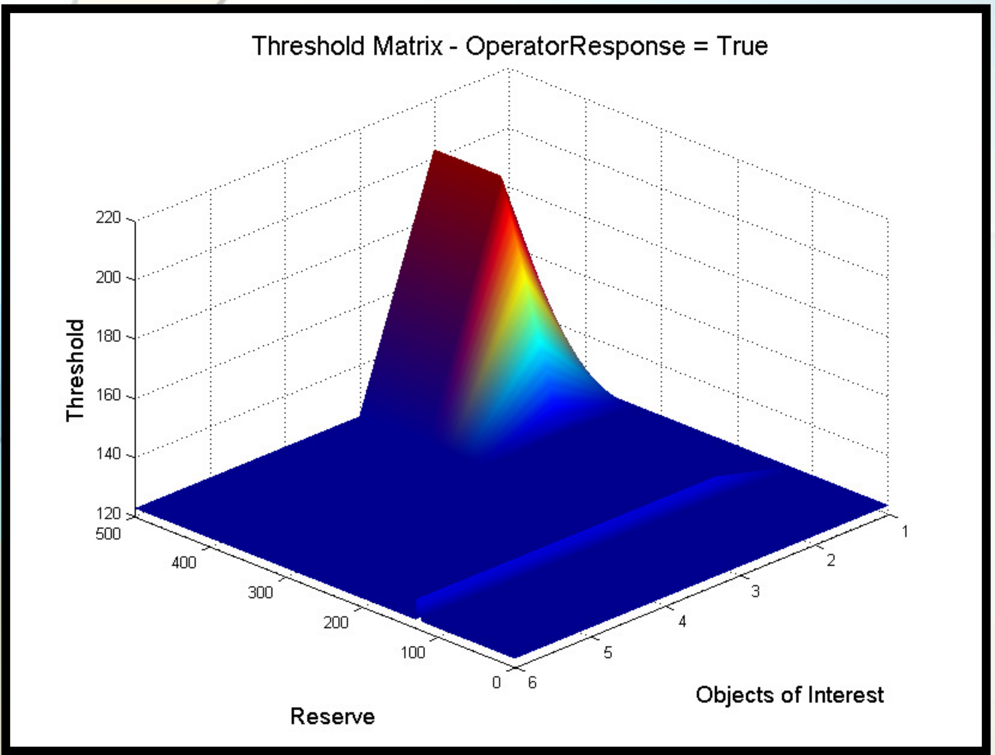
Essentially, if the expected revisit cost is less then the threshold the MAV will revisit the OOI.

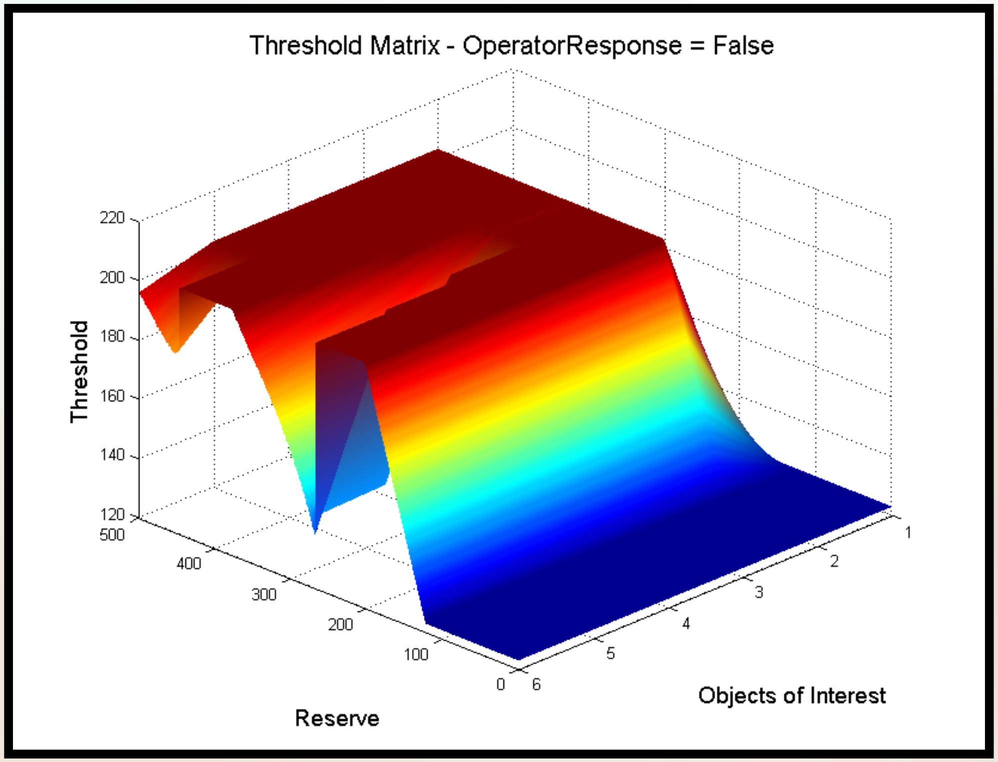
Method 1 Threshold Matrices



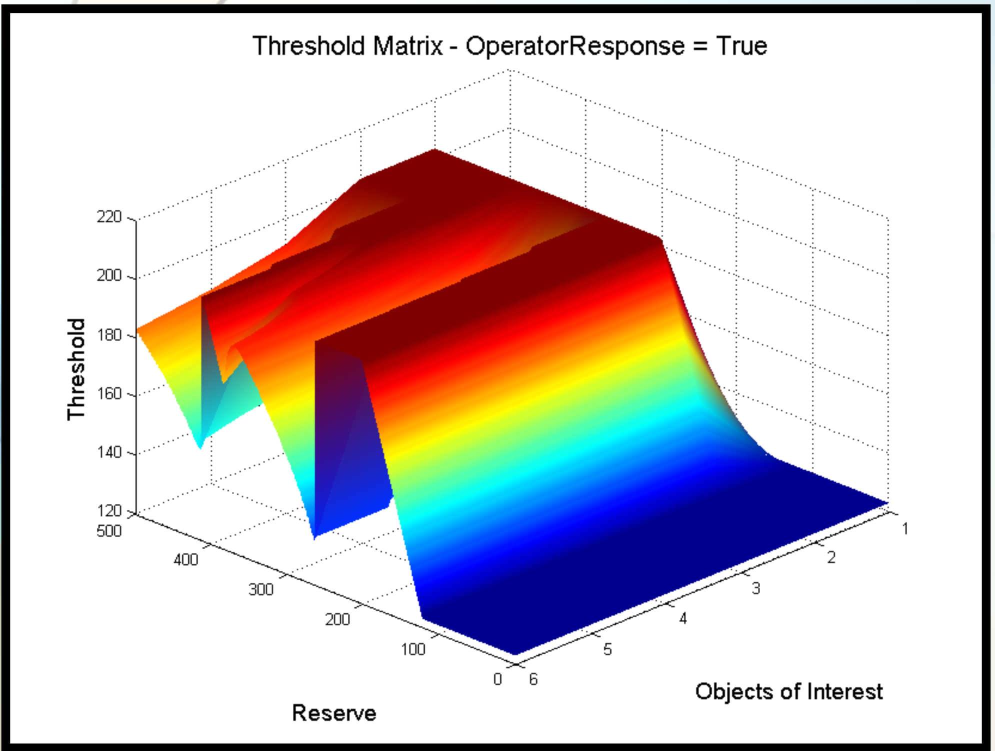


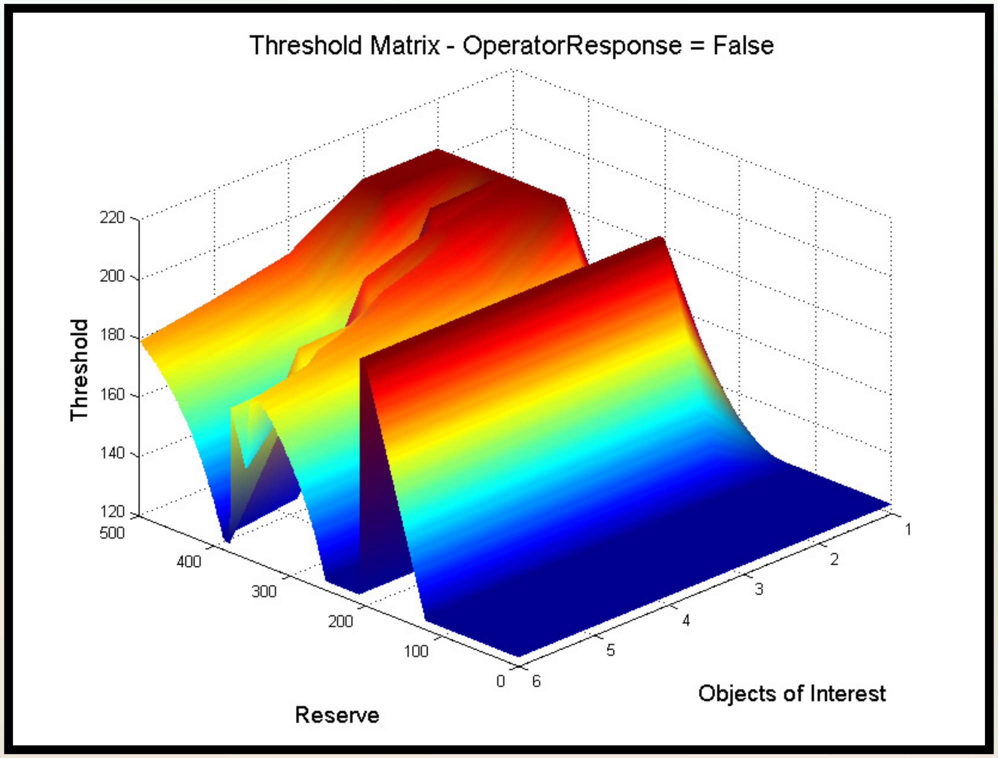
Method 2 Threshold Matrices





Method 3 Threshold Matrices





Measure of Performance

3 methods + benchmark

100 Monte Carlo simulations each (in MultiUAV2)

For each trial

If a revisit occurred

If the **t=T** and [**(r_1=T & r_2=F) or (r_1=F & r_2=T)]**
TrueScore++

If the **t=F** and **(r_1=F & r_2=F)**
FalseScore++

Determine mean value for scores
Normalize the scores and sum for overall score

Results and Conclusions

- Negative reward values causes saturation in the threshold function
- Large volume under a threshold surface for one case and small volume for the other will cause bias

	Benchmark	Method 1	Method 2	Method 3
True Score Mean	0.190	0.610	0.500	0.760
True Std Dev	0.419	0.803	0.674	0.842
False Score Mean	0.410	0.070	2.080	0.520
False Std Dev	0.911	0.432	2.977	1.453
Total Score	0.600	0.680	2.580	1.280
Std Dev	0.943	0.863	2.875	1.531
Adj. True	1.900	6.100	5.000	7.600
Adj. False	0.456	0.078	2.311	0.578
Adj. Total	2.356	6.178	7.311	8.178

- Method 2 had the best "False Score"
- Method 3 had the best " True Score"
- Method 3 had the best overall adjusted score

Reward Optimization

- The reward function outputs two values
  - Expected Reward given r_1=T
  - Expected Reward given r_1=F
- An optimization function could replace the reward function
  - It would modulate the expected rewards seeking optimal values
  - This would customize the controller for the mission at hand
  - This is feasible for scenarios where 24/7 coverage is required
- Issues
  - Absolute data about target truth is not known
    - May require a calibration stage with mock targets
  - Recalibration might be necessary for individual operators
    - The reward function is also sensitive to this

Future Work

- Improve the values within the discrete reward method
- Refine the information theory reward functions
- Conduct flight tests to measure the performance of the stochastic controller in a real stochastic environment
- Develop an optimization function
- Further development on the operator model