MDPvis: An Interactive Visualization for Testing Markov Decision Processes

Sean McGregor, Hailey Buckingham, Rachel Houtman, Claire Montgomery, Ronald Metoyer, and Thomas Dietterich
How did I get here?

- **2010**: Started with simulator building and optimization
- **2010 to Present**: Solve problems with slow and buggy software from foresters
- **2014**: Develop MDP visualizations for foresters
- **Today**: We also need tools for MDPs

Motivation

- Many sequential decision making problems combine complex models to optimize on Monte Carlo rollouts

Motivation

• Many sequential decision making problems combine complex models to optimize on Monte Carlo rollouts
• Models and MDP specification may be misspecified or poorly implemented

Motivation

• Many sequential decision making problems combine complex models to optimize on Monte Carlo rollouts
• Models and MDP specification may be misspecified or poorly implemented
• Want: better systems for understanding MDPs and testing for bugs

Outline

1. Wildfire Suppression MDP Example
   - Basic Introduction
   - Testing
2. MDPvis
   - Design
   - Testing Examples
   - MDPvis Use Case Study
   - Integrating Your Domain or Optimizer
3. Concluding
Starting in 1935, the United States adopted the “10 AM policy”

We need a more nuanced approach.


## Modeling Wildfire

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

Represents a challenging and general class of MDPs
- High Dimensional States
- Large State Space
- Integrates Several Simulators
\( P_0 \)
Start with Today’s Landscape

\[ P_0 \]
### MDPs: Basic Introduction

<table>
<thead>
<tr>
<th>Simulators</th>
<th>Optimizer</th>
<th>Rewards</th>
<th>Policy</th>
</tr>
</thead>
</table>

- **Generate an ignition and weather**

- **$P_0$**

---

**Notes:**
- This page from a presentation or document is focused on Markov Decision Processes (MDPs). The table lists components of an MDP, including simulators, optimizer, rewards, and policy. The diagram illustrates a process related to generating an ignition and weather scenario, likely part of a simulation or optimization process. The notation $P_0$ might represent an initial state or condition in the modeled system.
## MDPs: Basic Introduction

<table>
<thead>
<tr>
<th>Simulators</th>
<th>Optimizer</th>
<th>Rewards</th>
<th>Policy</th>
</tr>
</thead>
</table>

Generate an ignition and weather

$P_0$
MDPs: Basic Introduction

| Simulators | Optimizer | Rewards | Policy |

Select an Action

$P_0$
## MDPs: Basic Introduction

<table>
<thead>
<tr>
<th>Simulators</th>
<th>Optimizer</th>
<th>Rewards</th>
<th>Policy</th>
</tr>
</thead>
</table>

**Update Vegetation for Wildfire**

$P_0$
MDPs: Basic Introduction

<table>
<thead>
<tr>
<th>Simulators</th>
<th>Optimizer</th>
<th>Rewards</th>
<th>Policy</th>
</tr>
</thead>
</table>

Fire Suppression Effort

$95,000$

Fire Suppression Costs

$P_0$
### MDPs: Basic Introduction

<table>
<thead>
<tr>
<th>Simulators</th>
<th>Optimizer</th>
<th>Rewards</th>
<th>Policy</th>
</tr>
</thead>
</table>

- **Update Vegetation for harvest**
- **$20,000**
- **Harvest Revenue**

**$P_0$**

---

*OSU Oregon State University*
**MDPs: Basic Introduction**

<table>
<thead>
<tr>
<th>Simulators</th>
<th>Optimizer</th>
<th>Rewards</th>
<th>Policy</th>
</tr>
</thead>
</table>

Generate an ignition and weather

\[ P_0 \]
### MDPs: Basic Introduction

<table>
<thead>
<tr>
<th>Simulators</th>
<th>Optimizer</th>
<th>Rewards</th>
<th>Policy</th>
</tr>
</thead>
</table>

Select an Action

$P_0$ → Suppress

**OSU Oregon State University**
Update Vegetation for Wildfire
MDPs: Basic Introduction

<table>
<thead>
<tr>
<th>Simulators</th>
<th>Optimizer</th>
<th>Rewards</th>
<th>Policy</th>
</tr>
</thead>
</table>

Fire Suppression Effort

Fire Suppression Costs

$\text{(15,000)}$
<table>
<thead>
<tr>
<th>Simulators</th>
<th>Optimizer</th>
<th>Rewards</th>
<th>Policy</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_0$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Update Vegetation for Harvest

Harvest Revenue

$20,000$
<table>
<thead>
<tr>
<th>Simulators</th>
<th>Optimizer</th>
<th>Rewards</th>
<th>Policy</th>
</tr>
</thead>
</table>

(Continue Until Reaching the Horizon)
A High Dimensional Probabilistic Time Series

...And this is just one of many!
<table>
<thead>
<tr>
<th>Simulators</th>
<th>Optimizer</th>
<th>Rewards</th>
<th>Policy</th>
</tr>
</thead>
</table>

**Monte Carlo Rollouts**

![Monte Carlo Rollouts Diagram](image-url)
All visited states influence optimizer
MDPs: Basic Introduction

Simulators | Optimizer | Rewards | Policy

Update Policy

$P_0$ $\rightarrow$ $\rightarrow$ $\rightarrow$ $\rightarrow$ $\rightarrow$

$P_0$ $\rightarrow$ $\rightarrow$ $\rightarrow$ $\rightarrow$ $\rightarrow$

$P_0$ $\rightarrow$ $\rightarrow$ $\rightarrow$ $\rightarrow$ $\rightarrow$

$P_0$ $\rightarrow$ $\rightarrow$ $\rightarrow$ $\rightarrow$ $\rightarrow$
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
Testing and Debugging Process

1. Generate Rollouts

2. Visualize the data

3. Change Parameters
Outline

1. Wildfire Suppression MDP Example
   Basic Introduction
   Testing

2. MDPvis
   Design
   Testing Examples
   MDPvis Use Case Study
   Integrating Your Domain or Optimizer

3. Concluding
Introducing MDPvis
What are the elements of the MDPvis design?

Parameters

History

Distributions at Time Step

Distributions Through Time

State Snapshots
Parameter Areas

**Reward Specification**
- Discount
- Suppression Fixed Cost
- Suppression Variable Cost

**Model Modifiers**
- Years to simulate
- Futures to simulate
- Landscape Size
- Harvest Percent
- Slash Remaining
- Fuel Accumulation
- Fuel Load 6
- Fuel Load 24

**Policy Definition**
- Constant
- Date
- Days Left
- Temperature
- Wind Speed

**Exploration History**
- Expected Value 8: 46040.22
- Expected Value 9: 147446.26

[Optimize a New Policy]
### History Area

#### Reward Specification
- Discount
- Suppression Fixed Cost
- Suppression Variable Cost

#### Model Modifiers
- Years to simulate
- Futures to simulate
- Landscape Size
- Harvest Percent
- Blash Remaining
- Fuel Accumulaton
- Suppression Effect
- Use Original Bugs

#### Policy Definition
- Constant
- Date
- Days Left
- Temperature
- Wind Speed
- Fuel Load 8
- Fuel Load 24

#### Exploration History
- Rollouts are Current
- Expected Value $-640252.42
- Expected Value $-47446.26
- Expected Value $-42404.22

[Optimize a New Policy]
Visualization Areas

- Initial State Distribution
- Cumulative Suppression Cost
- Cumulative Growth
- Fuel Load, avg 24

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

\( \pi_1: \text{Let-Burn} \quad \pi_2: \text{Suppress-All} \quad \text{Comparison} \quad \pi_1 - \pi_2 \)
State Variable Distributions through Time
State Variable Distributions through Time

100th Percentile

50th Percentile

0th Percentile

Event Number

Discounted Reward

All Time Steps
State Variable Distributions through Time

$\pi_1$: Let-Burn

$\pi_2$: Suppress-All

Comparison

$\pi_1$ percentile is greater

$\pi_2$ percentile is greater
State Variable Distributions through Time

\( \pi_1: \text{Let-Burn} \)

\( \pi_2: \text{Suppress-All} \)

Comparison

\( \pi_1 - \pi_2 \)

Let-Burn is Always Better Across All Time Steps
State details

Allow MDP Simulator to Generate State Visualizations
Outline

1. Wildfire Suppression MDP Example
   - Basic Introduction
   - Testing
2. MDPvis
   - Design
   - Testing Examples
   - MDPvis Use Case Study
   - Integrating Your Domain or Optimizer
3. Concluding
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

Is the suppression decision sensitive to the date of the fire?

Interaction

1. Select states where fire is allowed to burn

Expectation

2. Date is a determinant of suppression choice

Buggy Result

3. Date does not determine suppression choice
Is the optimization sensitive to the reward signal?

**Interaction**

1. Zero-out harvest rewards
2. Re-optimize and generate rollouts

**Expectation**

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

**Buggy Result**

4. We spend money on suppression
Does the let-burn policy have bigger initial fires and smaller subsequent fires?

**Interaction**

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

**Expectation**

4. Fires will be larger in the present, and smaller in the future

**Buggy Result**

5. Fires are the same in the present, and larger in the future
Outline

1. Wildfire Suppression MDP Example
   - Basic Introduction
   - Testing
2. MDPvis
   - Design
   - Testing Examples
   - MDPvis Use Case Study
   - Integrating Your Domain or Optimizer
3. Concluding
Use Case Study of MDPvis

We tested a new wildfire policy domain

Wildfire Optimization Expert
(Faculty Research Assistant)

Visualization Developer
(Ph.D. Student in Computer Science)

New Fire Domain Developer
(Ph.D. Student in Forestry)

We found numerous bugs
Evaluation of MDPvis

Fires are not spreading east!
Hidden except in most extreme fire by harvests
Evaluation of MDPvis

Interaction
1. Compare rollouts from two policies

Expectation
2. Fire dates do not change between policies

Buggy Result
3. Policies choose the weather
Outline

1. Wildfire Suppression MDP Example
   Basic Introduction
   Testing
2. MDPvis
   Design
   Testing Examples
   MDPvis Use Case Study
   Integrating Your Domain or Optimizer
3. Concluding
Integrating MDPvis

4 HTTP Requests
1. /initialize
2. /rollouts
3. /optimize (optional)
4. /state (optional)
Integrating MDPvis

/initialize

```
"reward": [
  {"name": "Discount",
   "description": "The per-year discount",
   "current_value": 1, "max": 1, "min": 0, "units": "-"},
  {"name": "Suppression Fixed Cost",
   "description": "cost per day of suppression",
   "current_value": 500, "max": 999999, "min": 0, "units": "$"}
],

"transition": [
  {"name": "Years to simulate",
   "description": "how far to look into the future",
   "current_value": 10, "max": 150, "min": 0, "units": "Y"},
  {"name": "Futures to simulate",
   "description": "how many stochastic futures to generate",
   "current_value": 25, "max": 1000, "min": 0, "units": "#"}
],

"policy": [
  {"name": "Constant",
   "description": "for the intercept",
   "current_value": 0, "max": 10, "min": -10, "units": ""},
  {"name": "Date",
   "description": "for each day of the year",
   "current_value": 0, "max": 10, "min": -10, "units": ""}
]
```
Integrating MDPvis

/rollouts

```python
def rollouts(query):
    rollouts = []
    for rollout_number in range(0, 200):
        rollout = getRollouts(rollout_number, query)
        formatted_rollout = formatRollout(rollout)
        rollouts.append(formatted_rollout)
    return rollouts
```

/optimize

```python
def optimize(query):
    updated_parameters = optimize(query)
    return updated_parameters
```

/state

```python
def state(query):
    image_urls = getImages(query["rollout_number"], query)
    return image_urls
```
Integrating MDPvis

Connect the **remotely hosted** visualization to your **locally hosted** simulator and optimizer.
Conclusion

Figure 1: Learning curves for MBIE, Q-learning, and DDV as measured by confidence bounds on $V(s_0)$. 

Interactive Demo

MDPVis.github.io

* Not robust to many simultaneous requests
Thanks

- **Advisor:** Thomas Dietterich
- **Research Group:** Ronald Metoyer, Claire Montgomery, Rachel Houtman, Mark Crowley, Hailey Buckingham
- **Funder:** National Science Foundation

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

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.
Outline

1. Wildfire Suppression MDP Example
   - Basic Introduction
   - Testing
2. MDPvis
   - Design
   - Integrating Your Domain or Optimizer
   - Testing Examples
   - MDPvis Use Case Study
3. Concluding
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

**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

**Buggy Result**
4. The percentiles are unpredictable

**Interaction**
2. Assign the policy to the historical policy (suppress all)
1. **Build understanding** of how policy performs
2. **Explore distributions and filter** to interesting rollouts
3. Easy integration of **your custom visualizations**
4. **Shorter experiment/analysis cycle** by connecting tools directly to implementation
5. Parameterizations are **shareable**
6. **Simple integration** with existing domains