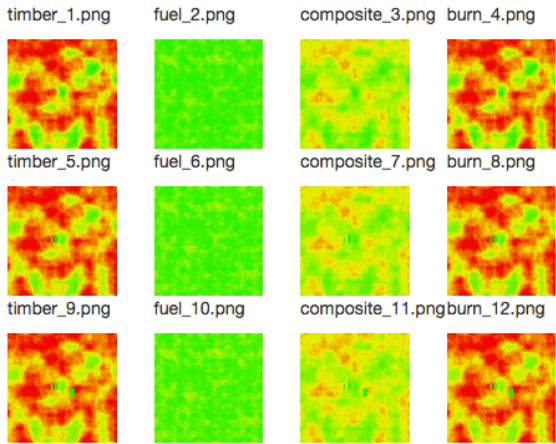


# Facilitating Testing and Debugging of Markov Decision Processes with Interactive Visualization

Sean McGregor, Hailey Buckingham,  
Thomas G. Dietterich, Rachel Houtman,  
Claire Montgomery, and Ronald Metoyer

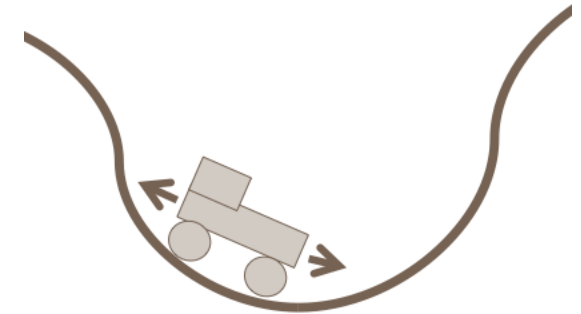
# What are Markov Decision Processes (MDPs)? Sequential Decision Making Under Uncertainty



Wildfire Suppression



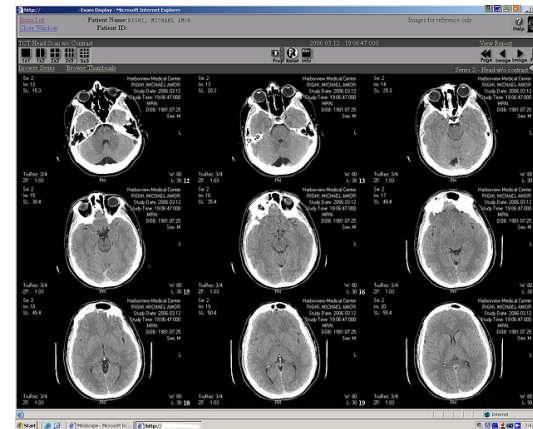
Autonomous Helicopter<sup>0</sup>



Mountain Car



Logistics<sup>1</sup>



Medical Diagnosis<sup>2</sup>

## Outline

1. **Markov Decision Processes (MDPs)**
  - | Basic Introduction
  - | Testing
2. **MDPvis**
  - | Design
  - | Testing Examples
  - | MDPvis Use Case Study
3. **Concluding**

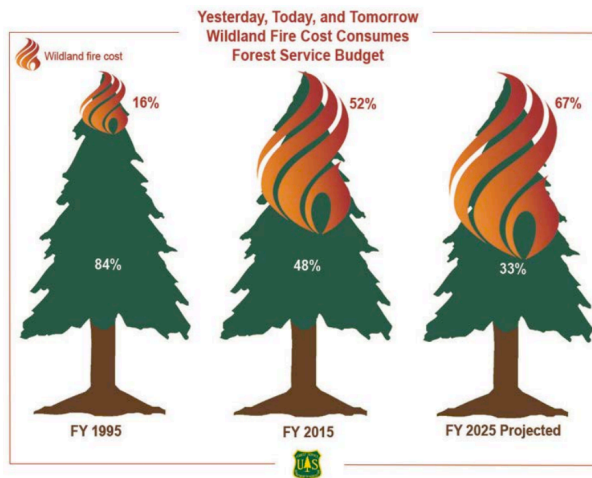
Notation,  $M = \langle S, A, P, R, \gamma, P_0 \rangle$

$S$	All States of the World
$P_0$	Starting State Distribution
$A$	Available Actions
$R(s, a)$	Rewards
$\gamma \in (0, 1)$	Discount
$P$	State Transition Probabilities (Simulators)
$\pi(s) \rightarrow a$	Policy

Puterman, M. (1994). Markov Decision Processes: Discrete Stochastic Dynamic Programming (1st ed.). Wiley-Interscience.

## Motivating Domain of Wildfire

Starting in 1935, the United States adopted the “**10 AM policy**”



We need a more nuanced approach.



Houtman, R. M., Montgomery, C. A., Gagnon, A. R., Calkin, D. E., Dietterich, T. G., McGregor, S., & Crowley, M. (2013). Allowing a Wildfire to Burn: Estimating the Effect on Future Fire Suppression Costs. *International Journal of Wildland Fire*, 22(7), 871–882.

<http://www.fs.fed.us/sites/default/files/2015-Fire-Budget-Report.pdf>

## Modeling Wildfire

$S$	All the possible configurations of trees / ignitions
$P_0$	A snapshot of the current forest, with a random fire
$A$	Suppress or let-burn
$R(s, a)$	Timber harvest, Suppression Expense
$\gamma \in (0, 1)$	0.96 (Forest Service Standard)
$P$	Several Simulators
$\pi(s) \rightarrow a$	Suppress all fires

Represents a challenging and more general class of MDPs

- High Dimensional States
- Large State Space
- Integrates Several Simulators

Simulators

Optimizer

Rewards

Policy

 $P_0$  ●

Simulators

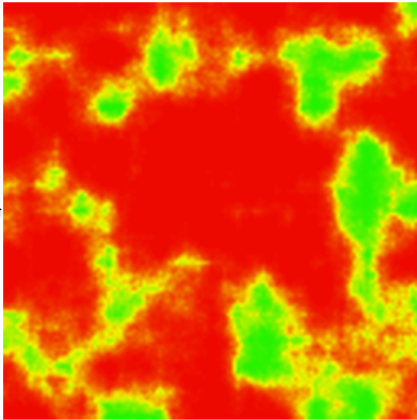
Optimizer

Rewards

Policy

Start with Today's Landscape

$P_0$  →





Simulators

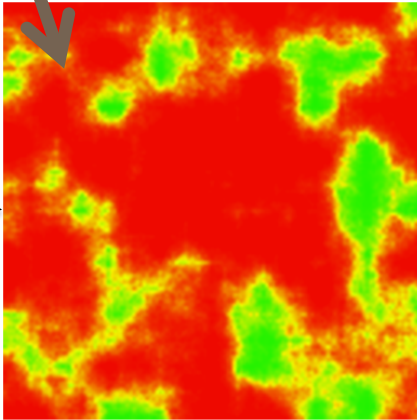
Optimizer

Rewards

Policy

Generate an ignition and weather

$P_0$  →



Simulators

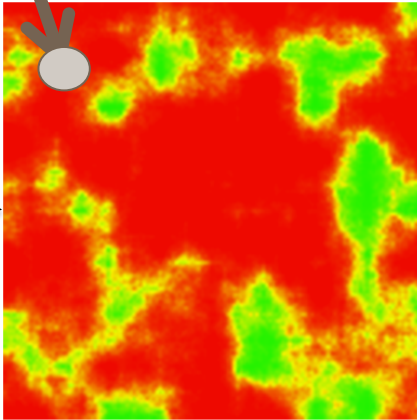
Optimizer

Rewards

Policy

Generate an ignition and weather

$P_0$  →



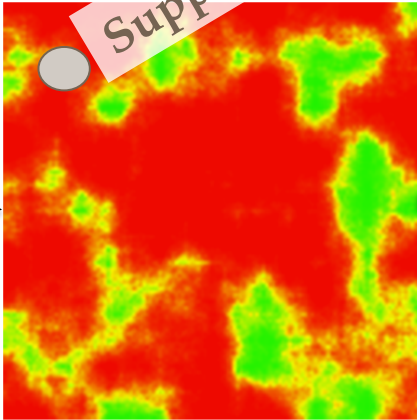
Simulators

Optimizer

Rewards

Policy

Select an Action

 $P_0$ 

Simulators

Optimizer

Rewards

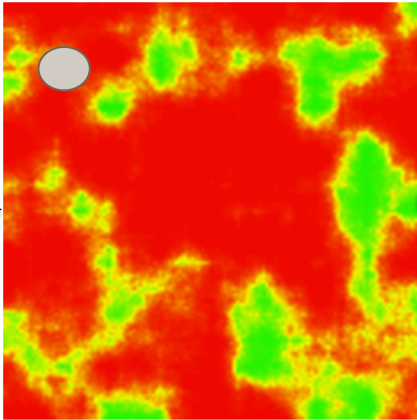
Policy

Fire Suppression Effort

Fire Suppression Costs

\$(95,000)

$P_0$



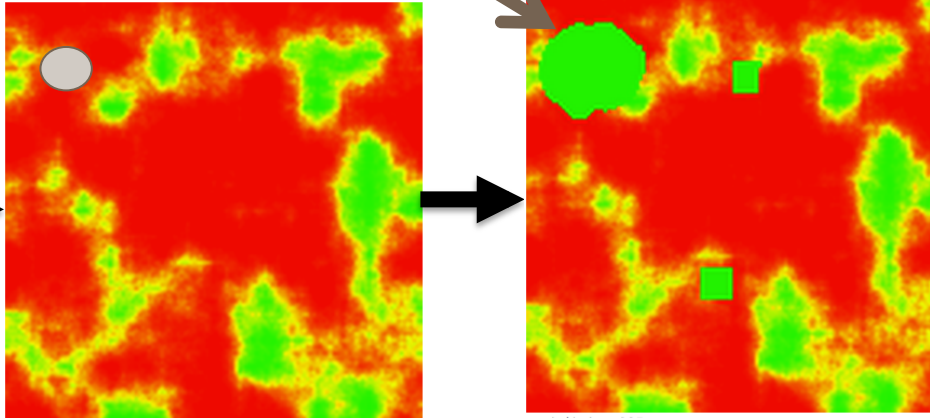
Simulators

Optimizer

Rewards

Policy

Update Vegetation for Wildfire

 $P_0$ 

Simulators

Optimizer

Rewards

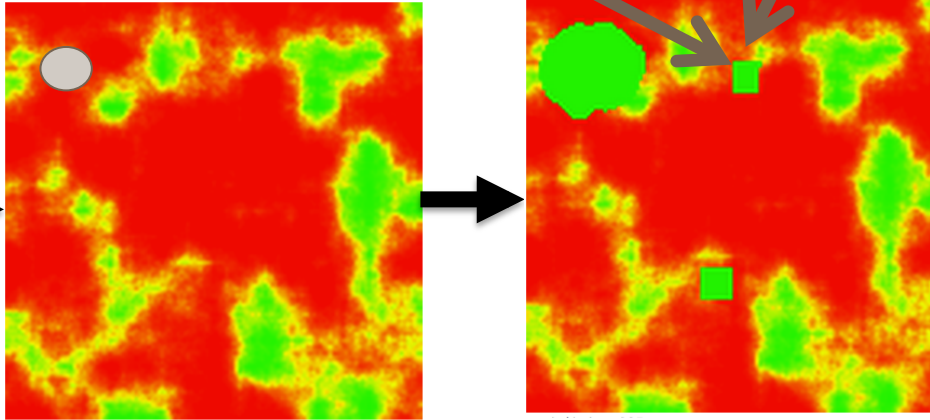
Policy

Update Vegetation for harvest

\$20,000

Harvest Revenue

$P_0$



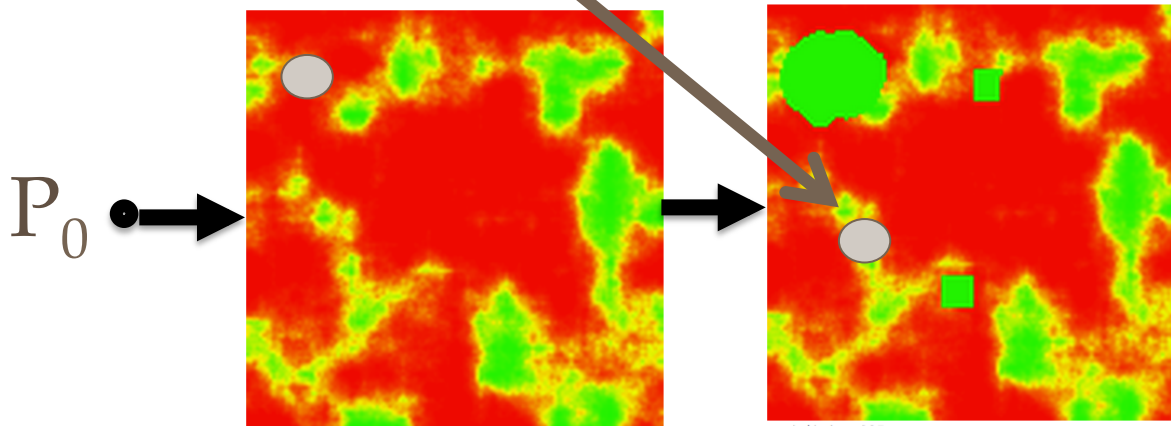
Simulators

Optimizer

Rewards

Policy

Generate an ignition and weather



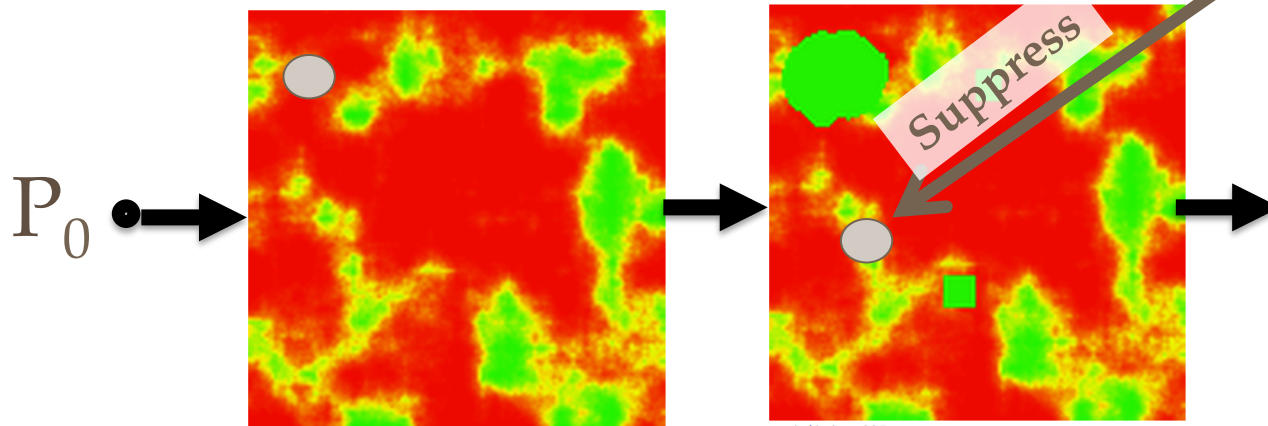
Simulators

Optimizer

Rewards

Policy

Select an Action





Simulators

Optimizer

Rewards

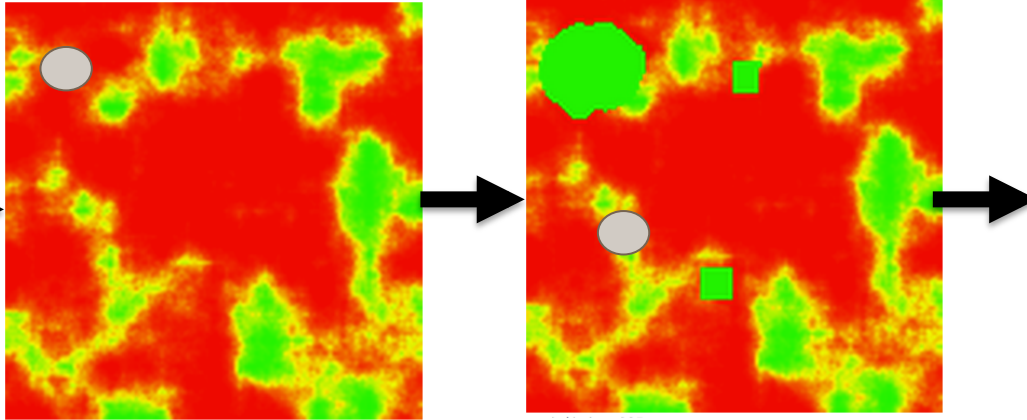
Policy

Fire Suppression Effort

Fire Suppression Costs

\$(15,000)

$P_0$



Simulators

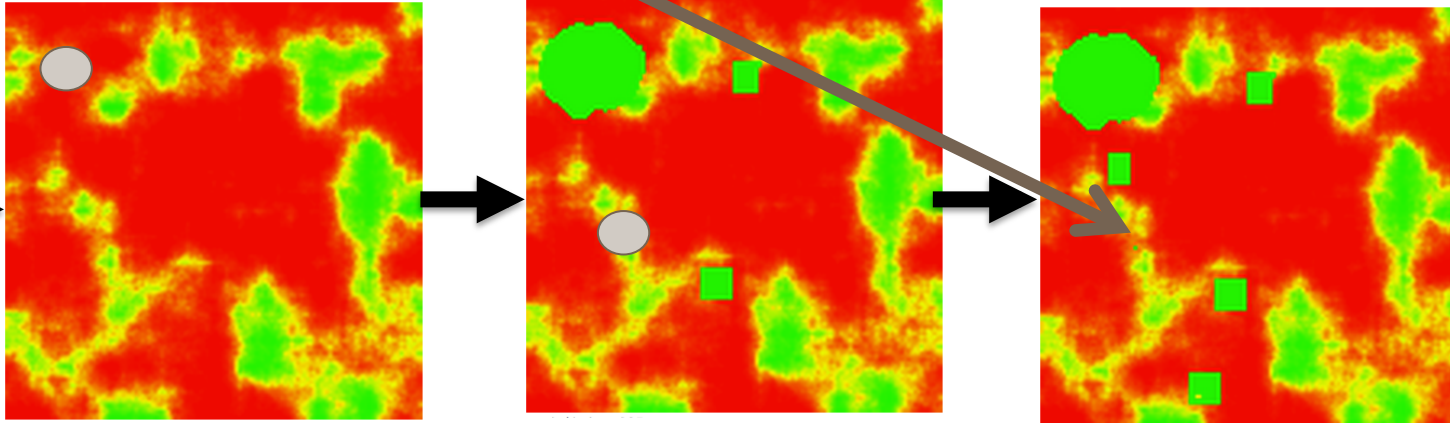
Optimizer

Rewards

Policy

Update Vegetation for Wildfire

$P_0$



Simulators

Optimizer

Rewards

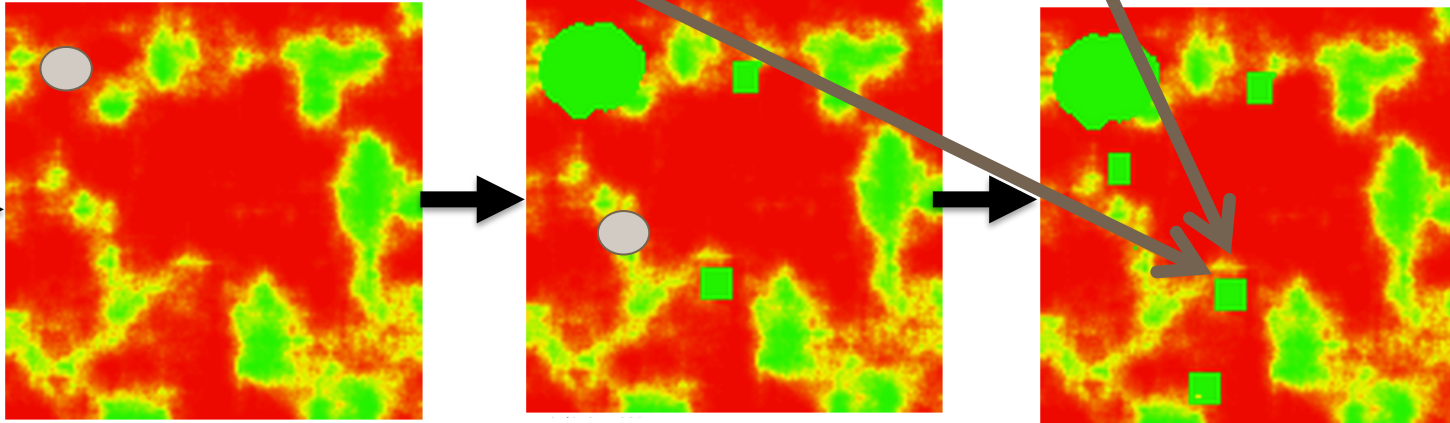
Policy

Update Vegetation for Harvest

Harvest Revenue

\$20,000

$P_0$



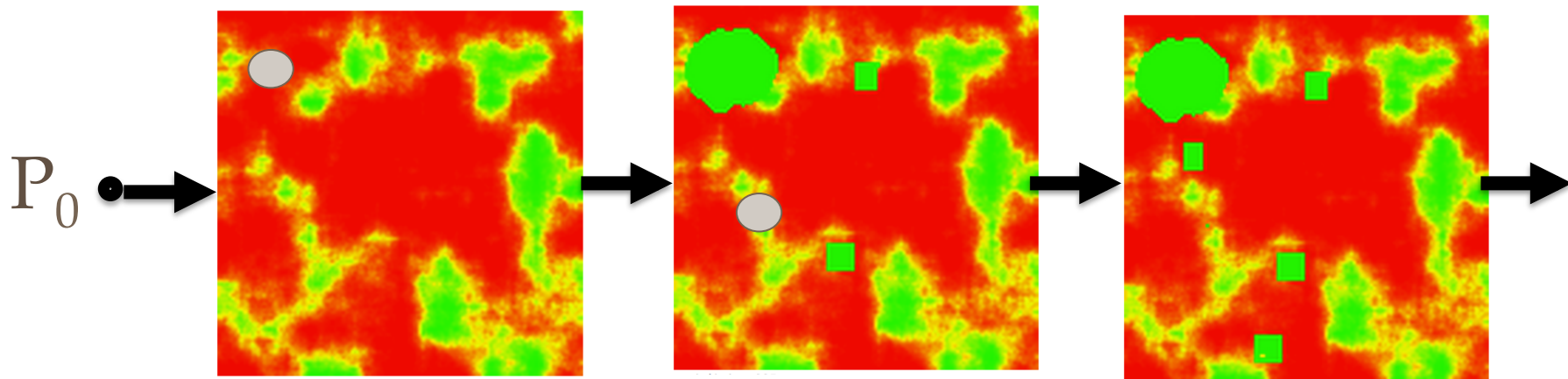
Simulators

Optimizer

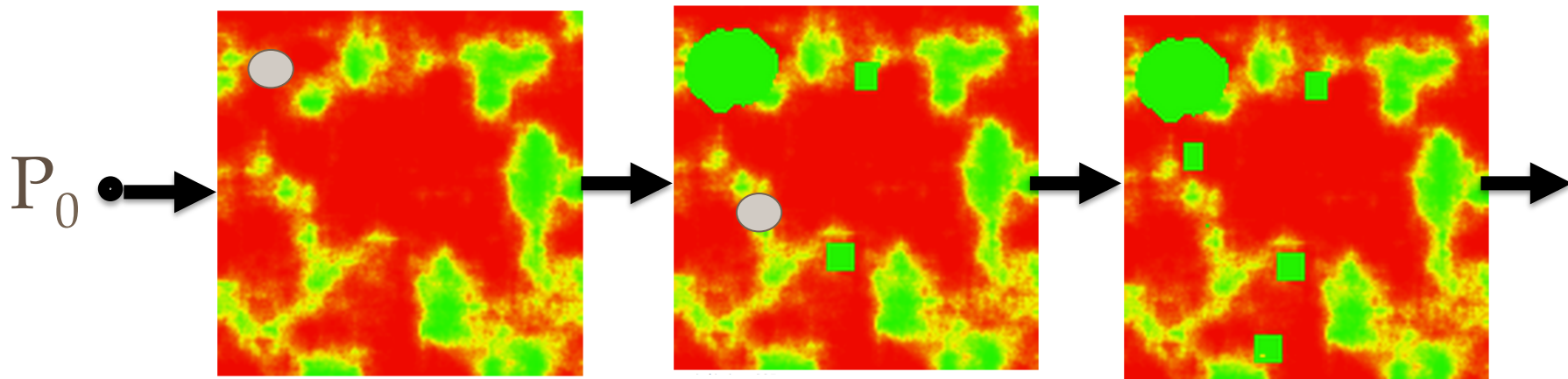
Rewards

Policy

(Continue Until Reaching the Horizon)



## A High Dimensional Probabilistic Time Series



...And this is just one of many!

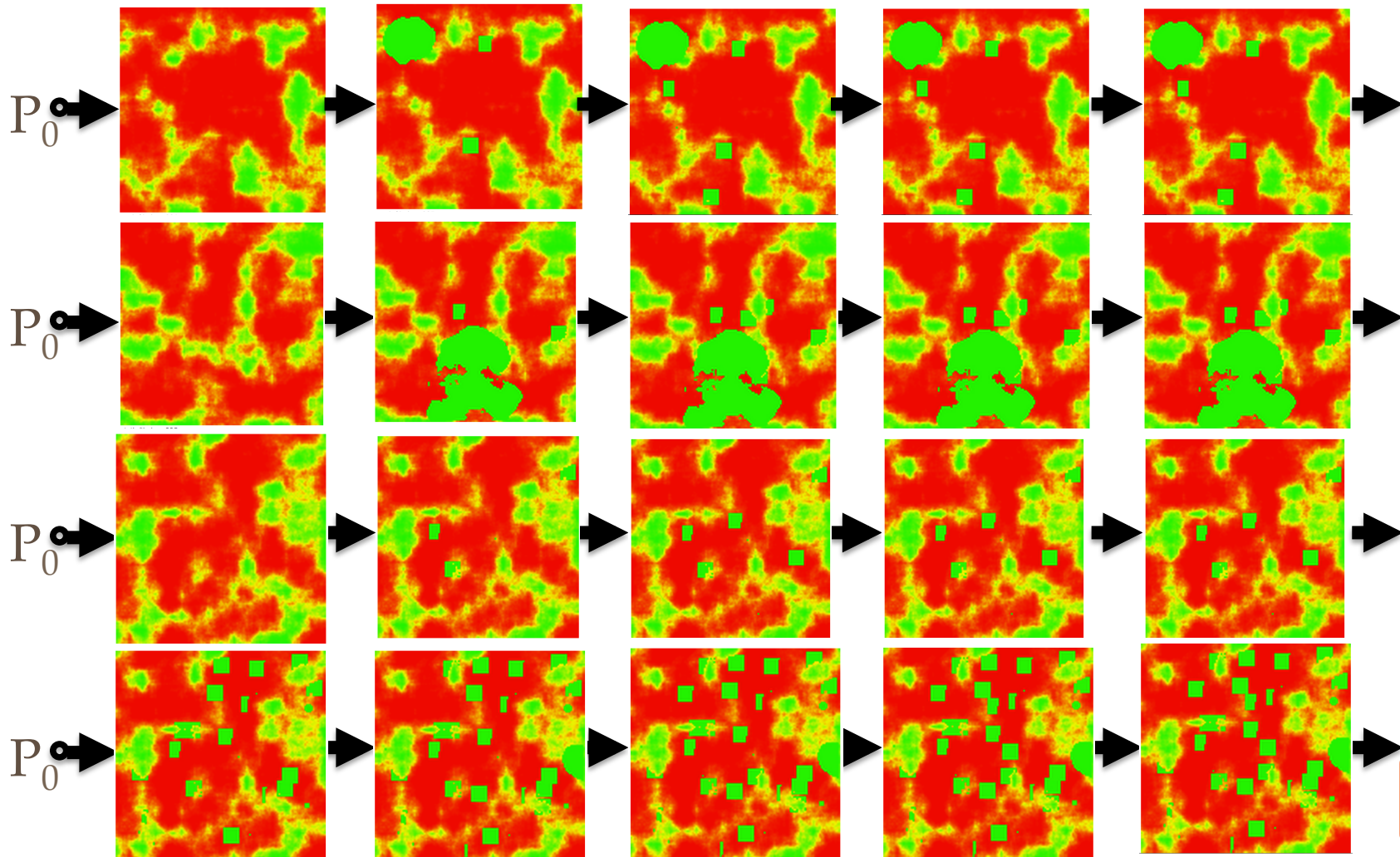
Simulators

Optimizer

Rewards

Policy

## Monte Carlo Rollouts



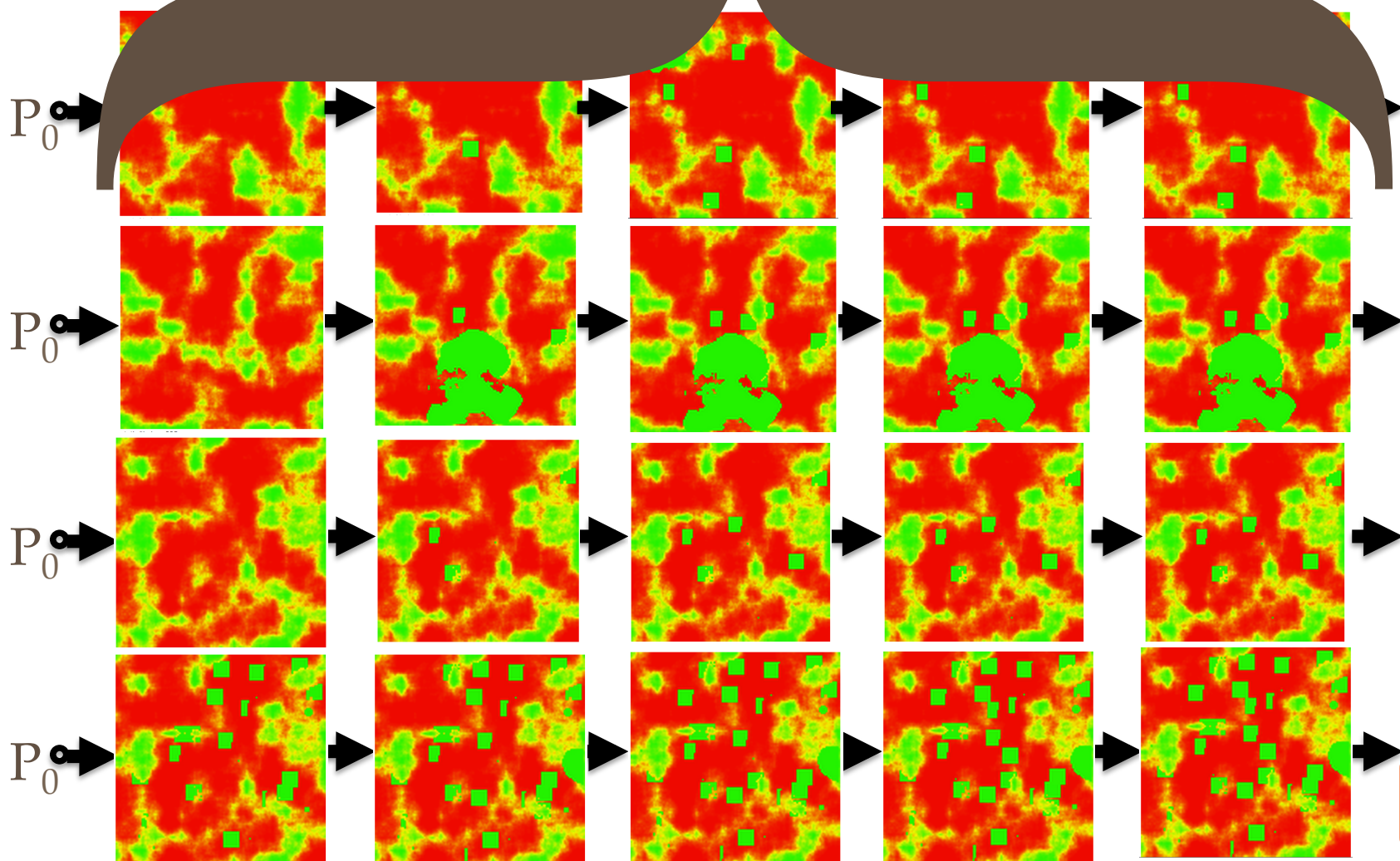
Simulators

Optimizer

Rewards

Policy

All visited states influence optimizer



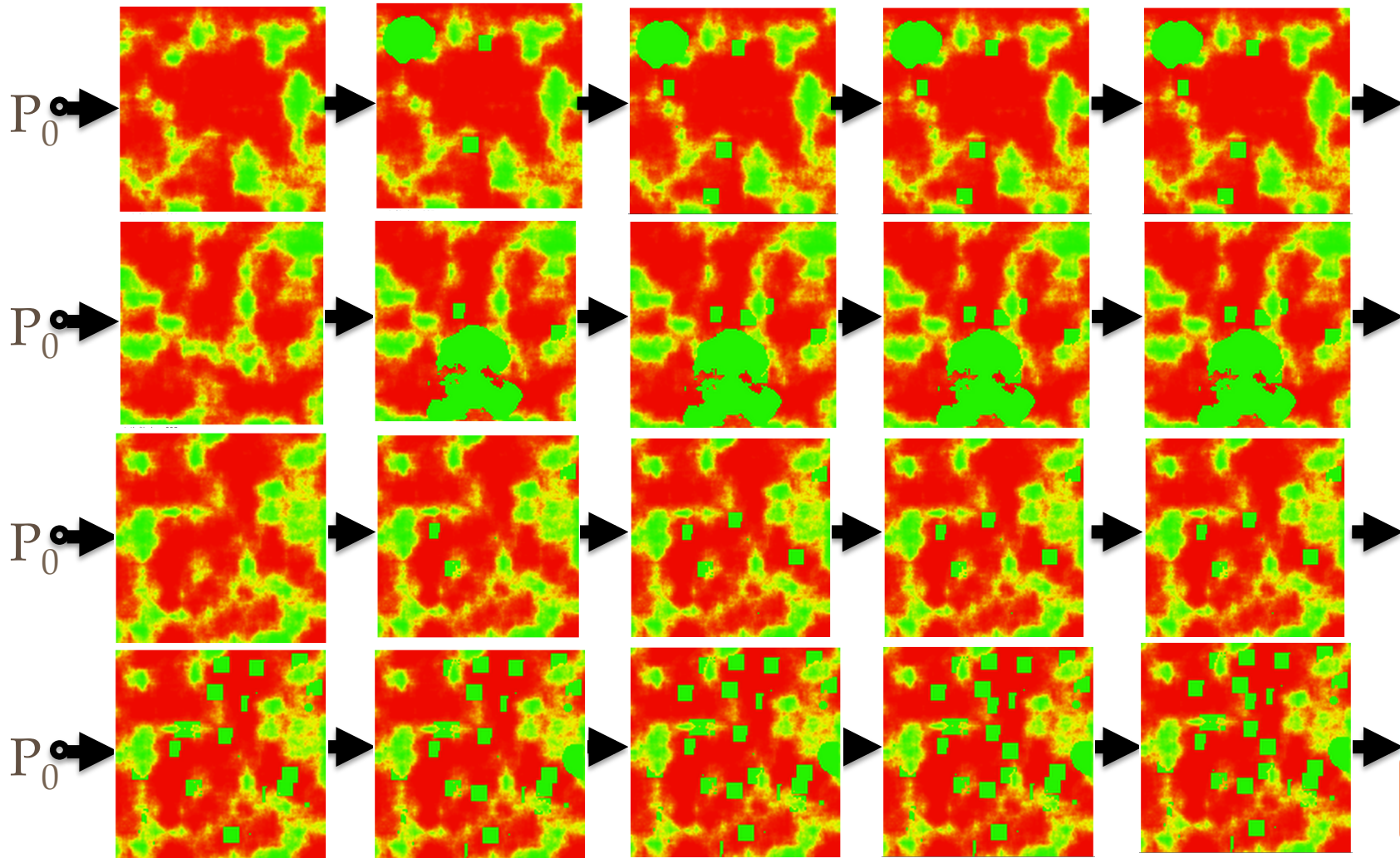
Simulators

Optimizer

Rewards

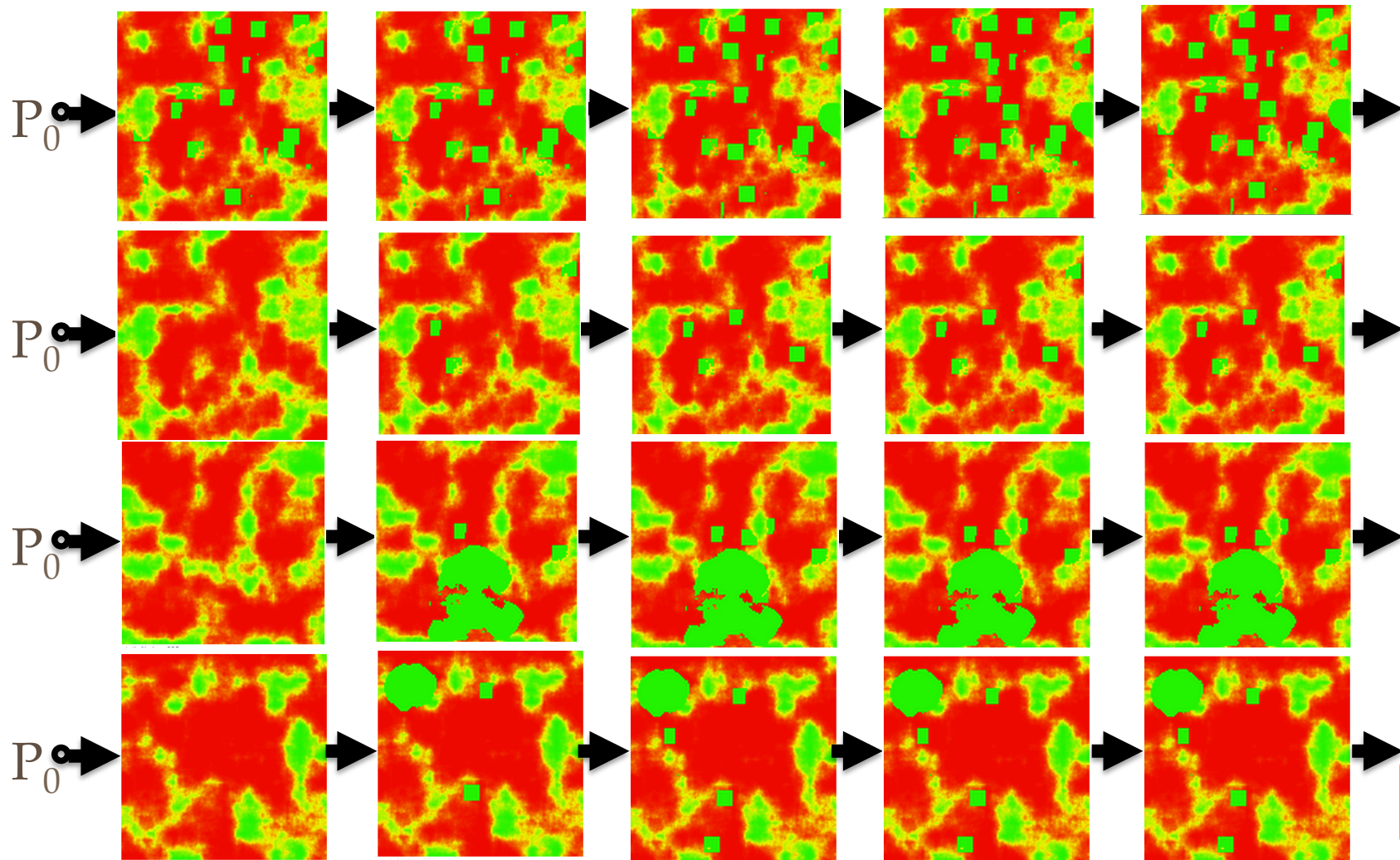
Policy

Update Policy





# The Rollout Distribution Changes!



## MDP Testing Challenges

- Bugs are **probabilistically** expressed in a **high dimensional temporal dataset**.
- The **dataset changes** with changes to parameters.
- The **optimizer sees more of the state and policy space** than the user.

Testing requires  
exploring rollouts and  
parameters

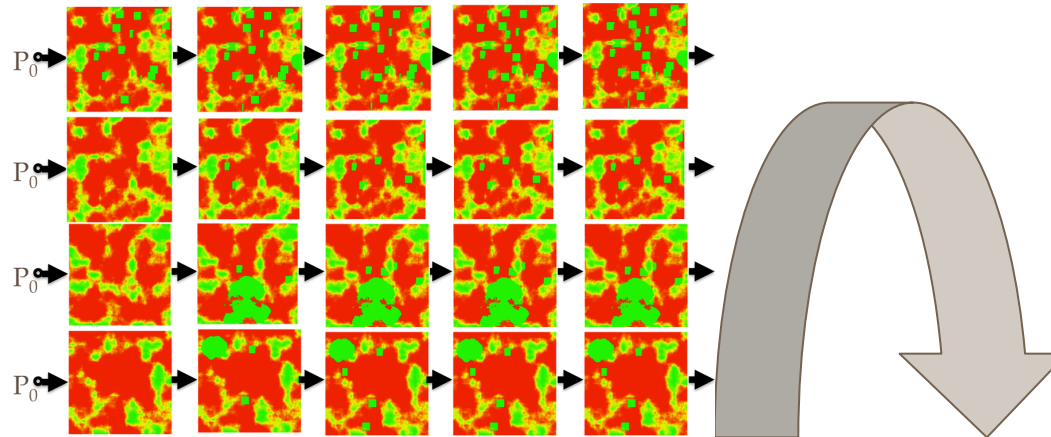
# MDP Debugging and Fault Isolation

- Deactivate/modify components to isolate fault
  - e.g. Balance reward magnitude and frequency

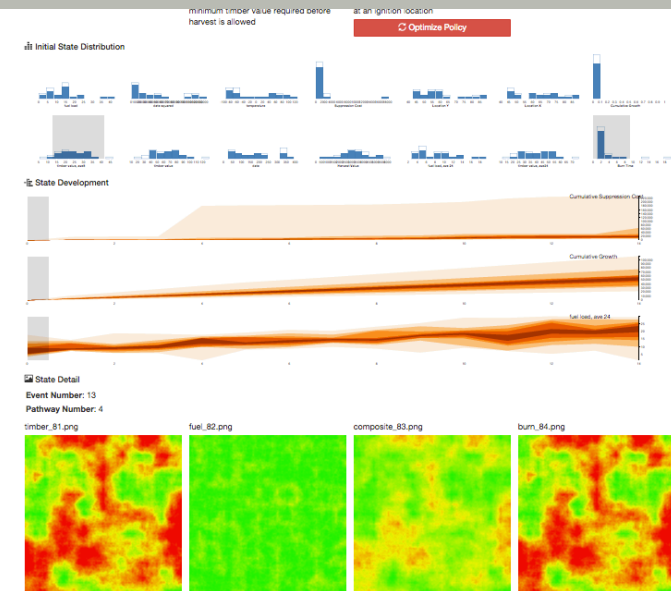
**Debug MDP  
specification and  
integration with  
parameter changes**

# Testing and Debugging Process

## 1. Generate Rollouts



## 2. Visualize the data



## 3. Change Parameters

**\$ Reward Specification**

Discount: -1

Suppression Fixed Cost: \$ 500

Suppression Variable Cost: \$ 500

**⚙ Model Modifiers**

Years to simulate: 20

Futures to simulate: 40

Landscape Size: 129

Harvest Percent: 0.95

Slash Remaining: 10

Fuel Accumulation: 2

Suppression Effect: 0.5

Use Original Bugs: 0

**⚙ Policy Definition**

Constant: 0

Date: 0

Days Left: 0

Temperature: 0

Wind Speed: 0

Fuel Load 8: 0

Fuel Load 24: 0

[Optimize a New Policy](#)

**📄 Exploration History**

Rollouts are Current

Expected Value \$ -640252.42  
[View Rollout Set 3](#) [Compare To](#)

Expected Value \$ -47446.26  
[View Rollout Set 2](#) [Compare To](#)

Expected Value \$ -42404.22  
[View Rollout Set 1](#) [Compare To](#)

## Outline

1. Markov Decision Processes (MDPs)
  - | Basic Introduction
  - | Testing
2. **MDPvis**
  - | Design
  - | Testing Examples
  - | MDPvis Use Case Study
3. Concluding

# Introducing MDPvis

### \$ Reward Specification

- 1 Discount

\$ 500 Suppression Fixed Cost

\$ 500 Suppression Variable Cost

### ⚙ Model Modifiers

Y 20 Years to simulate

# 40 Futures to simulate

# 129 Landscape Size

% 0.95 Harvest Percent

# 10 Slash Remaning

# 2 Fuel Accumulation

% 0.5 Suppression Effect

- 0 Use Original Bugs

### 🔗 Policy Definition

0 Constant

0 Date

0 Days Left

0 Temperature

0 Wind Speed

0 Fuel Load B

0 Fuel Load 24

↻ Optimize a New Policy

### 📋 Exploration History

✔ Rollouts are Current

Expected Value \$ -640252.42  
[View Rollout Set 3](#) Compare To

Expected Value \$ -47446.26  
[View Rollout Set 2](#) Compare To

Expected Value \$ -42404.22  
[View Rollout Set 1](#) Compare To

#### 📊 Initial State Distribution

#### 📈 State Development

#### 🖼 State Detail

Event Number: 13  
Pathway Number: 4

timber\_81.png

fuel\_82.png

composite\_83.png

burn\_84.png

# What are the elements of the MDPvis design?

**\$ Reward Specification**

Discount: -1

Suppression Fixed Cost: \$ 500

Suppression Variable Cost: \$ 500

**⚙ Model Modifiers**

Years to simulate: 20

Futures to simulate: 40

Harvest Percent: 0.95

Slash Remaning: 10

Fuel Accumulation: 2

Suppression Effect: 0.5

Use Original Bugs: 0

**⚡ Policy Definition**

Constant: 0

Date: 0

Days Left: 0

Temperature: 0

Wind Speed: 0

Fuel Load 8: 0

Fuel Load 24: 0

[Optimize a New Policy](#)

**☰ Exploration History**

Rollouts are Current

Expected Value \$ -47446.26

Expected Value \$ -42404.22

## Parameters

## Distributions at Time Step

## Distributions Through Time

**📄 State Detail**

Event Number: 13

Pathway Number: 4

timber\_81.png

fuel\_82.png

composite\_83.png

burn\_84.png

## State Snapshots

# Parameter Areas

## \$ Reward Specification

- Discount ⓘ
- \$  Suppression Fixed Cost ⓘ
- \$  Suppression Variable Cost ⓘ

## ⚙ Model Modifiers

- Y  Years to simulate ⓘ
- #  Futures to simulate ⓘ
- #  Landscape Size ⓘ
- %  Harvest Percent ⓘ
- #  Slash Remaning ⓘ
- #  Fuel Accumulation ⓘ
- %  Suppression Effect ⓘ
- ~  Use Original Bugs ⓘ

## ⚡ Policy Definition

- Constant ⓘ
- Date ⓘ
- Days Left ⓘ
- Temperature ⓘ
- Wind Speed ⓘ
- Fuel Load 8 ⓘ
- Fuel Load 24 ⓘ

[↻ Optimize a New Policy](#)

## ☰ Exploration History

✓ Rollouts are Current

Expected Value \$ -640252.42  
[View Rollout Set 3](#) [Compare To](#)

Expected Value \$ -47446.26  
[View Rollout Set 2](#) [Compare To](#)

Expected Value \$ -42404.22  
[View Rollout Set 1](#) [Compare To](#)



# History Area

## \$ Reward Specification

- 1 Discount ⓘ
- \$ 500 Suppression Fixed Cost ⓘ
- \$ 500 Suppression Variable Cost ⓘ

## ⚙ Model Modifiers

- Y 20 Years to simulate ⓘ
- # 40 Futures to simulate ⓘ
- # 129 Landscape Size ⓘ
- % 0.95 Harvest Percent ⓘ
- # 10 Slash Remaining ⓘ
- # 2 Fuel Accumulation ⓘ
- % 0.5 Suppression Effect ⓘ
- 0 Use Original Bugs ⓘ

## ↔ Policy Definition

- 0 Constant ⓘ
- 0 Date ⓘ
- 0 Days Left ⓘ
- 0 Temperature ⓘ
- 0 Wind Speed ⓘ
- 0 Fuel Load 8 ⓘ
- 0 Fuel Load 24 ⓘ

↻ Optimize a New Policy

## ☰ Exploration History

✓ Rollouts are Current

**Expected Value \$ -640252.42**

View Rollout Set 3 Compare To

**Expected Value \$ -47446.26**

View Rollout Set 2 Compare To

**Expected Value \$ -42404.22**

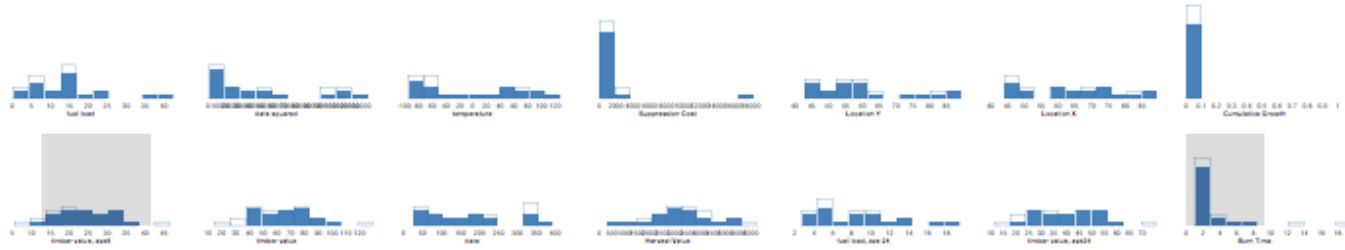
View Rollout Set 1 Compare To

Discount: The per-year discount  
 500 Suppression Fixed Cost: cost per day of suppression  
 500 Suppression Variable Cost: cost per hectare of suppression  
 15 Years to simulate: how far to look into the future  
 -25 Futures to simulate: how many stochastic futures to generate  
 0.9 Harvest Percent: timber harvest as a percent of annual increment  
 50 Minimum Timber Value: the minimum timber value required before harvest is allowed  
 0 Days Left: for each day left in the year  
 0 Temperature: for air temperature at the time of an ignition  
 0 Wind Speed: for wind speed at the time of an ignition  
 0 Timber Value: for the timber value at an ignition location

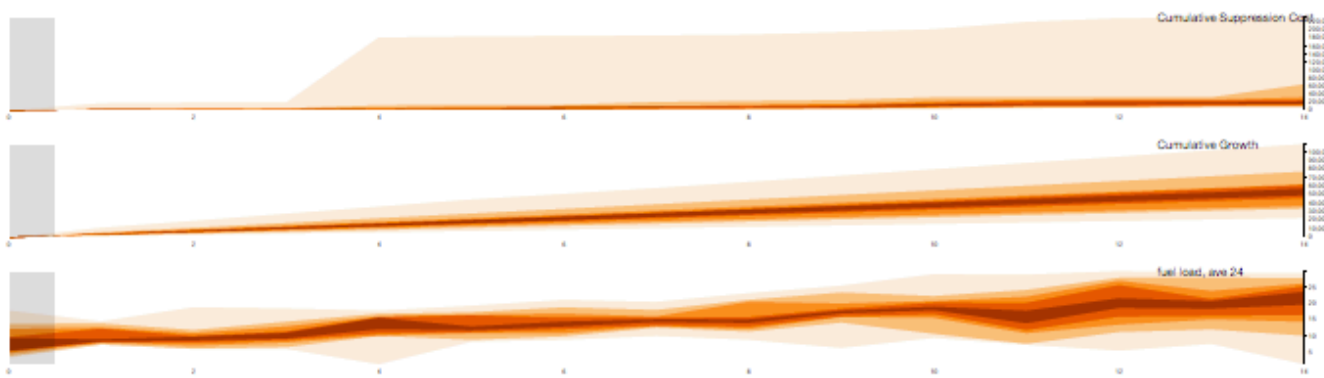
Rollouts are Current  
 Expected Value \$ 18354.79  
[View Rollout Set 2](#) [Compare To](#)  
 Expected Value \$ -10385.19  
[View Rollout Set 1](#) [Compare To](#)

# Visualization Areas

### Initial State Distribution



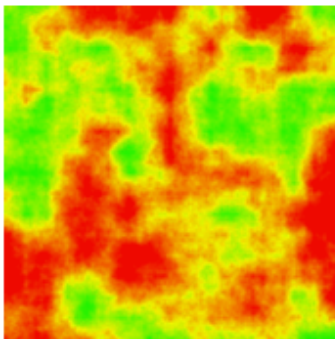
### State Development



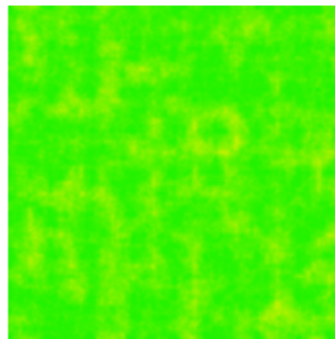
### State Detail

Event Number: 13  
 Pathway Number: 4

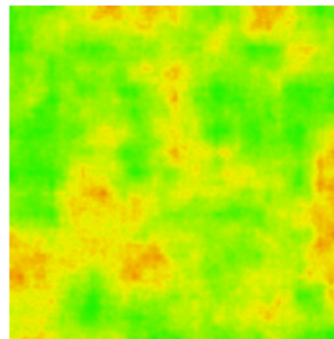
timber\_81.png



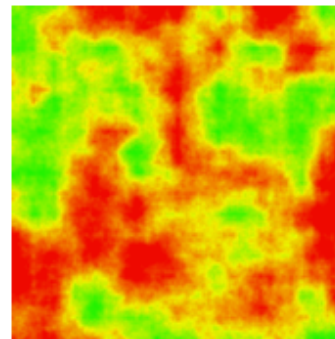
fuel\_82.png



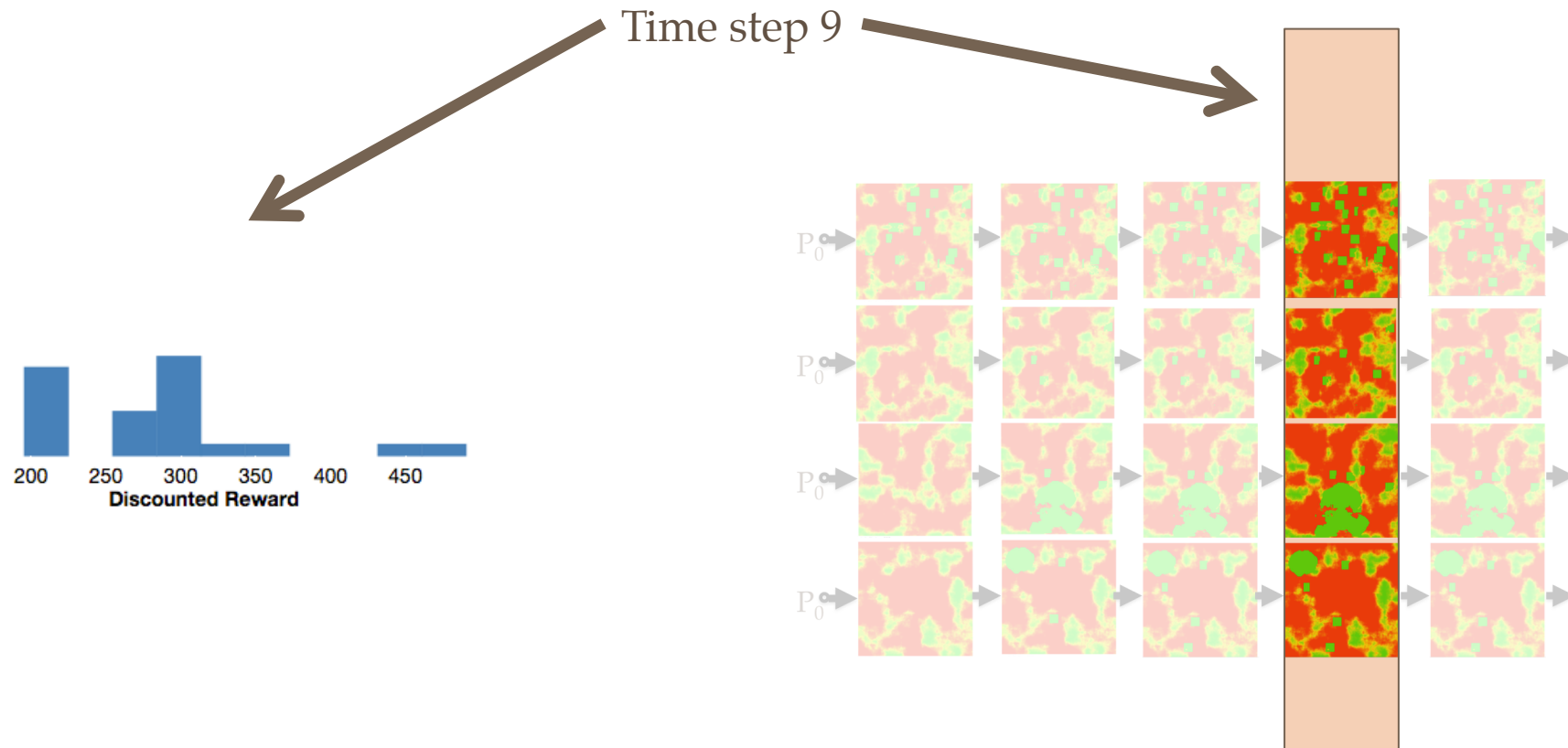
composite\_83.png



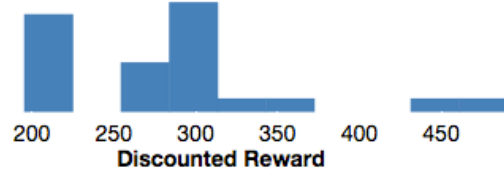
burn\_84.png



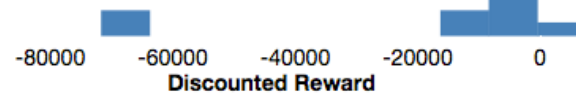
# State Variable Distributions for a Fixed Time Step



# State Variable Distributions for a Fixed Time Step



$\pi_1$ : Let-Burn



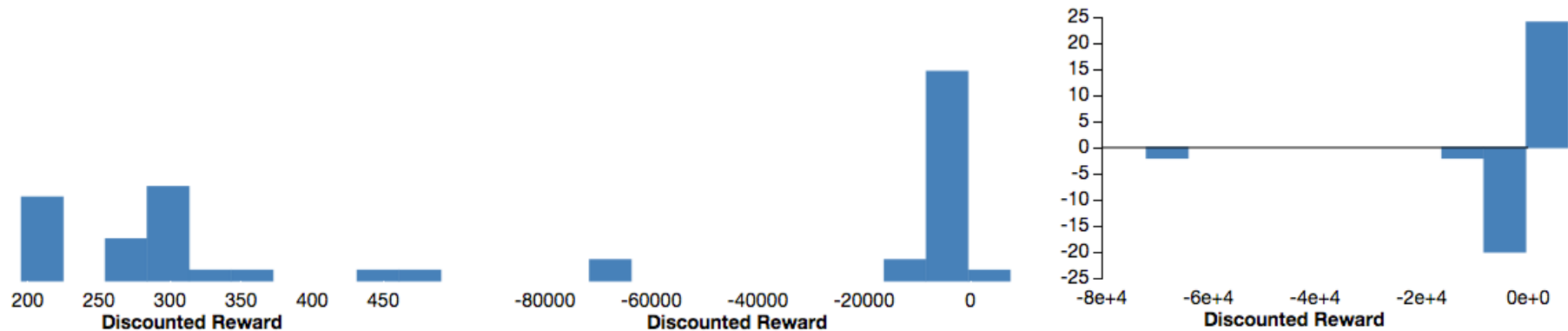
$\pi_2$ : Suppress-All



Comparison

$\pi_1 - \pi_2$

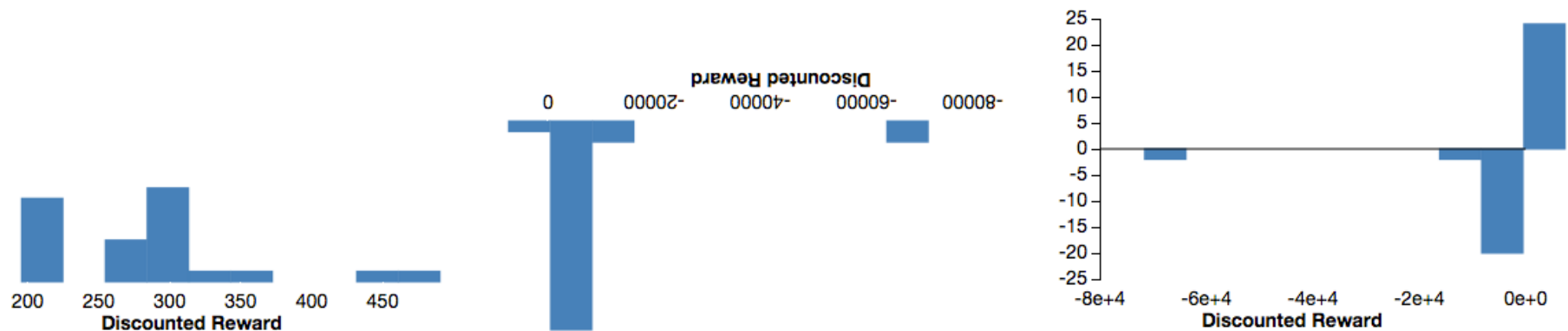
# State Variable Distributions for a Fixed Time Step



Comparison

$$\pi_1 - \pi_2$$

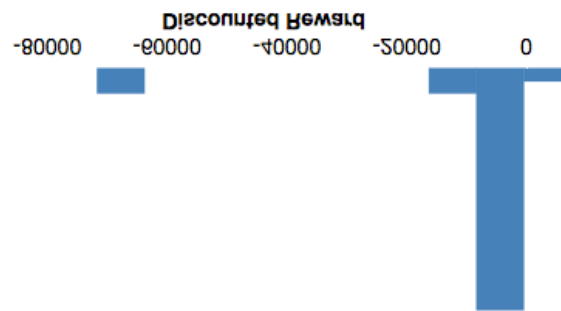
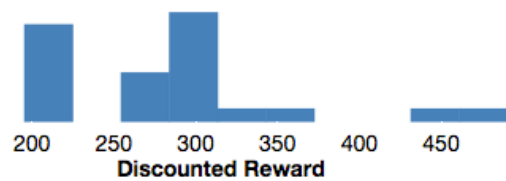
# State Variable Distributions for a Fixed Time Step



## Comparison

$$\pi_1 - \pi_2$$

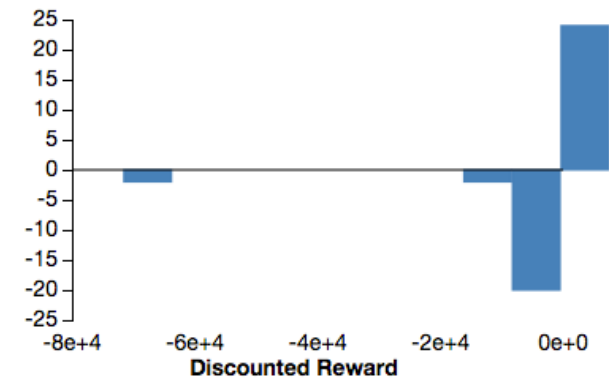
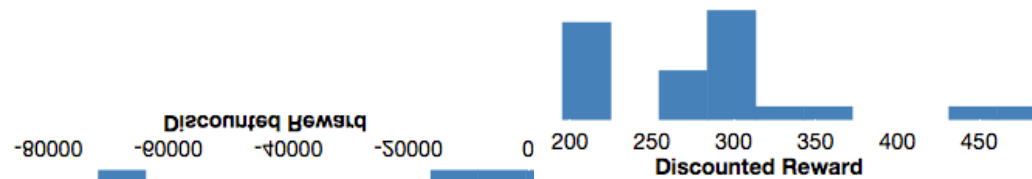
# State Variable Distributions for a Fixed Time Step



Comparison

$$\pi_1 - \pi_2$$

# State Variable Distributions for a Fixed Time Step



## Comparison

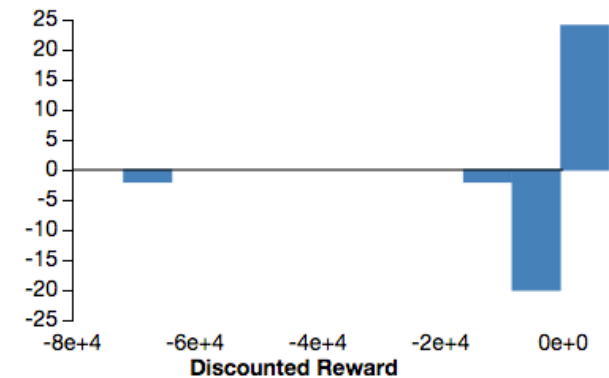
$$\pi_1 - \pi_2$$



## State Variable Distributions for a Fixed Time Step



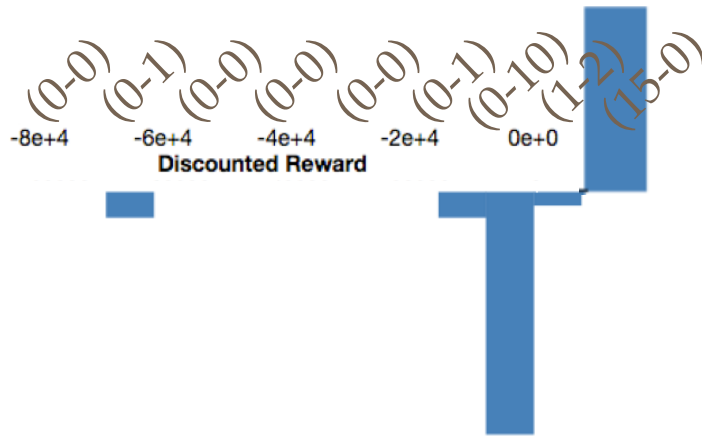
Rescale



Comparison

$$\pi_1 - \pi_2$$

# State Variable Distributions for a Fixed Time Step

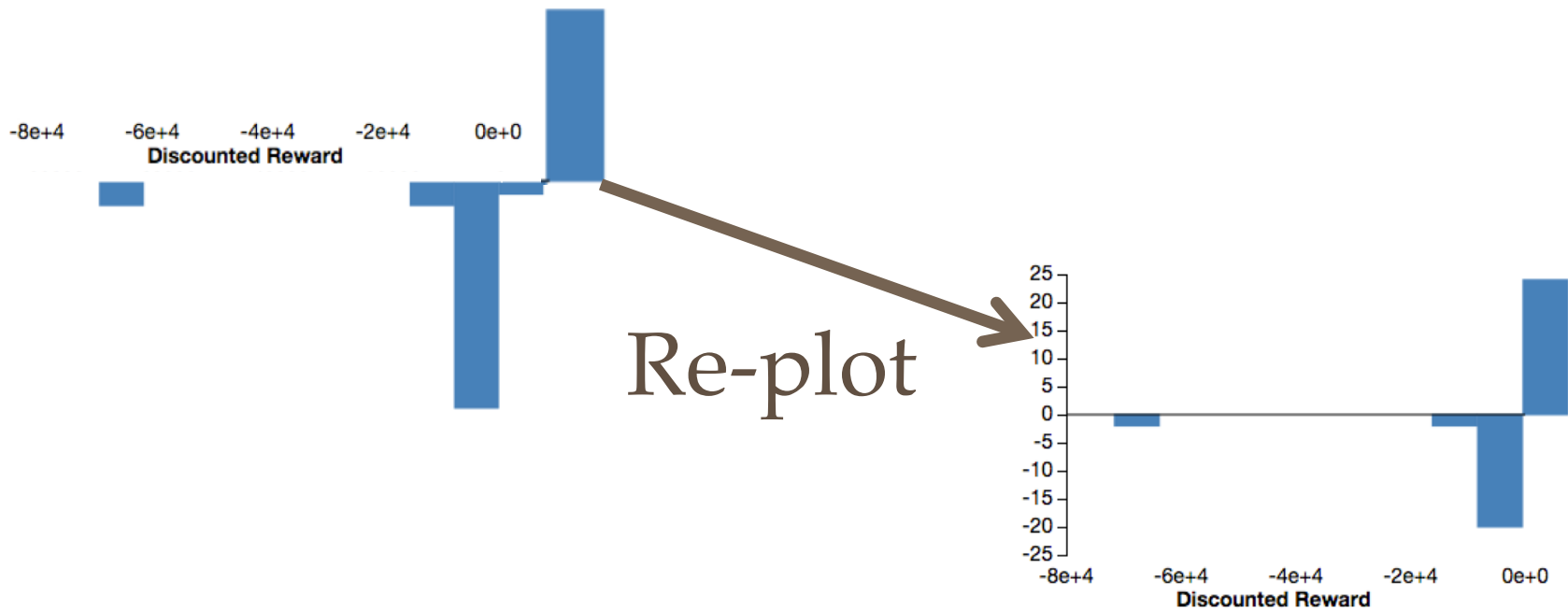


Take Difference in Counts

Comparison

$$\pi_1 - \pi_2$$

## State Variable Distributions for a Fixed Time Step



Comparison

$$\pi_1 - \pi_2$$

# State Variable Distributions for a Fixed Time Step



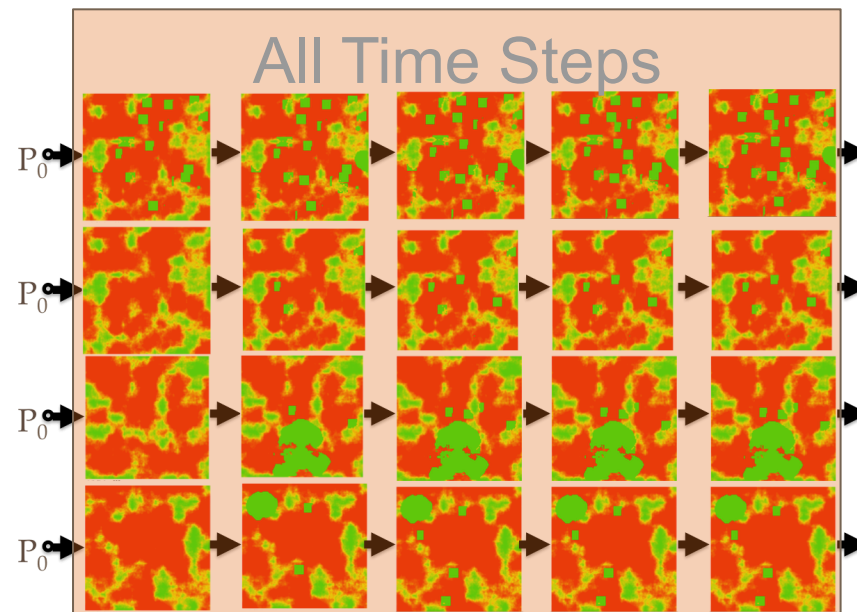
Let-Burn  
 Dominates  
 Suppress-All in  
 this time step



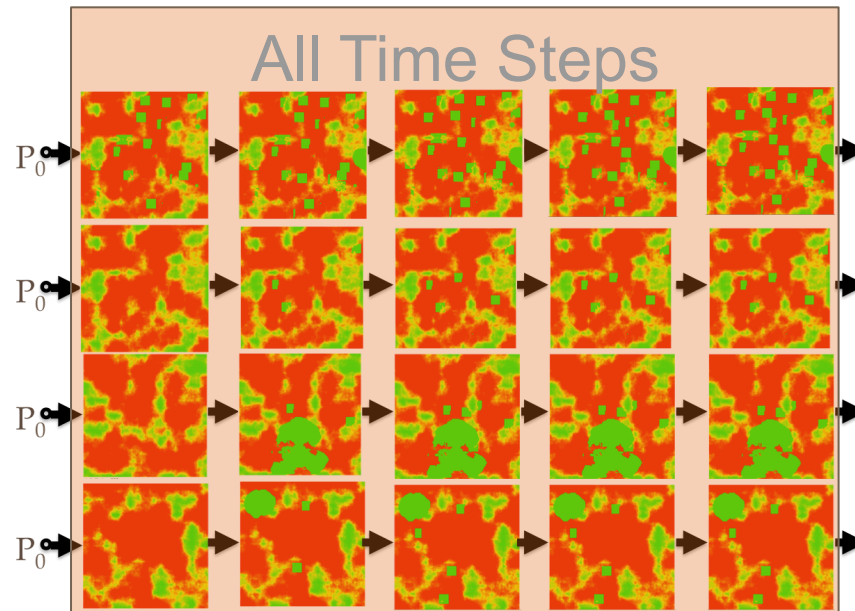
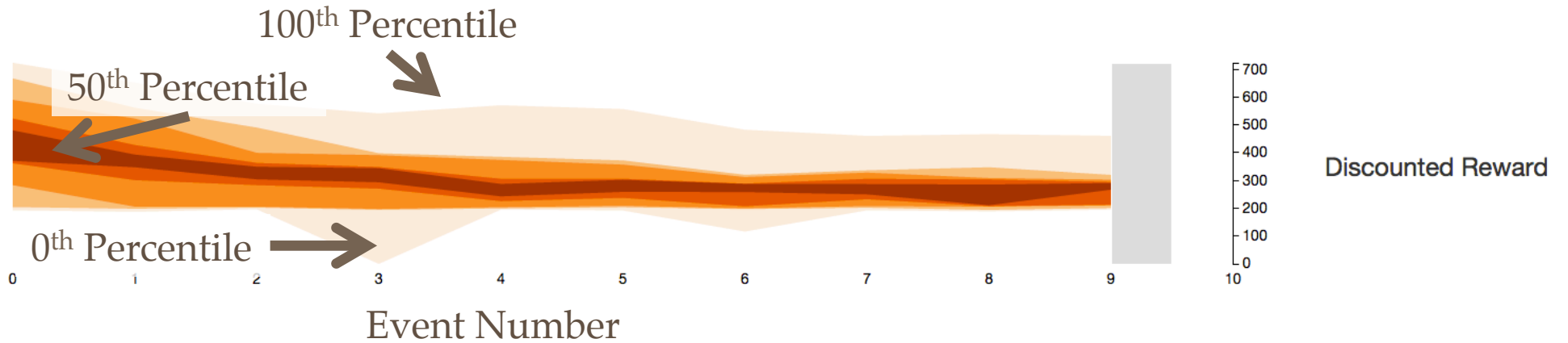
Comparison

$$\pi_1 - \pi_2$$

# State Variable Distributions through Time

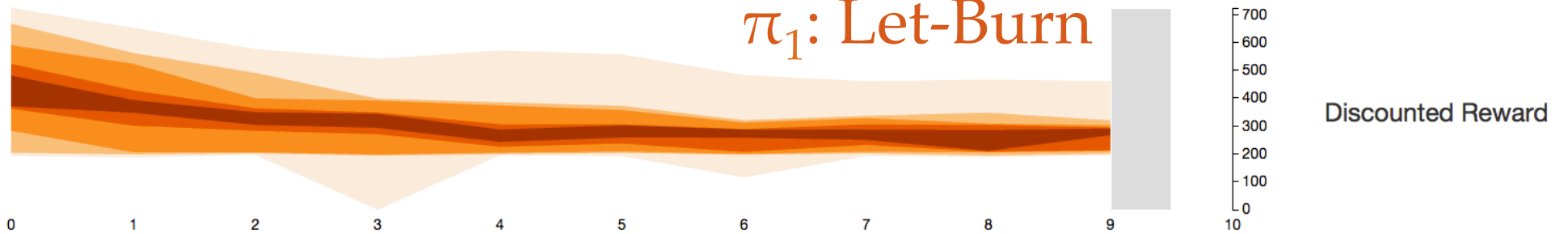


# State Variable Distributions through Time

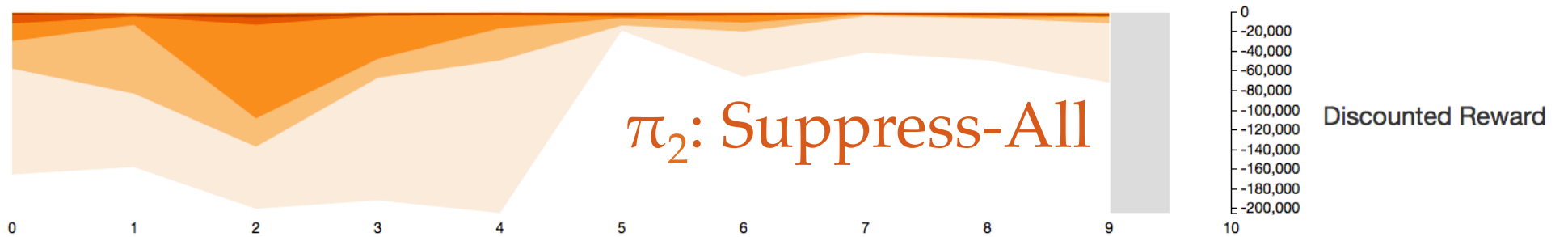


# State Variable Distributions through Time

$\pi_1$ : Let-Burn

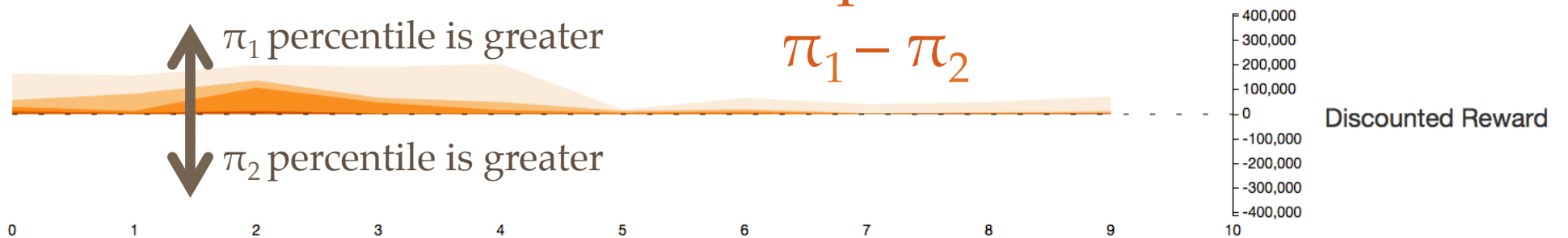


$\pi_2$ : Suppress-All



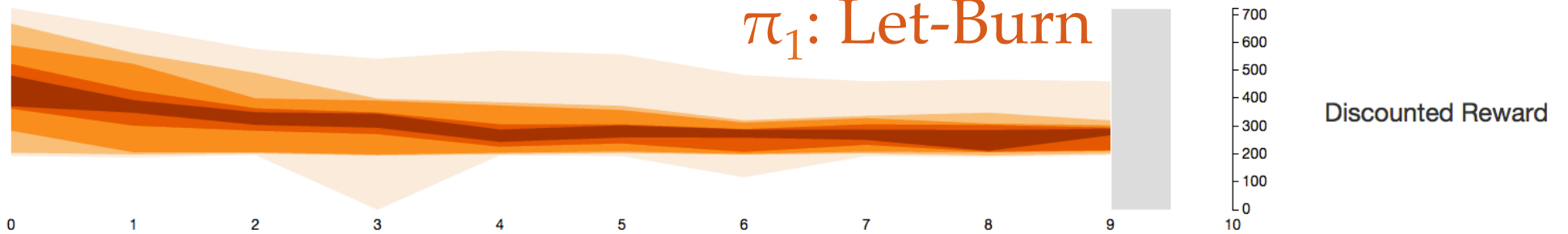
Comparison

$\pi_1 - \pi_2$

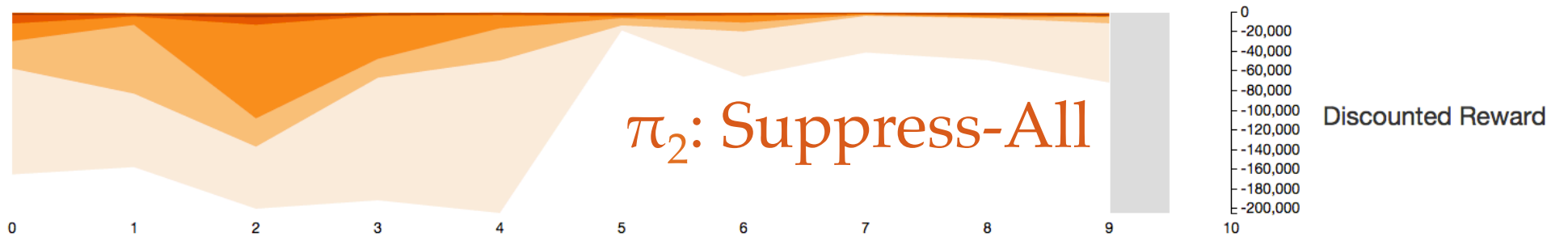


# State Variable Distributions through Time

$\pi_1$ : Let-Burn



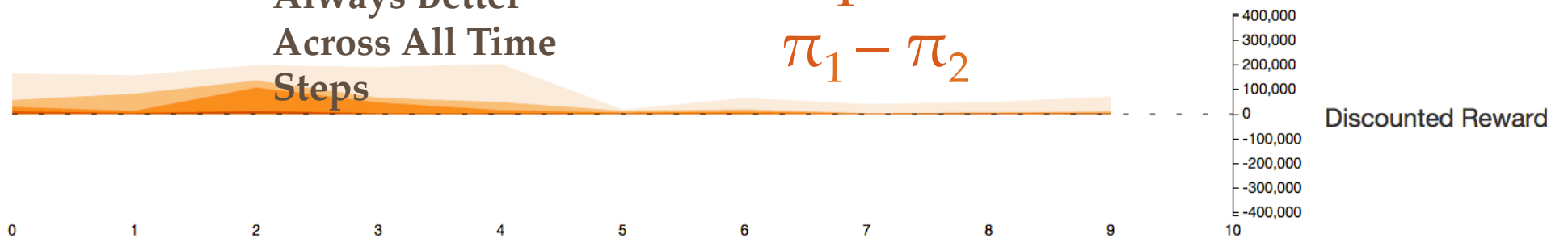
$\pi_2$ : Suppress-All



Let-Burn is  
Always Better  
Across All Time  
Steps

Comparison

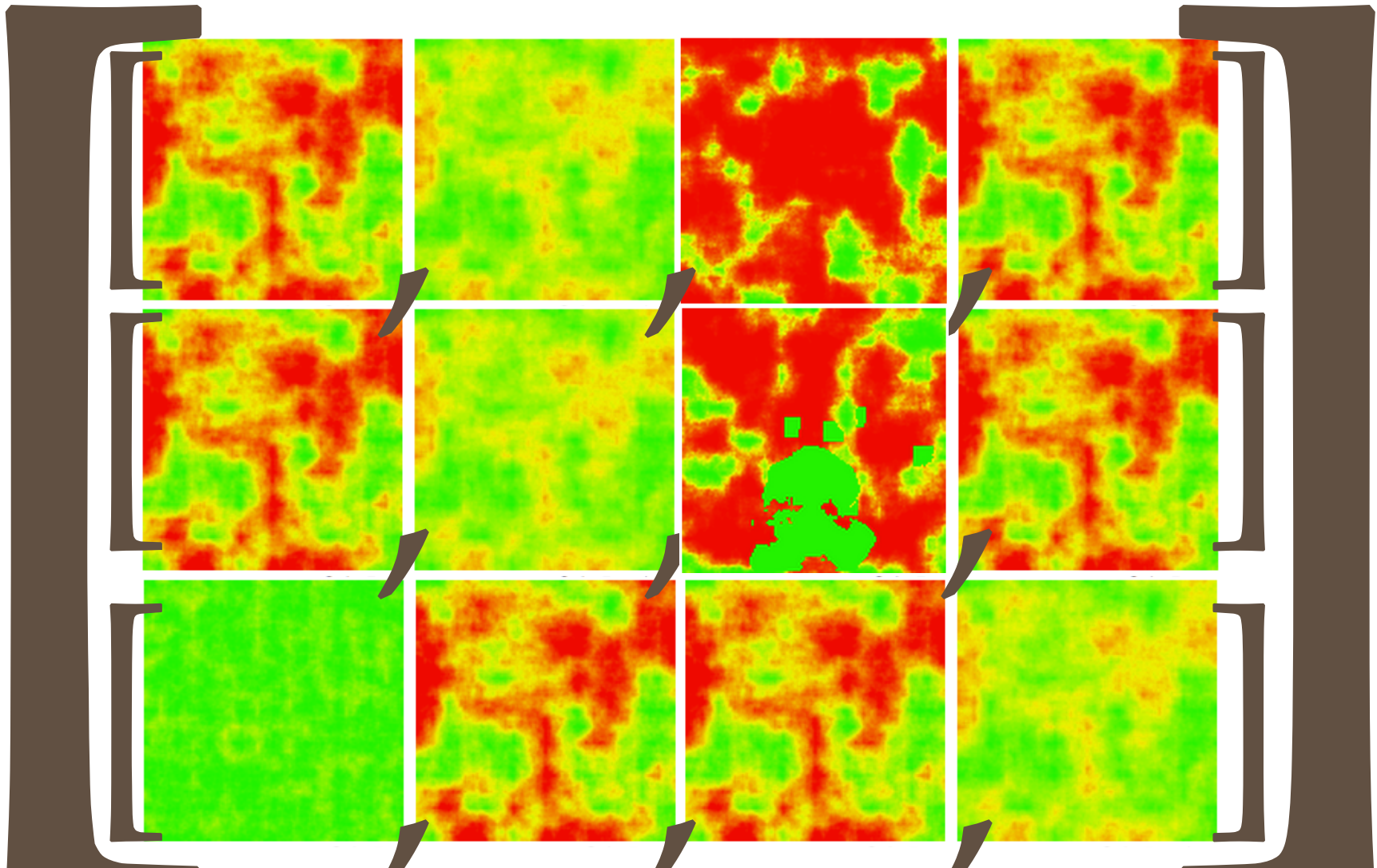
$\pi_1 - \pi_2$





## State details

Allow MDP Simulator to Generate State Visualizations



## Parameter Space Analysis (PSA)

*“[PSA] is the systematic variation of model input parameters, generating outputs for each combination of parameters, and investigating the relation between parameter settings and corresponding outputs.”*

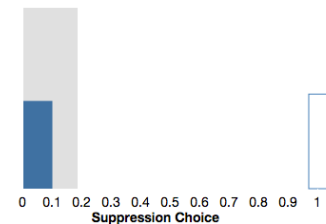
**Categories**  
Sensitivity  
Optimization  
Outliers  
Partition  
Uncertainty  
Fitting

Sensitivity · Optimization · Outliers · Partition · Uncertainty · Fitting

*Is the policy Sensitive to the state?*

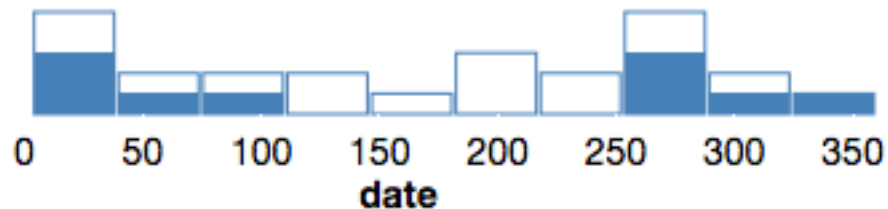
## Interaction

1. Brush suppression choice to select Let-Burn



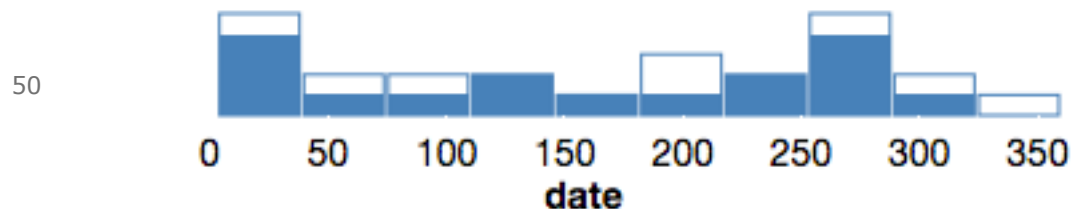
## Expectation

2. Date is a determinant of suppression choice



## Buggy Result

3. Date does not determine suppression choice



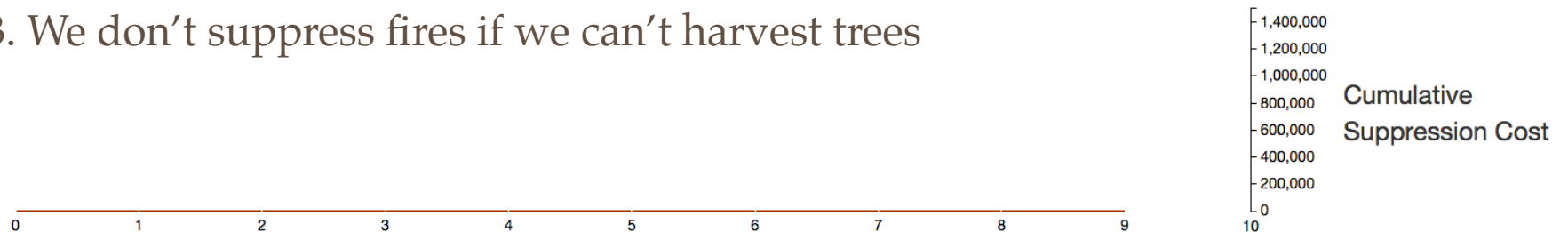
Sensitivity · **Optimization** · Outliers · Partition · Uncertainty · Fitting*Is the optimization sensitive to the reward signal?*

## Interaction

1. Zero-out harvest rewards %  Harvest Percent ⓘ2. Re-optimize and generate rollouts [↻ Optimize a New Policy](#)

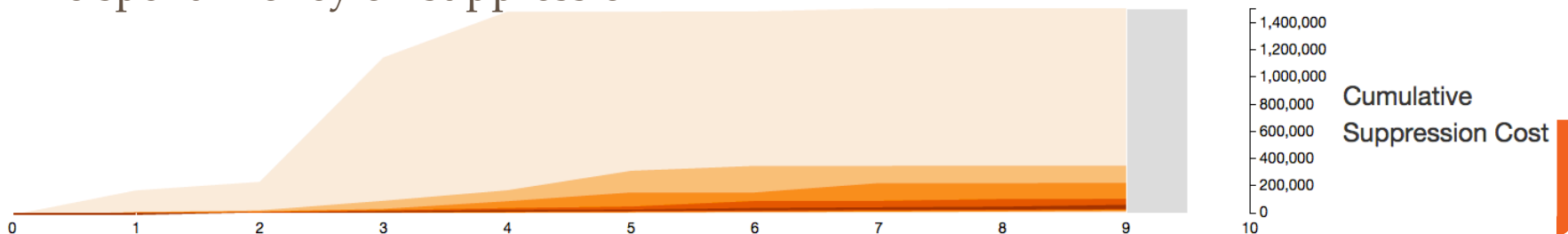
## Expectation

3. We don't suppress fires if we can't harvest trees



## Buggy Result

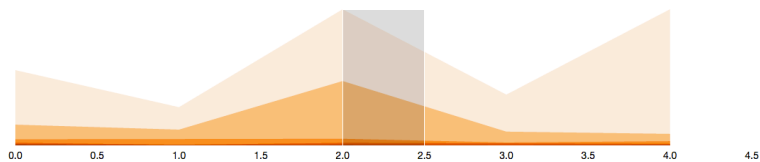
4. We spend money on suppression



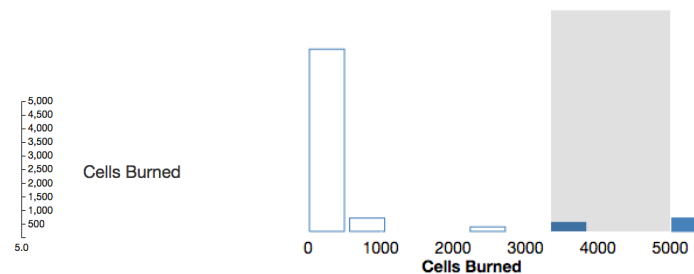
Sensitivity · Optimization · **Outliers** · Partition · Uncertainty · Fitting*Are the largest fires realistic?*

## Interaction

1. Change the year to the one with the largest fire

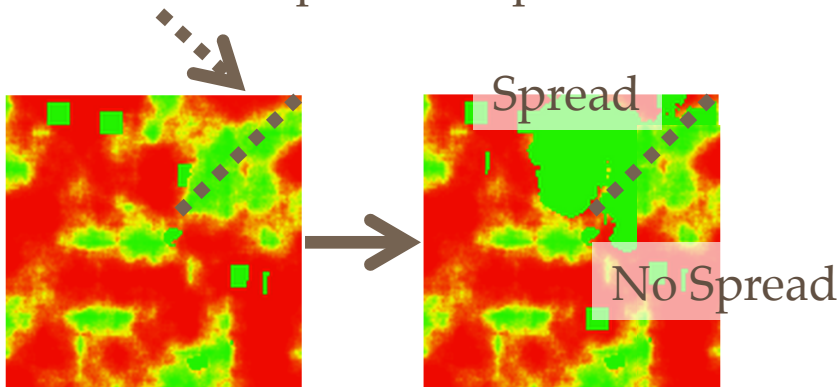


2. Brush histogram to view largest fires



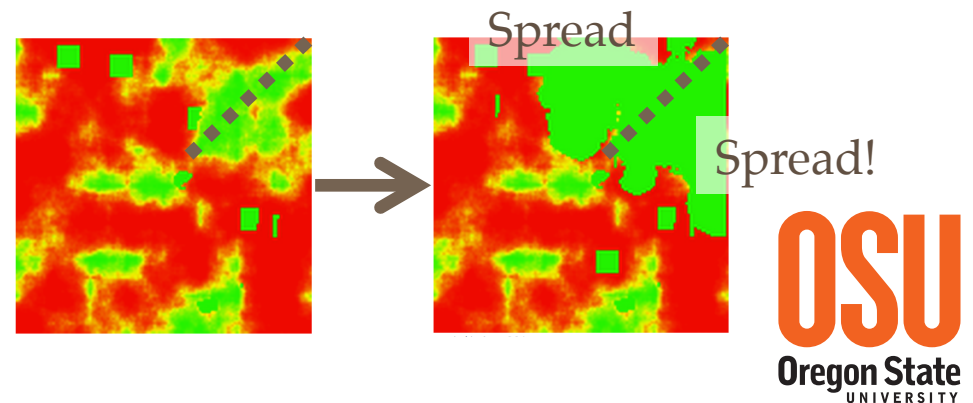
## Expectation

3. Fire break prevents spread



## Buggy Result

4. Fire break doesn't prevent spread



Sensitivity · Optimization · Outliers · **Partition** · Uncertainty · Fitting*Do policies partition the state space?*

## Interaction

1. Generate suppress-all rollouts
2. Generate let-burn-all rollouts
3. Click the “compare rollouts” button

Policy Definition Policy Definition

Expected Value \$ -570788.61

View Rollout Set 5 Compare To

Expected Value \$ 9129.08

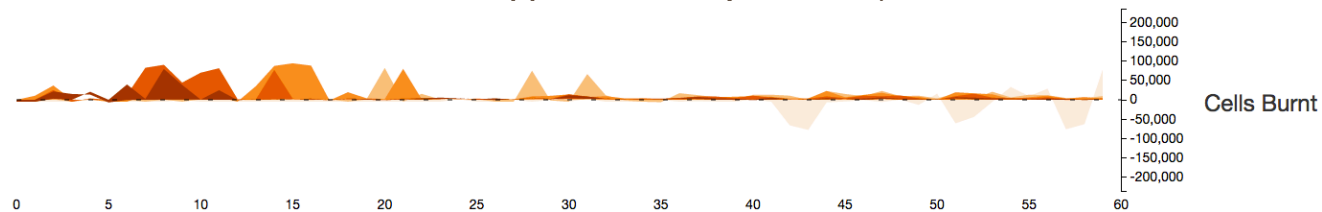
View Rollout Set 4 Compare To

10 Constant ③ -10 Constant ③

0 Date ③ 0 Date ③

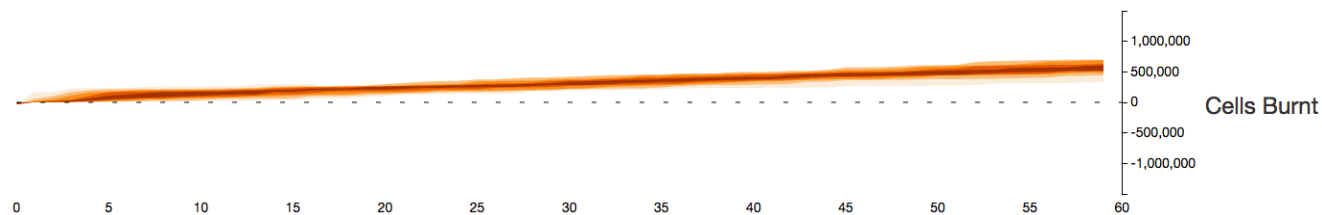
## Expectation

4. Let-burn-all fires will be larger in the present, and smaller in the future



## Buggy Result

5. Let-burn-all fires are the same in the present, and larger in the future



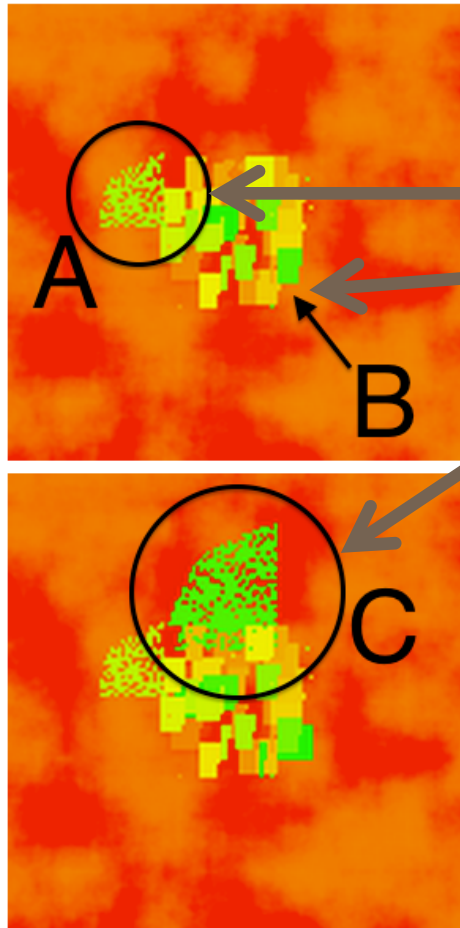
## Use Case Study of MDPvis

### **We tested a new wildfire policy domain**

- Visualization Developer: 1 Ph.D. Student in Computer Science
- New Fire Domain Developer: 1 Ph.D. Student in Forestry
- Wildfire Optimization Expert: 1 faculty research assistant

We found numerous bugs

## Evaluation of MDPvis



Viewed Largest Fires in Rollouts

Second Largest Fire

Harvest Areas

Largest Fire

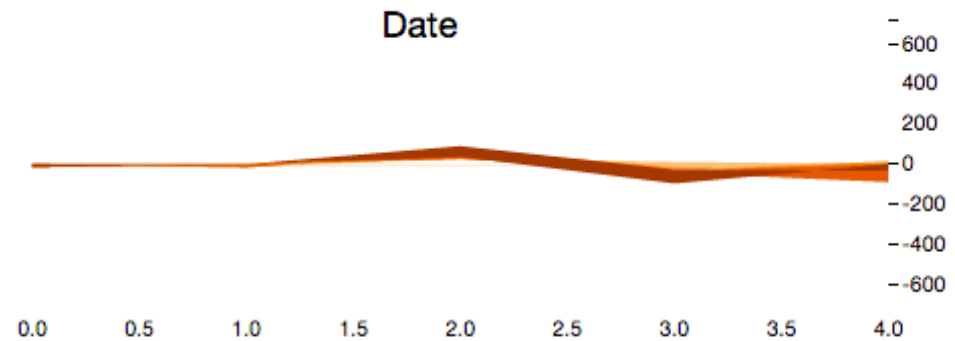
**Fires are not spreading east!**

Hidden except in most extreme fire by harvests



## Evaluation of MDPvis

1. **Compare:** Same model with different policies
2. **Expect:** Same ignition date in both rollout sets.
3. **Actual:** Policies change the weather.



## Conclusion

# Summary

We need visualization IDEs for MDPs!

## Interactive Demo

# MDPVis.github.io

\* Not robust to many *simultaneous* requests



## Thanks

- **Reviewers:** <you know who you are>
- **Advisor:** Thomas Dietterich
- **Research Group:** Ronald Metoyer, Claire Montgomery, Rachel Houtman, Mark Crowley, Hailey Buckingham
- **Funder:** National Science Foundation



# MDPVis.github.io

This material is based upon work supported by the National Science Foundation under Grant No. 1331932.

59

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

End.

# Questions?



# MDPVis.github.io

Contact

Email: [VLHCC@SeanBMcGregor.com](mailto:VLHCC@SeanBMcGregor.com)

Twitter: [@SeanMcGregor](https://twitter.com/SeanMcGregor)

End.

# Come to the Full Demo!



## MDPVis.github.io

Contact

Email: [VLHCC@SeanBMcGregor.com](mailto:VLHCC@SeanBMcGregor.com)

Twitter: [@SeanMcGregor](https://twitter.com/SeanMcGregor)

Sensitivity · Optimization · Outliers · Partition · **Uncertainty** · Fitting*How consistent is the policy for small changes to the model?*

## Interaction

1. Optimize and generate rollouts
2. Add air tankers to the model
3. Optimize and generate rollouts
4. Click the “Compare Rollouts” button

Optimize a New Policy

 **Model Modifiers**

% 0.5 Suppression Effect ⓘ

Optimize a New Policy

Expected Value \$ -570788.61

View Rollout Set 5

Compare To

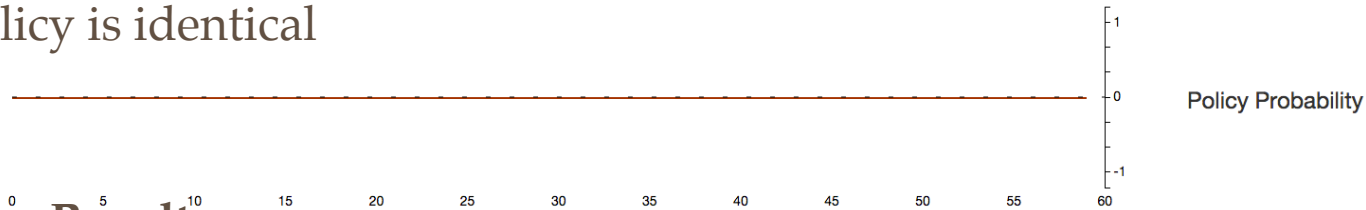
Expected Value \$ 9129.08

View Rollout Set 4

Compare To

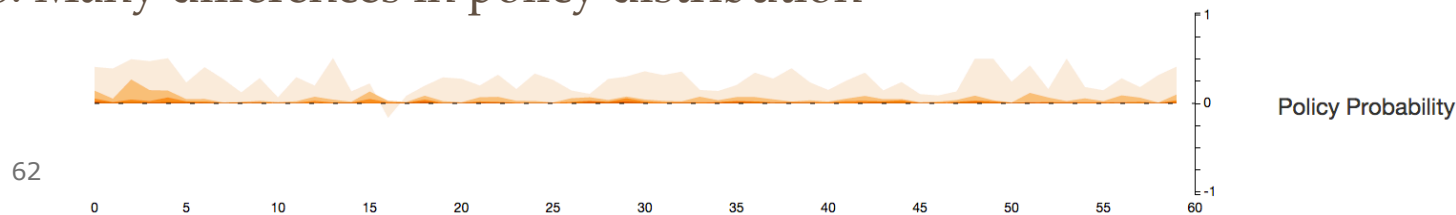
## Expectation

5. Policy is identical



## Buggy Result

6. Many differences in policy distribution



*Does the growth rate match the historical dataset?*

**Pre-Process**

1. Add a variable for the growth percentile within the historic data

**Expectation**

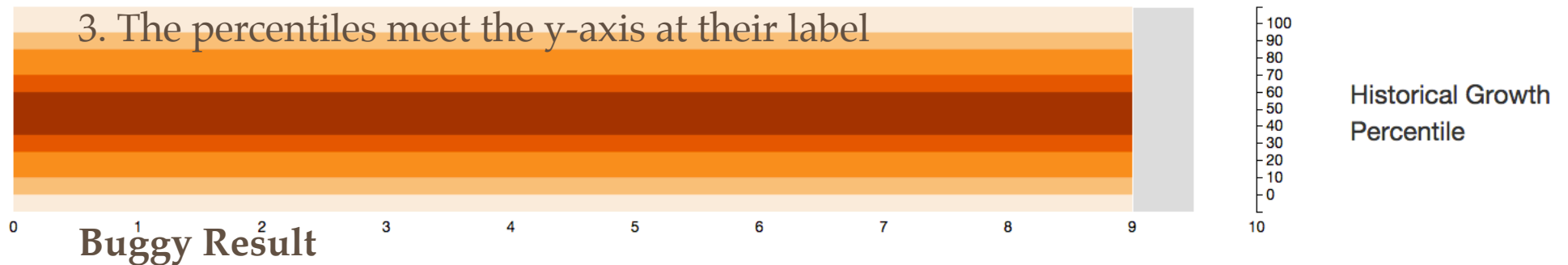
3. The percentiles meet the y-axis at their label

**Interaction**

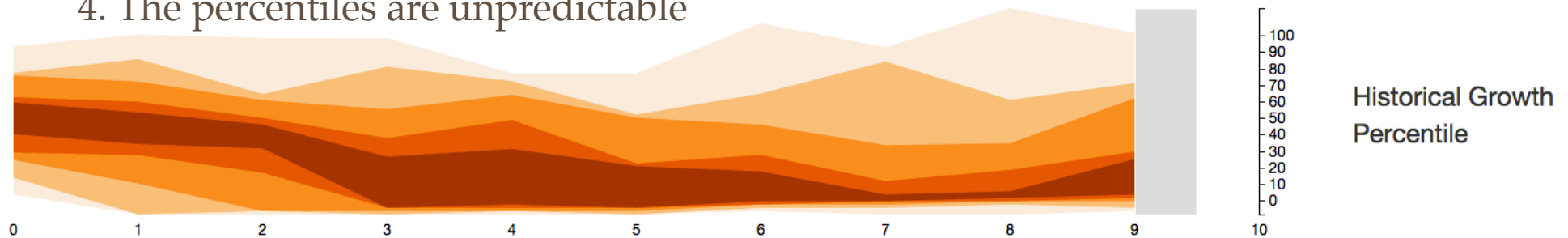
2. Assign the policy to the historical policy (suppress all)

**Policy Definition**

- 999 Constant ⓘ
- 0 Date ⓘ
- 0 Fuel Load ⓘ
- 0 Fuel Load 8 ⓘ



4. The percentiles are unpredictable





## Let's Construct a Simple MDP: "Pixel Forest"

