# Towards Adaptive Training of Agent-based Sparring Partners for Fighter Pilots

Brett Israelsen [1] Nisar Ahmed [1] Kenneth Center [2] Roderick Green [2] Winston Bennett Jr. [3]

[1]University of Colorado, Boulder, and
Cooperative Human-robot Intelligence (COHRINT) Lab, and
Research and Engineering Center for Unmanned Vehicles (RECUV)

[2]Orbit Logic Incoporated

[3]AFRL Warfighter Readiness Research Division, Wright Patterson AFB

# Motivation & Background

# Motivation

- Pilots need recurring training
- Physical training is expensive
- The "Not-so-grand Challenge" [4]
- Artificial Intelligences (AIs) should:
  - Remain challenging by adapting
  - Assess and adapt to skill level
  - Can be used in Live-virtual-constructive (LVC) simulations

# Our Contributions

**Vision: In the future Artificial Intelligences will be partners for training humans**

Challenges

1. AIs are trained to optimize engagement metrics which are black-box
2. Training AI can be expensive
3. Engagement outcomes are volatile
4. Humans exploit weaknesses of AIs

Novel Contributions:

- Provide model of objective function, with uncertainty, to AI
- Gaussian Process Bayesian Optimization (GPBO)
- Hybrid Repeat/Multi-point Sampling (HRMS) GPBO

# Background & Previous Work

Previous Work:

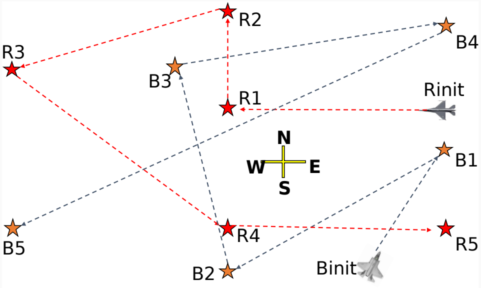
- Real-time air combat AI [5]
- Optimizing aerial combat [13, 12, 19, 7]
- Adaptive gaming agents [1, 11, 14]

Shortcomings:

- Don't take expense of training into account
- Focused on strategic level – not individual behaviors
- Prediction of outcomes and uncertainty in un-explored configurations

# Methodology

# Framing the Problem




- 1 on 1 aerial combat between blue and red
- AI has tunable behaviors
  - Intercept Speed (intspeed)
  - Launch Delay (launch)
  - Weapon Select Delay (select)
  - Min/max azimuth (min/max_az)
- Record engagement outcome metrics
  - Time to kill (TTK)
  - Energy Management
  - Mission objectives met
  - Time on offense/defense
- Limited to 300 seconds
- Terminates if red/blue is eliminated

# Framing the Problem (continued)

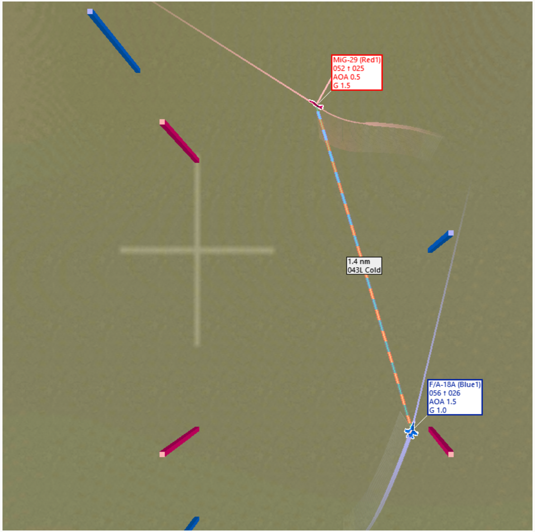
**Problem Statement: Efficiently train parametrized AI pilot to spar with human pilots**

- Investigate Blue:intspeed and Blue:launch
- Optimize TTK:

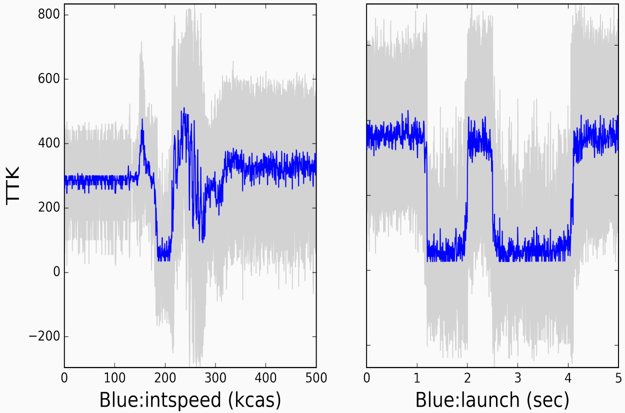
$$
\begin{aligned}
TTK &= SimTime \quad \text{if Blue survived} \\
TTK &= 600 - SimTime \quad \text{otherwise}
\end{aligned}
$$

- Two AIs: Blue$\rightarrow$ adaptive, Red$\rightarrow$ static
- Only require AI to have behavioral parameters
- Optimize and model TTK simultaneously

# Video – Typical non-optimized engagement



# The outcome TTK metric is volatile



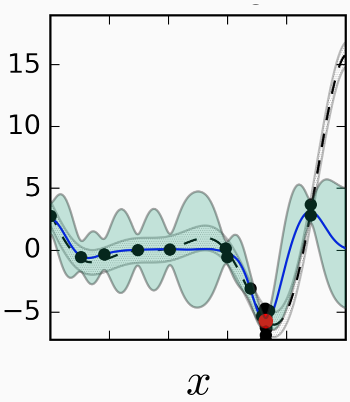

- High noise
- Discontinuities
- Noise magnitude varies
- Volatile due to: weather, sensor noise, human error, and others

# Bayesian Optimization

- **Bayesian Optimization = Optimizing a (black-box) function while learning about it.**
- Bayesian optimization → surrogate or model-based
- Useful for expensive functions
- Acquisition functions selects next location
- Explore/exploit

**Algorithm 1** Standard Bayesian Optimization

```
Initialize surrogate function with n training inputs
while termination criteria not met do
    x_t = argmax_X a(f(X))
    Evaluate y(x_t)
    Add y(x_t) and x_t to the surrogate function f(X) and update
end while
```



# Mathematical Formulation – Surrogate Function

- **Surrogate: global model of black-box function**
- Gaussian Process [16]

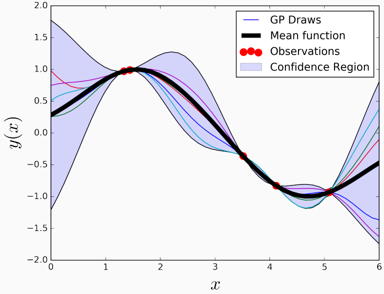
$$
\begin{aligned}
f(x) &\sim \mathcal{GP}(m(x), k(x, x')) \\
m(x) &= \mathbb{E}[f(x)] \\
k(x, x') &= \mathbb{E}[(f(x) - m(x))(f(x') - m(x'))]
\end{aligned}
$$

Inference:

$$
\begin{aligned}
f_* | X_*, X, f &\sim \mathcal{N}(m, \sigma^2) \\
\mu_c(X_*) &= K(X_*, X)\, K(X, X)^{-1} f \\
{}^2_c(X_*) &= K(X_*, X_*) \\
&\quad - K(X_*, X)\, K(X, X)^{-1} K(X, X_*))
\end{aligned}
$$

- Màtern Kernel:

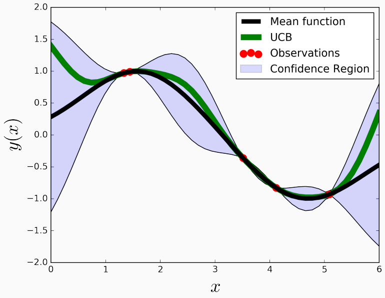
$$
k_{\nu=3/2}(r) = \left(1 + \frac{\sqrt{5}r}{l}\right) \exp\left(-\frac{\sqrt{3}r}{l}\right)
$$



# Mathematical Formulation – Acquisition Function

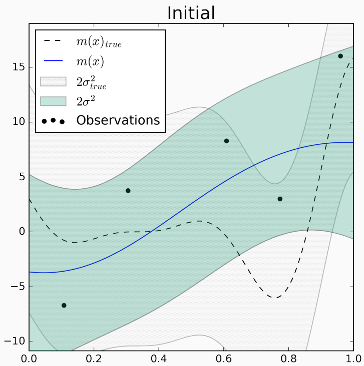
- Next experiment at $\text{argmax}_x \, a(\cdot)$
- Use global optimizer – DIRECT [9]
- **Cheaper to optimize $a(\cdot)$ than the objective function**
- There are many acquisition functions:
  - Expected Improvement (EI)[10]
  - Upper Confidence Bound (UCB)[18]
  - Thompson Sampling (TS)
- Example: Upper Confidence Bound (UCB)

$$a(\cdot)_{\text{UCB}(x,\beta)} = \mu(x) + \beta\sigma(x)$$



# Practical Considerations

- Bootstrapping GP (i.e. $n = 10d$) [8]
- Need "good" surrogate to guide the optimization
- Also, helps give model of outcomes
- The GP parameters need to be learned, using MLE or MAP or X-validation



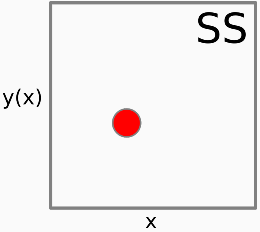
# Persisting Challenges and Existing Methods

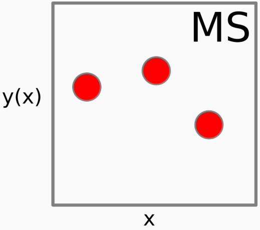
Existing Methods:

- Single Samples (SS) are standard
- Multi-point sampling (MS)
- Batch $a(\cdot)$ – qEI, GP-UCB-PE, and TS [17, 2]

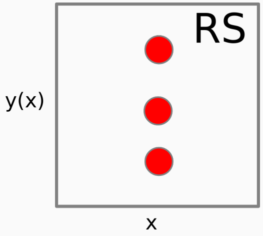
Persisting challenges:

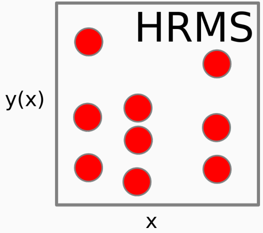
- Hope GP stays "good"
- Trade-off: little data (expensive functions) vs more data (good surrogate)
- Sensitive to low observation:noise ratio



# Hybrid Repeat/Multi-point Sampling




- Repeated Sampilng (RS) [15, 3]
- Use both RS and MS simultaneously
- Local and Global knowledge every iteration
- Sampling strategy $\implies$ use any batch $a(\cdot)$

# HRMS (continued)

- What makes this method work? We still want to investigate further.
- RS causes singular covariance matrix
- Hard to perform inference – requires matrix inverse
- RS forces the noise hyperparameter to be relevant

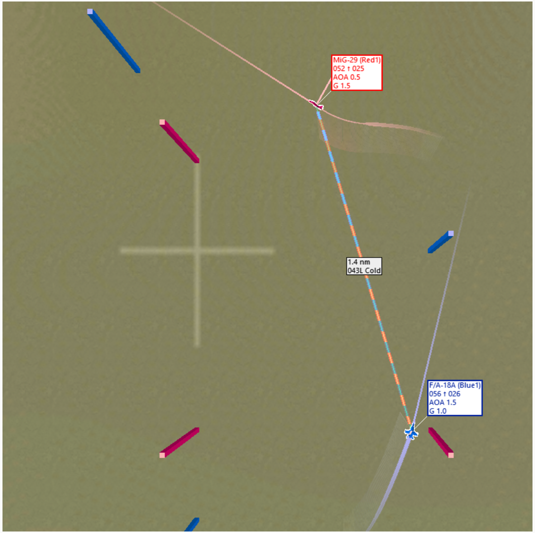
$$\ln p(y|X, \theta, k_i) = -\frac{1}{2}(y - m)^T K_y(\theta, k_i)^{-1}(y - m) - \frac{1}{2}\ln |K_y(\theta, k_i)| - \frac{D}{2}\ln 2\pi$$

$$\mathrm{K}_y(\theta, k_i) = K(\theta, k_i) + \sigma_n^2 I$$

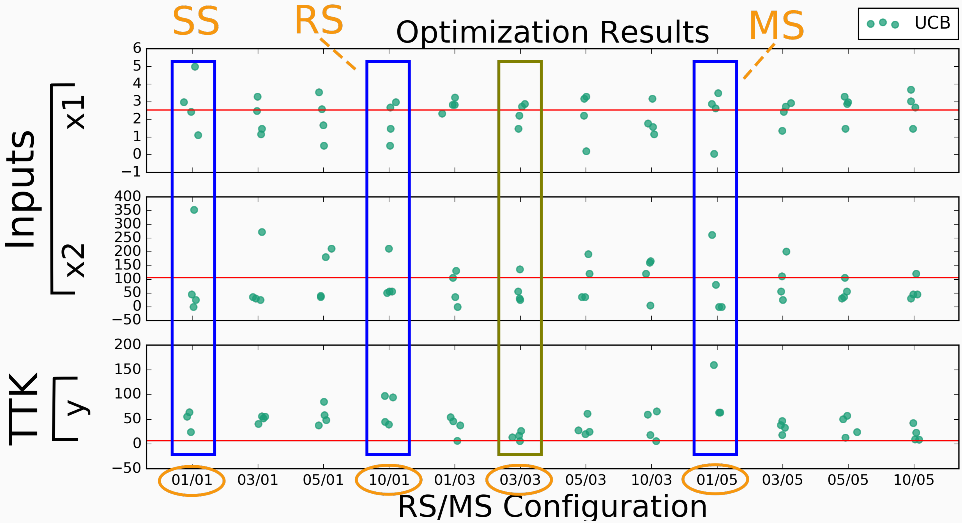
# Results

# Experiments

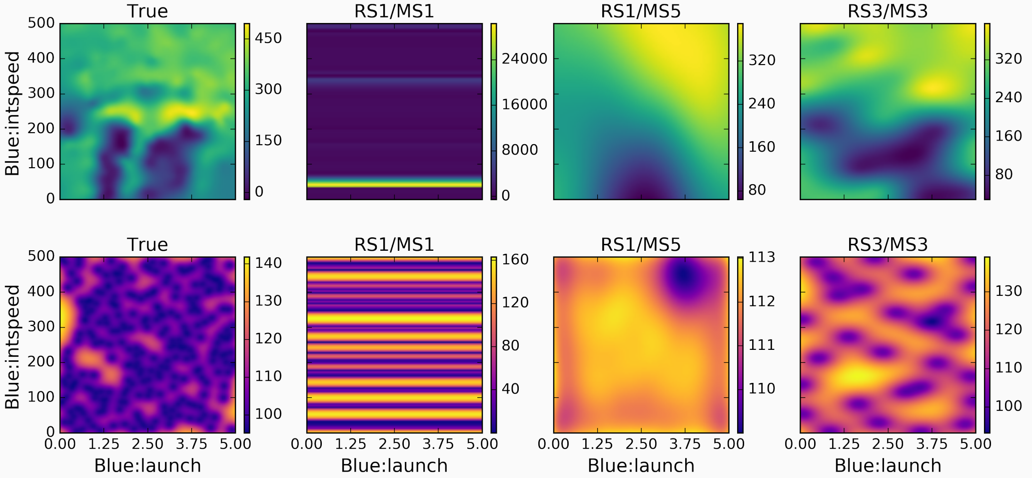
- Run experiments with different RS/MS configurations
- Evaluate parameter and objective estimates
- Compare surfaces of final surrogate models



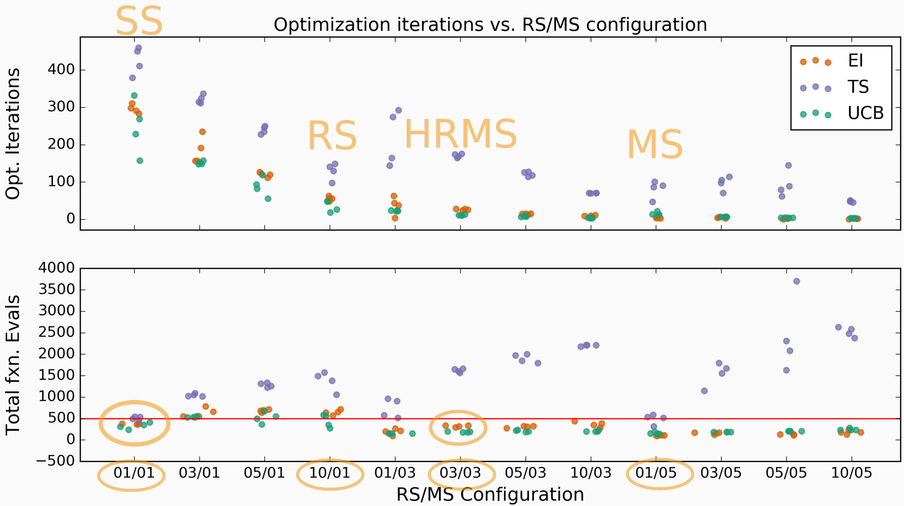
# More Repeatable Optimization



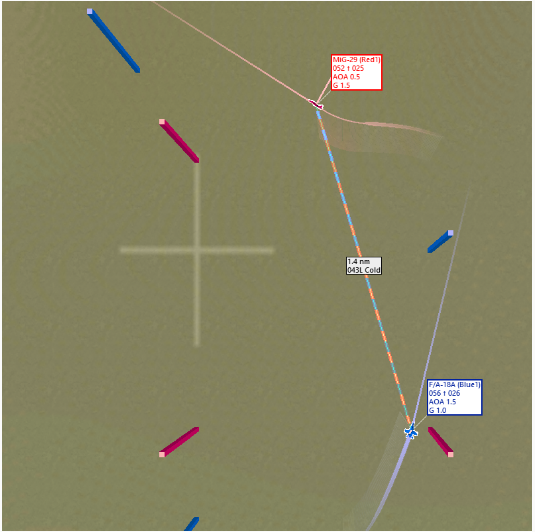
# Better Surrogate Surface



# Better Results, Less Experiments



# Video – TTK 23.4 Seconds



# Summary, Conclusions & Future Work

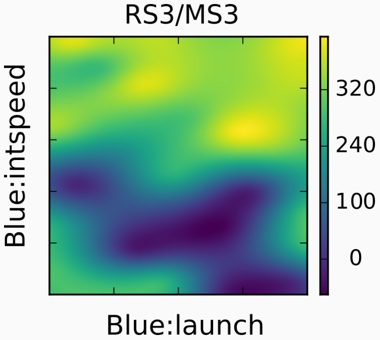
# Summary and Conclusions

Summary:

- Adaptive artificial agents are important for training humans
- Find optimum behaviors *and* map objective function with Bayesian Optimization
- Approach is generally applicable for all AI decision making

Conclusions:

- AI jet pilot behaviors can be optimized using GPBO
- HRMS makes a higher quality surrogate model
- Better model $\rightarrow$ better optimization
- Does not require more fxn evals
- Better model $\rightarrow$ better adaptation
- Model assists in interpreting decisions



# Future Work

Aerial Combat:

- Adapt multiple agents simultaneously
- Include humans in training
- Use AI as tutor/instructor
- Multi-objective learning

GPBO Theory:

- More detailed analysis of why RS helps
- Online adaptation of experimentation
- How to select RS and MS
- Fully Bayesian learning

# Questions?

# References

Nicholas Cole, Sushil J. Louis, and CHris Miles. “Using a genetic algorithm to tune first-person shooter bots”. In: *Proceedings of the 2004 Congress on Evolutionary Computation (IEEE Cat. No.04TH8753)* 1 (June 2004), pp. 139–145.

Emile Contal et al. “Parallel Gaussian process optimization with upper confidence bound and pure exploration”. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol. 8188 LNAI. PART 1. Springer, 2013, pp. 225–240. ISBN: 9783642409875. arXiv: arXiv:1304.5350v3.

# References II


- Carroll Croarkin and Paul Tobian. "e-Handbook of Statistical Methods". In: *NIST/SEMATECH, Available online: http://www. itl. nist. gov/div898/handbook* (2016).
- Margery J Doyle and Antoinette M Portrey. "Rapid Adaptive Realistic Behavior Modeling is Viable for Use in Training". In: *Proceedings of the 23rd Conference on Behavior Representation in Modeling and Simulation (BRIMS)*. 2014.
- N Ernest et al. "Genetic Fuzzy based Artificial Intelligence for Unmanned Combat Aerial Vehicle Control in Simulated Air Combat Missions". In: *J Def Manag* 6 (2016).
- Alexander Forrester, Andras Sobester, and Andy Keane. *Engineering design via surrogate modelling: a practical guide*. John Wiley & Sons, 2008.

# References III


P.G. Gonsalves and J.E. Burge. "Software Toolkit for Optimizing Mission Plans (STOMP)". en. In: *Collection of Technical Papers - AIAA 1st Intelligent Systems Technical Conference*. Vol. 1. American Institute of Aeronautics and Astronautics, Sept. 2004, pp. 391–399. ISBN: 156347719X.

Deng Huang. "Experimental planning and sequential kriging optimization using variable fidelity data". PhD thesis. 2005, 120–120 p. ISBN: 9780496944156; 0496944150.

Donald R. Jones, Computer Law, and Computer Law. "Lipschitzian Optimization Without the Lipschitz Constant". In: *Journal of Optimization Theory and Applications* 79.1 (1993), pp. 157–181.

Donald R. Jones, Matthias Schonlau, and J William. "Efficient Global Optimization of Expensive Black-Box Functions". In: *Journal of Global Optimization* 13.4 (1998), pp. 455–492. ISSN: 09255001.

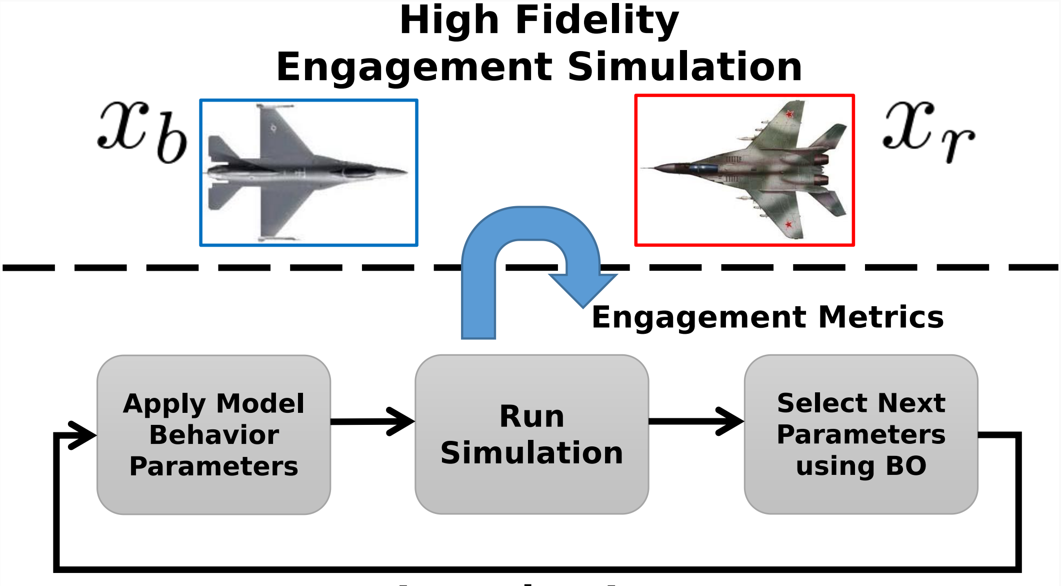
# References IV


- Chishyan Liaw et al. "Evolving a Team in a First-Person Shooter Game By Using a Genetic Algorithm". In: *Applied Artificial Intelligence* 27.3 (Mar. 2013), pp. 199–212. ISSN: 0883-9514.
- Sandeep Mulgund, Karen Harper, and Greg Zacharias. "Large-Scale Air Combat Tactics Optimization Using Genetic Algorithms". en. In: *Journal of Guidance, Control, and Dynamics* 24.1 (Jan. 2001), pp. 140–142. ISSN: 0731-5090.
- S. Mulgund et al. "Air combat tactics optimization using stochastic genetic algorithms". In: *SMC'98 Conference Proceedings. 1998 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No.98CH36218)*. Vol. 4. Oct. 1998, pp. 3136–3141. ISBN: 0-7803-4778-1.
- Nasri Othman et al. "Simulation-based optimization of StarCraft tactical AI through evolutionary computation". In: *2012 IEEE Conference on Computational Intelligence and Games, CIG 2012*. Sept. 2012, pp. 394–401. ISBN: 9781467311922.

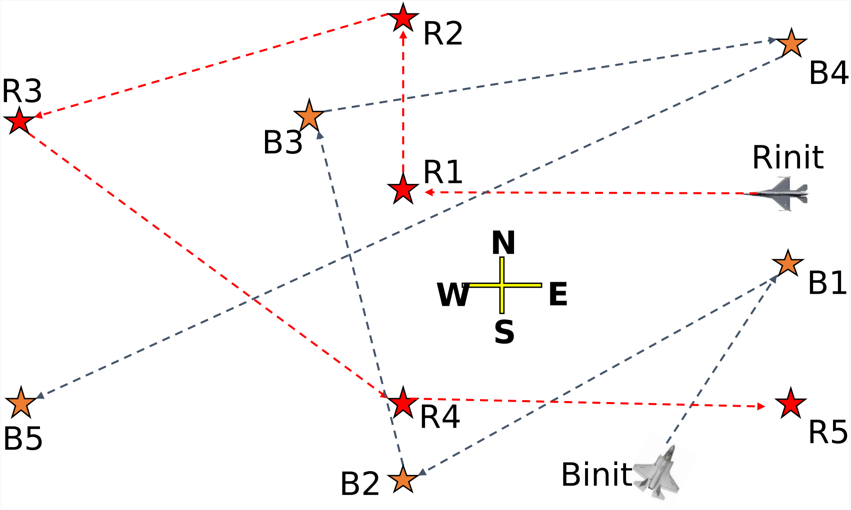
# References V


- Thomas Pyzdek and Paul A Keller. *Quality engineering handbook*. CRC Press, 2003.
- Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian processes for machine learning*. Adaptive computation and machine learning. Cambridge, Mass: MIT Press, 2006, p. 248. ISBN: 026218253X. arXiv: 026218253X.
- M Schonlau. "Computer experiments and global optimization". PhD thesis. University of Waterloo, 1997, p. 143.
- Niranjan Srinivas et al. "Gaussian Process Optimization in the Bandit Setting: No Regret and Experimental Design". In: (June 2010). arXiv: 0912.3995.
- Wen-Hai Wu and Qingdao Branch. "Air Combat Decision-making for Cooperative Multiple Target Attack using Heuristic Adaptive Genetic Algorithm". In: *Proceedings of the 4th International Conference on Machine Learning and Cybernetics*. Vol. 1. Aug. 2005, pp. 473–478. ISBN: 0780390911.

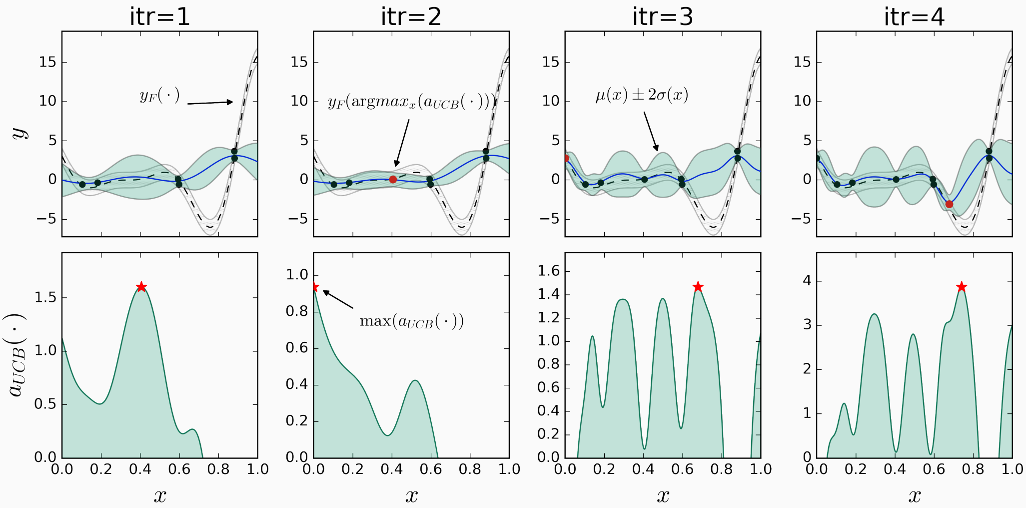
# Learning Loop



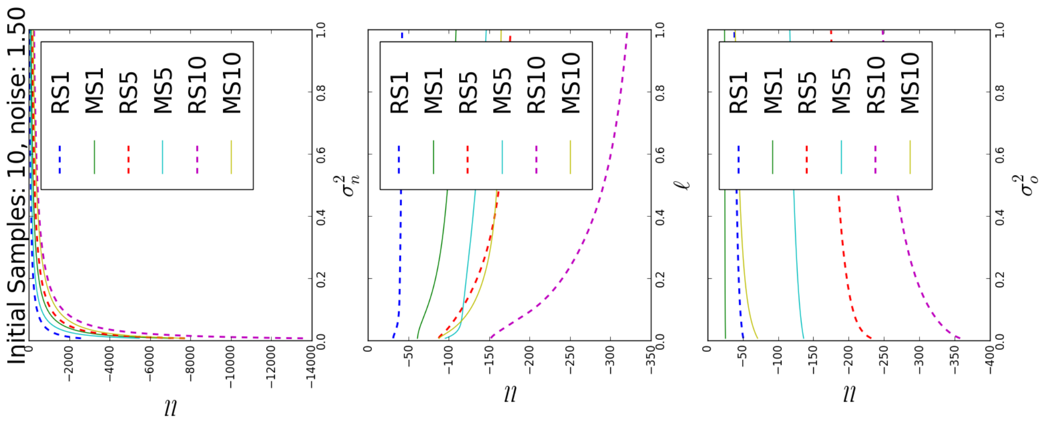
# Engagement Space



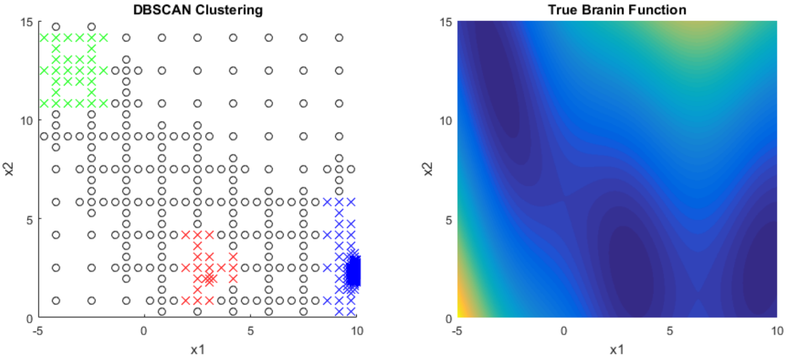
# A Toy Example: Forrester function [6]



# HRMS effect on log-likelihood



# DIRECT



DBSCAN Clustering
15
10
5
0
-5
0
5
10
x1
x2
True Branin Function
15
10
5
0
-5
0
5
10
x1
x2


(a)

(b)