Equilibrium Policy Gradients for Spatiotemporal Planning

Mark Crowley
crowley@eecs.oregonstate.edu

Ph.D Computer Science, University of British Columbia
Postdoc at Oregon State University

Western Forest Economists Conference 2012
Newport, Oregon
June 12, 2012

Spatiotemporal Planning

Decision making where actions need to be taken at many locations at each point in time

- We want to maximize some utility to find the best policy possible in the time available
- Demonstration Problem: Forestry Management under cutting constraints
  - spatial harvest constraints
  - flow constraints
  - presence of Mountain Pine Beetle

Goal

- Provide useful **tools** for human planners to use as a first approximation or as a sanity check
- Allow people to deal with larger problems
- Model uncertainty directly
- Utilize existing simulations/dynamics
- Produce a general tool for spatial planning in large domains with complex values like ecology/ forestry

Translations

Sometimes we use different terms to mean the same things, a short list

- State - current landscape
- Features - properties, attributes which describe the current state of the world
- Rewards - values, objectives, costs
- Cells - locations, stands, management units

Origin of Topic

- Discussions with researchers at UBC Forestry who were looking for ways to use off-the-shelf modelling tools in planning
- Need from Forestry planning to handle uncertainty and spatial correlations
- Planning in the presence of Mountain Pine Beetle infestation demands spatially correlated actions

Richer Methods Needed

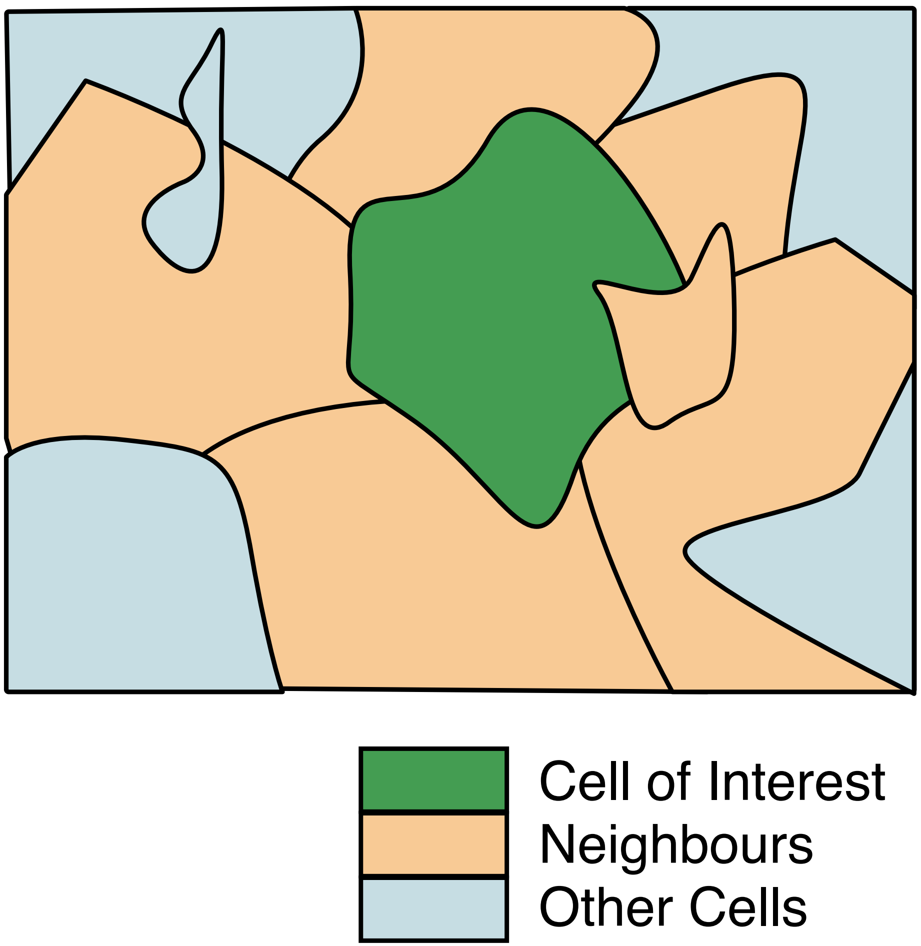
- There is a need for scalable, model-free planning methods that can take advantage of existing simulators for natural resource planning problems. [Forsell et al., 2009][Kimmins et al., 2005] [Baskent and Keles, 2005]

Forestry Planning Problem

A *landscape* is divided up into *cells (stands)*.

Each cell has :

- actions : what to do in a cell
- a set of neighbouring cells
- features : describing the local conditions

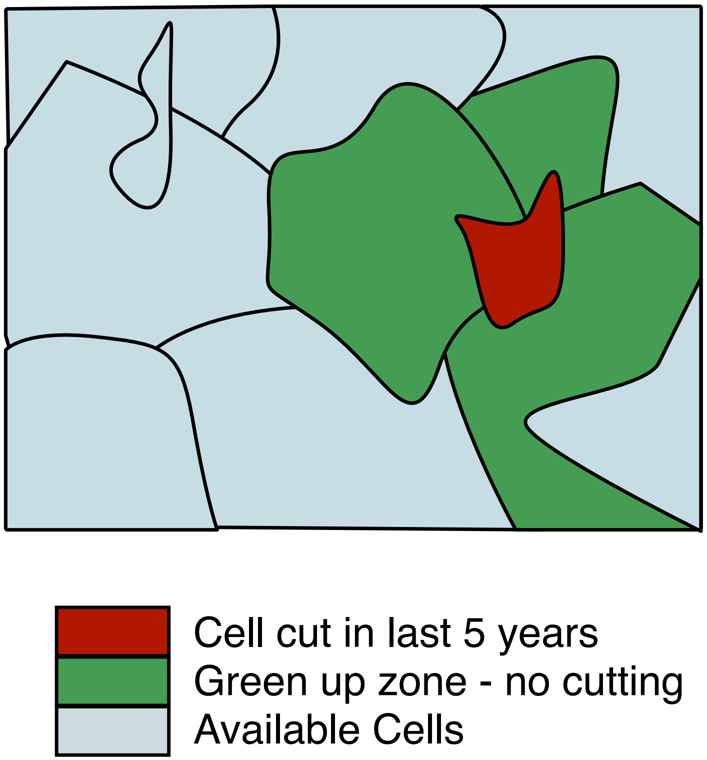


Cell Actions

- clear cut - cut all trees in a cell this year
- don't cut - do nothing this year

Constraints on Actions:

- Don't cut adjacent cells
- Annual allowable cut
- Even flow constraints
- Budget

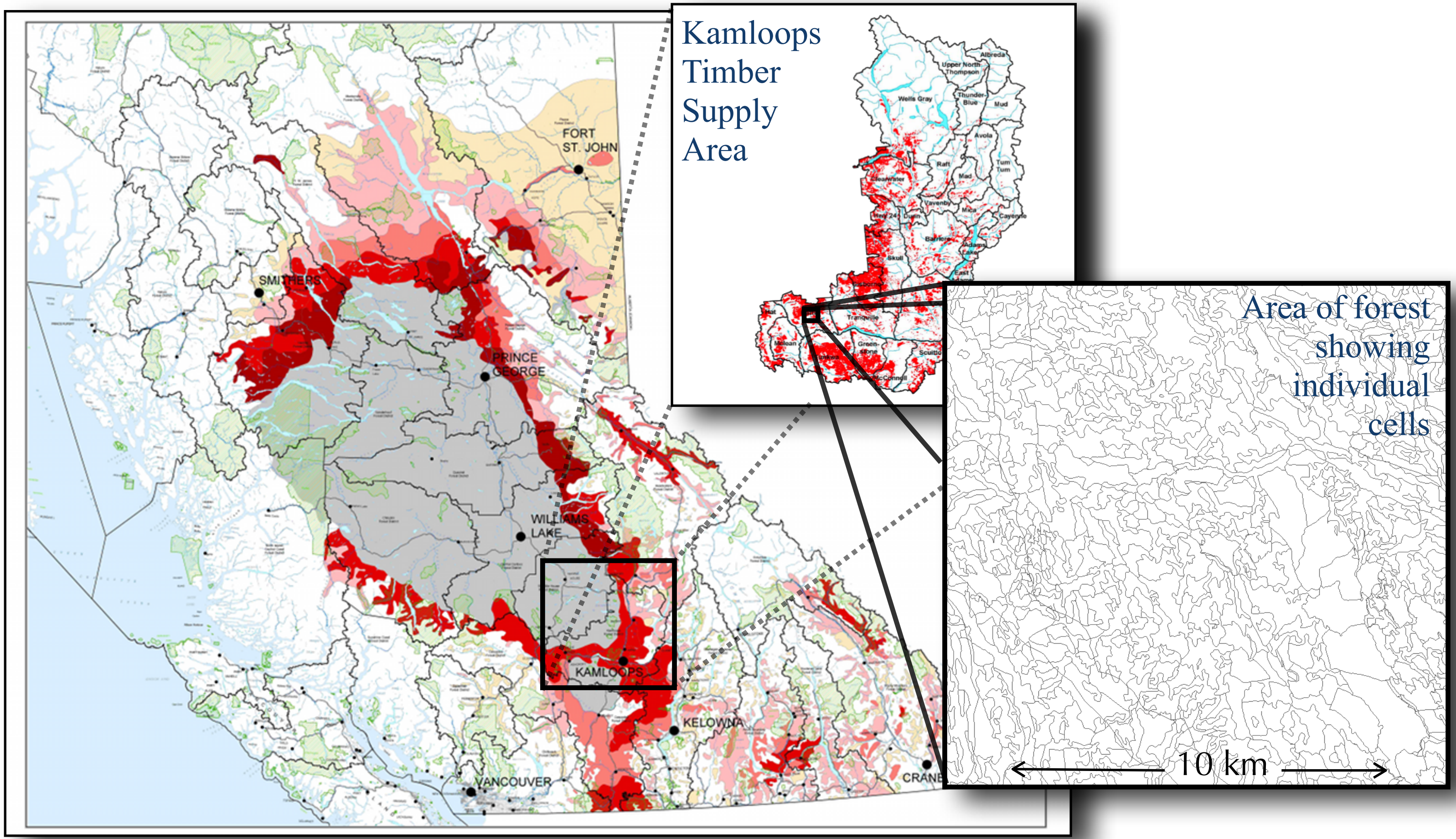


Features of Cells

Number of features generally small ~30

- area - continuous (1-50 ha)
- **age of trees** - discrete (0-300 years)
- **volume** - continuous (0 - 100,000 m^3)
- slope, soil type, climatic zone
- dominant trees species - discrete (pine, fir, birch, spruce)
- presence of roads, water, protected wildlife
- MPB infestation risk, fire risk - discrete (low, med, high)
- **max adjacent volume** - real (0 - 100,000 m^3)
- **any adjacent cut** - Boolean

Very Large Scale



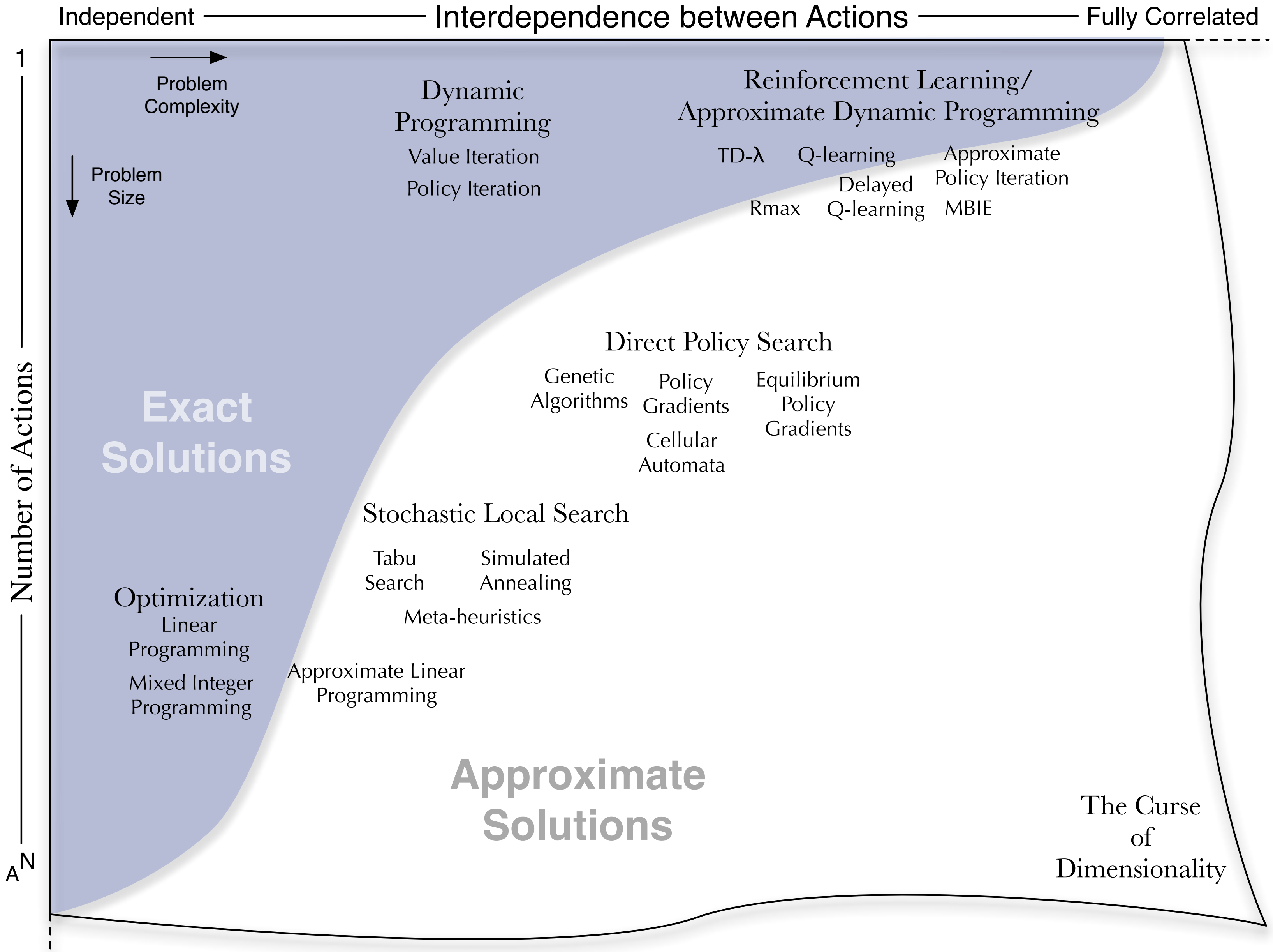
Common Size : ~100,000 cells ~300 years

Spatial Planning as an Markov Decision Process (MDP)

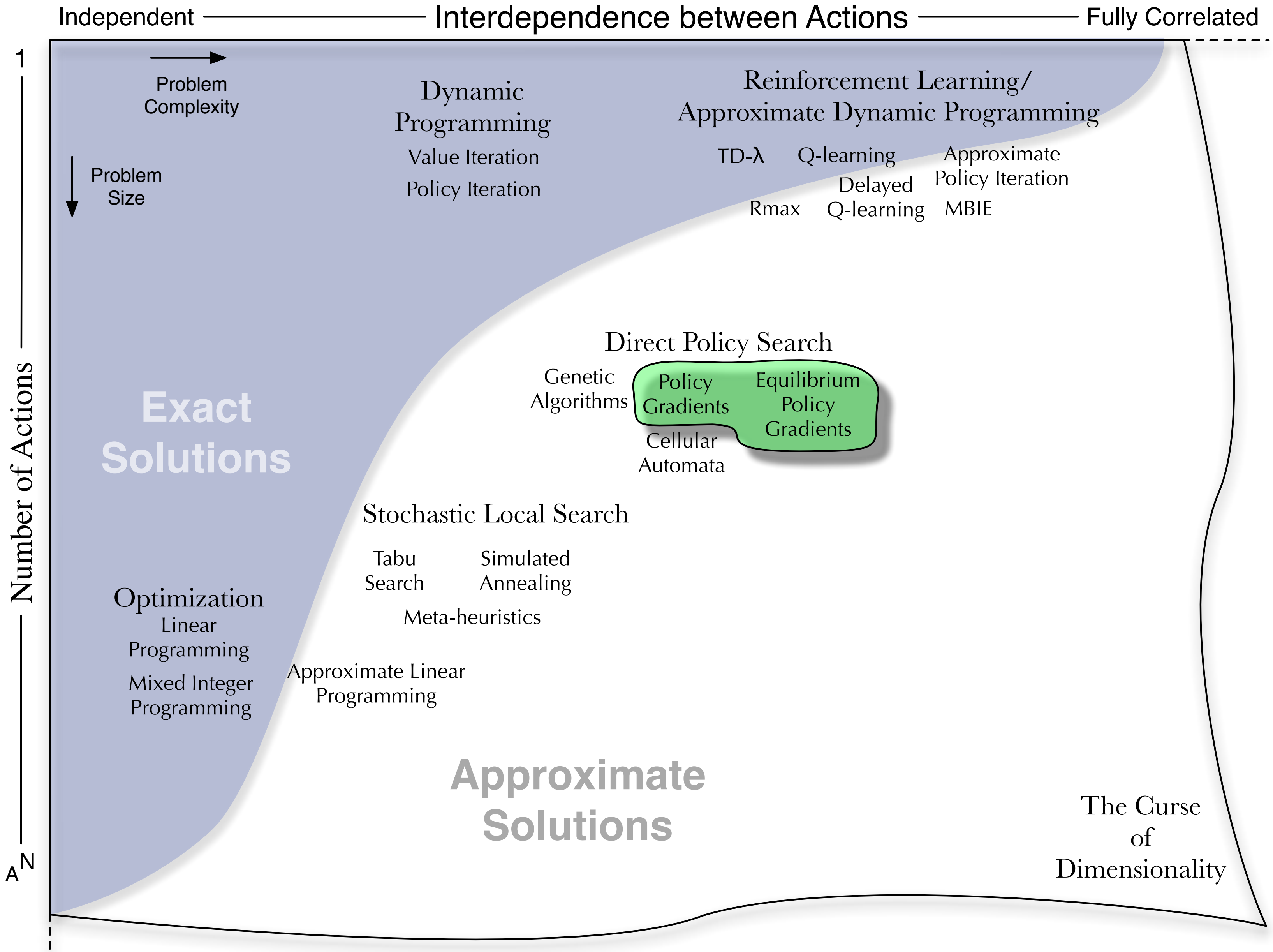
States	$\mathbf{s} \in \mathbf{S}$
Actions	$\mathbf{a} \in \mathbf{A}$
Transition Dynamics	$\mathcal{T}(\mathbf{s}^t, \mathbf{a}^t, \mathbf{s}^{t+1}) : \mathbf{S} \times \mathbf{A} \rightarrow \delta(\mathbf{S})$
Rewards	$r(\mathbf{s}^t, \mathbf{a}^t, \mathbf{s}^{t+1}) : \mathbf{S} \times \mathbf{A} \times \mathbf{S} \rightarrow \Re$

Policy	$\Pi(\mathbf{a}\vert\mathbf{s}) : \mathbf{S} \rightarrow \delta(\mathbf{A})$

Map of Algorithms for Spatial Planning



Map of Algorithms for Spatial Planning



Spatial Planning as an MDP

Cells	C		
Cell States	$\mathbf{s}_c \in S$	Landscape States	$\mathbf{S}$ is S^C
Cell Actions	$\mathbf{a}_c \in A$	Landscape Actions	$\mathbf{A}$ is A^C
Transition Dynamics	$\mathcal{T}(\mathbf{s}^t, \mathbf{a}^t, \mathbf{s}^{t+1}) : \mathbf{S} \times \mathbf{A} \rightarrow \delta(\mathbf{S})$		
Rewards	$r(\mathbf{s}^t, \mathbf{a}^t, \mathbf{s}^{t+1}) : \mathbf{S} \times \mathbf{A} \times \mathbf{S} \rightarrow \Re$		

Policy	$\Pi(\mathbf{a}\|\mathbf{s}) : \mathbf{S} \rightarrow \delta(\mathbf{A})$

Challenge I : Cannot Enumerate States or Actions

Consider a simple problem with

- 1000 cells
- 10 Binary Features
- Binary Actions

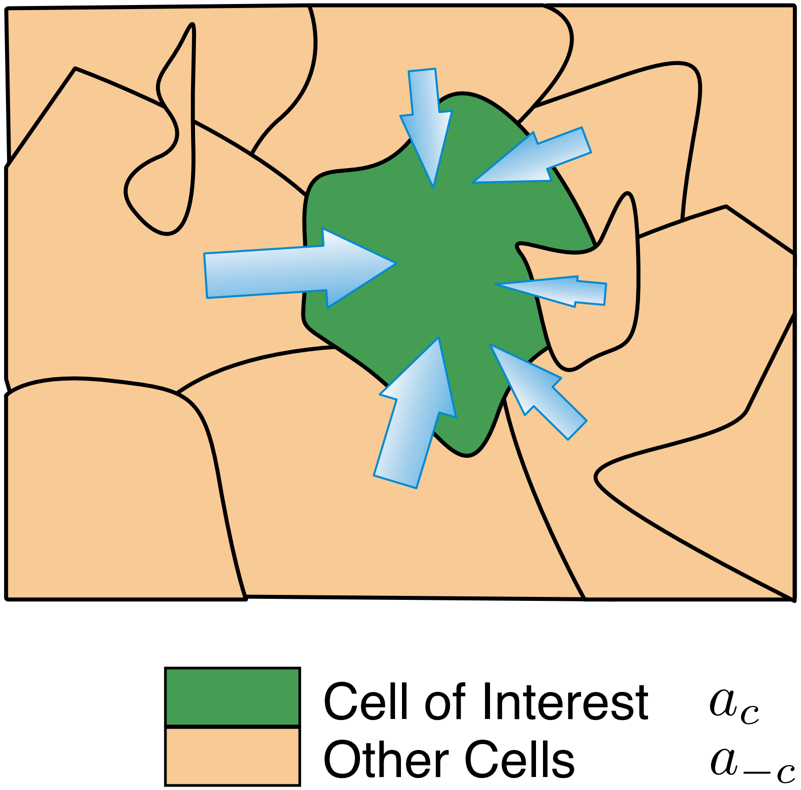
Number of ...	at each cell	entire landscape
actions	2	$2^{1000} \approx 10^{300}$
states	2^{10}	$(2^{10})^{1000} \approx 10^{3000}$

Challenge II : Cannot Analyse Dynamics Directly

- Best models are external simulators built by experts in the domain.
- Often designed to explore scenarios by manually adjusting parameters.
- We want to be able to treat the simulator as a black box.

Challenge III : Cannot Treat Locations as Independent

- spatial constraints - no adjacent cutting
- spatial dynamics - MPB spread
- non-local rewards - global cutting quota



Policy Gradient Planning

PG Planning is : Gradient descent in policy parameter space

$$\Pi(\mathbf{a}|\mathbf{s},\theta) : \mathbf{S} \rightarrow \delta(\mathbf{A})$$

All we need to be able to do is:

- requires sampling policy $\Pi(\mathbf{a}|\mathbf{s},\theta)$
- acquiring next state from $\mathcal{T}(\mathbf{s}^t, \mathbf{a}^t, \mathbf{s}^{t+1})$
- compute rewards on these samples
- compute the gradient of $\Pi(\mathbf{a}|\mathbf{s},\theta)$

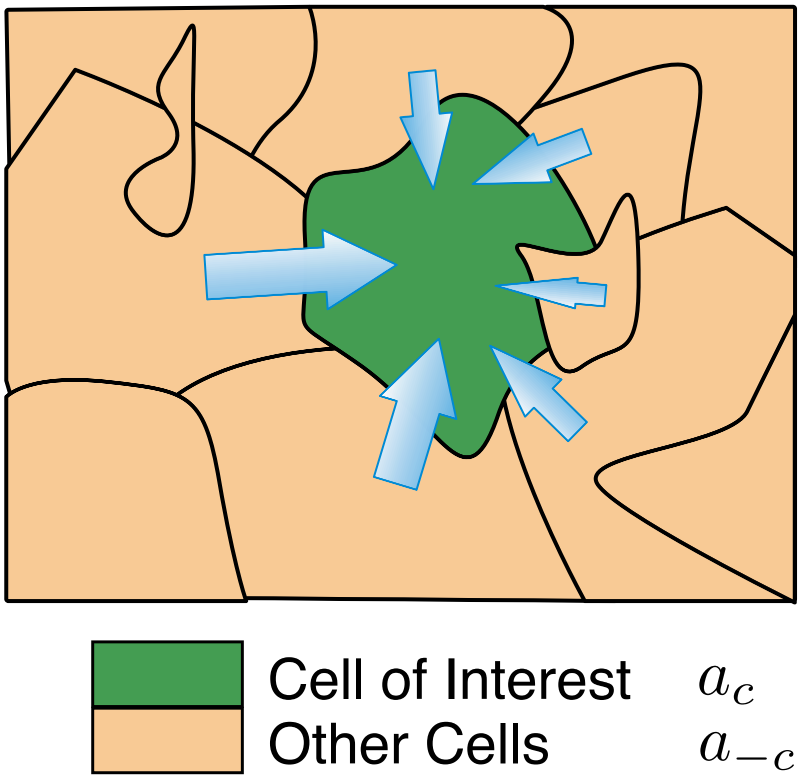
Modelling a Spatial Policy

1. **cell policy** - distribution over actions for a single cell given local state as well other relevant states and actions elsewhere

2. **landscape policy** - distribution over landscape actions given states of all cells

Cell Policy

$$\pi_c(a_c | a_{-c}, \mathbf{s}, \theta)$$



Cell Policy means:

Given the state of the forest, what action would an agent take at this location if the actions at all other locations were already decided ?

Cell Policy

Parameters are weights for each feature x action

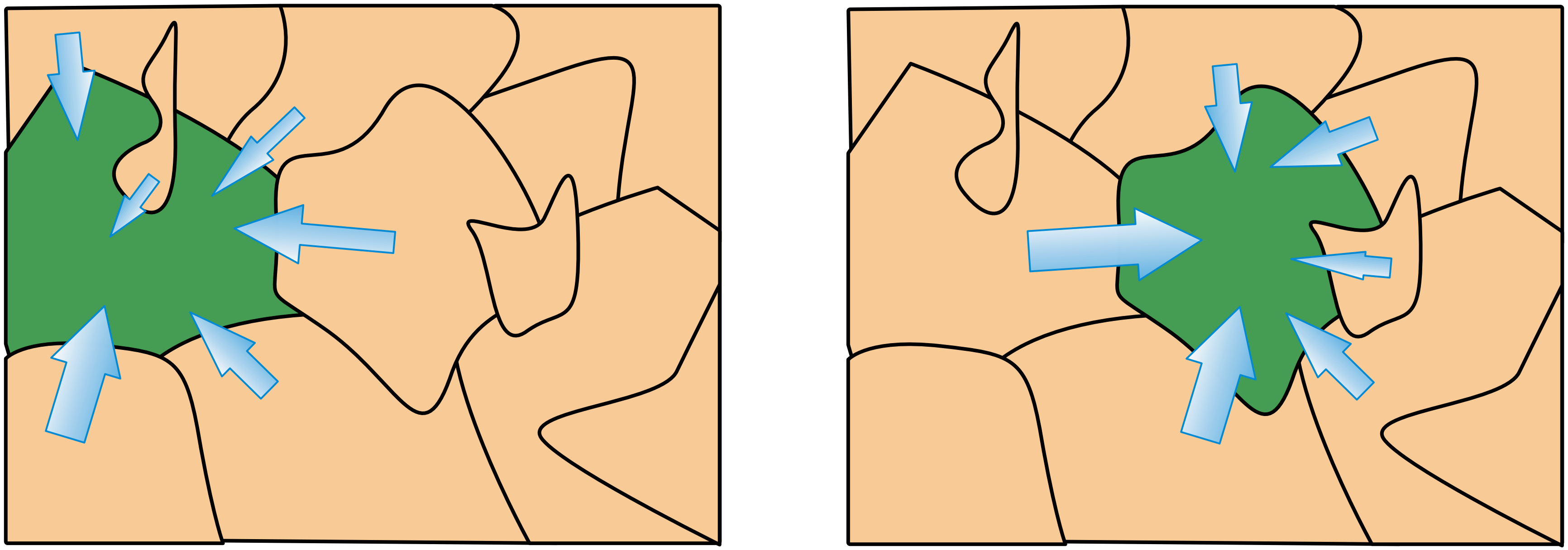
$$\theta_f(a) : \mathcal{F} \times A \to \Re$$

Features

Actions

$\theta_f(a)$	**Age**	**Largest Neigh**	**Any Adj Cut**	**Volume**
cut	2.0	-1.98	0.71	0.29
no cut	-1.6	6.97	3.85	4.79

Spatial Landscape Policy



Question: How can we define a consistent Landscape policy using local cell policies?

Equilibrium Landscape Policy

- A landscape policy $\Pi(\mathbf{a}|\mathbf{s},\theta)$ can be modelled as the **equilibrium** of a Markov chain defined by interacting local distributions $\pi_c(a_c|a_{-c},\mathbf{s},\theta)$
- **sampling**: acquire a landscape action given the current landscape state
- **gradient**: how does a change in one parameter affect the probability of a given landscape action

Sampling the Equilibrium Landscape Policy

construct a Markov chain with a node for each action over many sample steps

$$\pi_c(a_c | a_{-c}, \mathbf{s}, \theta)$$

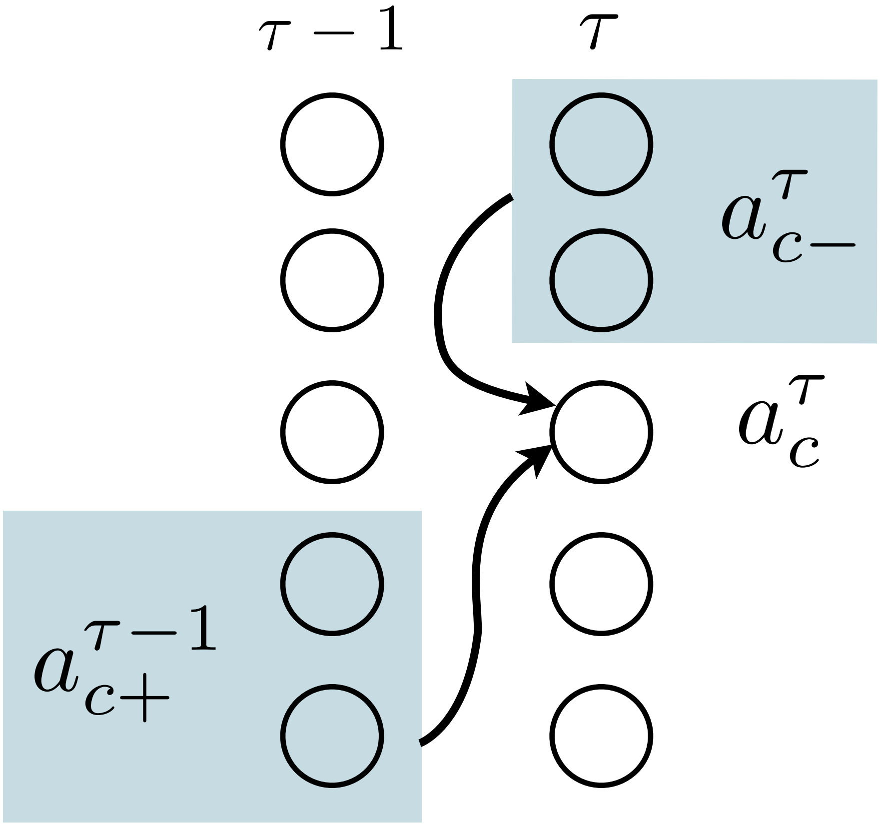
becomes ↓

$$\pi_c(a_c | a^{\tau}_{c-} \cup a^{\tau-1}_{c+}, \mathbf{s}, \theta)$$

↓

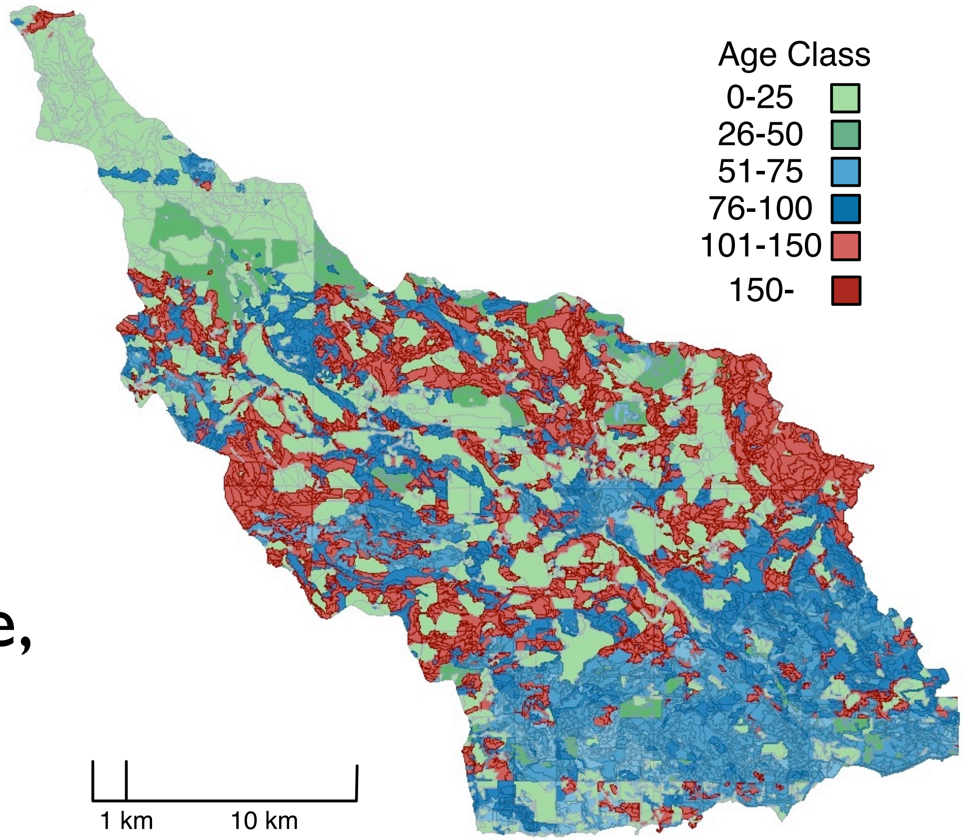
use Gibbs sampling to estimate landscape policy

$$\Pi(\mathbf{a} | \mathbf{s}, \theta)$$

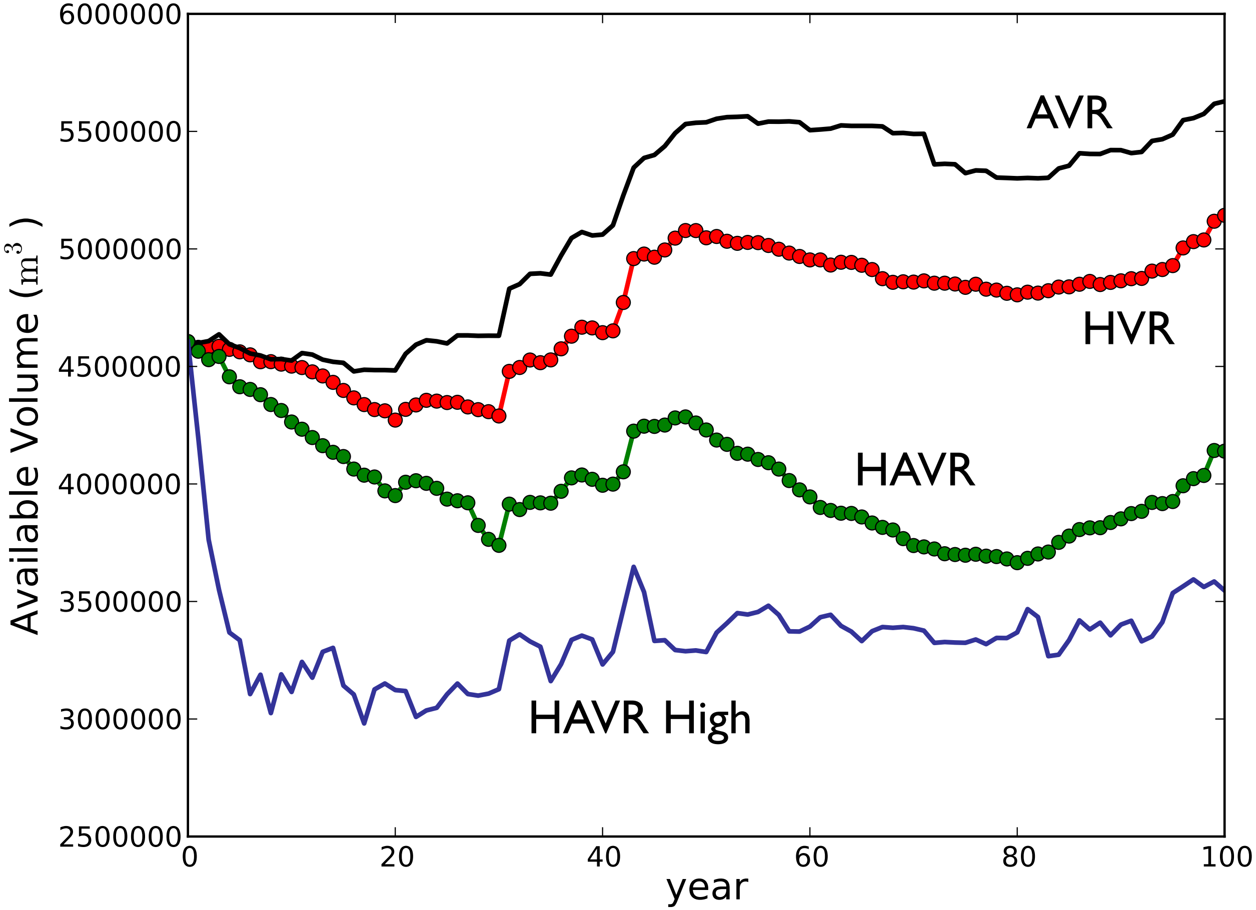


Experiment

- dynamics - **FSSAM**
  (BC Forest Service Spatial Analysis Model)
- 1880 cells
- actions : cut, nocut
- features : Age, Volume Largest Neighbour Volume, Any Adjacent Cut
- 100 year planning horizon
- 10 policy updates
- 500 MCMC steps after burn-in



Total Forest Volume Comparisons



Reward Models	
HVR	Penalize Irregular Harvest
AVR	Penalize Irregular Available Volume
HAVR	Penalize Both
All models reward total harvest volume and penalize adjacency violation	

Final Policy Parameters

$\theta_f(a)$ Parameters - Initial Values

Action	Age	Max AV	AnyAdj	Volume
`Cut`	1.0	1.0	1.0	1.0
`NoCut`	5.0	5.0	5.0	5.0

$\theta_f(a)$ Parameters - `HAVR`

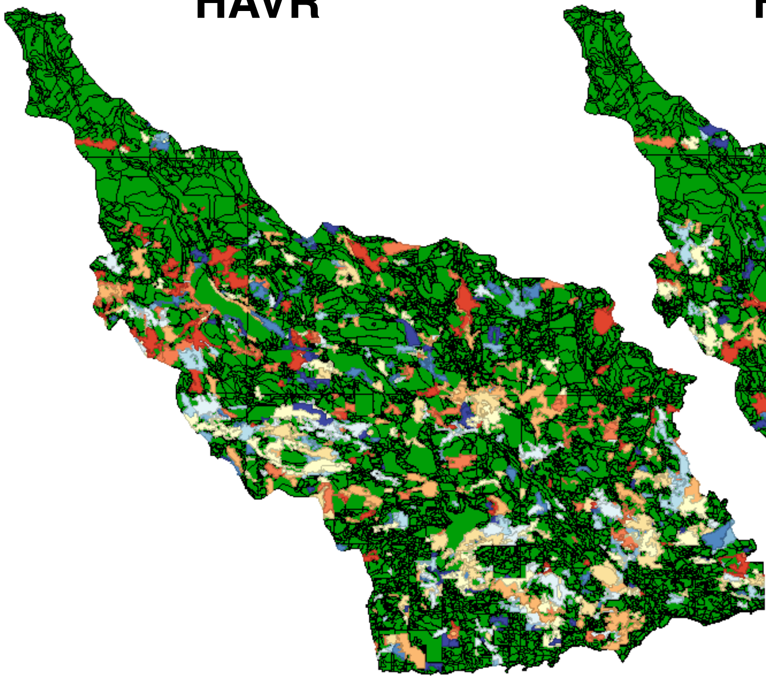
Action	Age	Max AV	AnyAdj	Volume
`Cut`	-1.55	-1.98	0.71	0.29
`NoCut`	7.82	6.97	3.85	4.79

$\theta_f(a)$ Parameters - `HAVR-High`

Action	Age	Max AV	AnyAdj	Volume
`Cut`	-11.88	-3.78	0.63	5.49
`NoCut`	15.27	9.05	4.16	-0.40

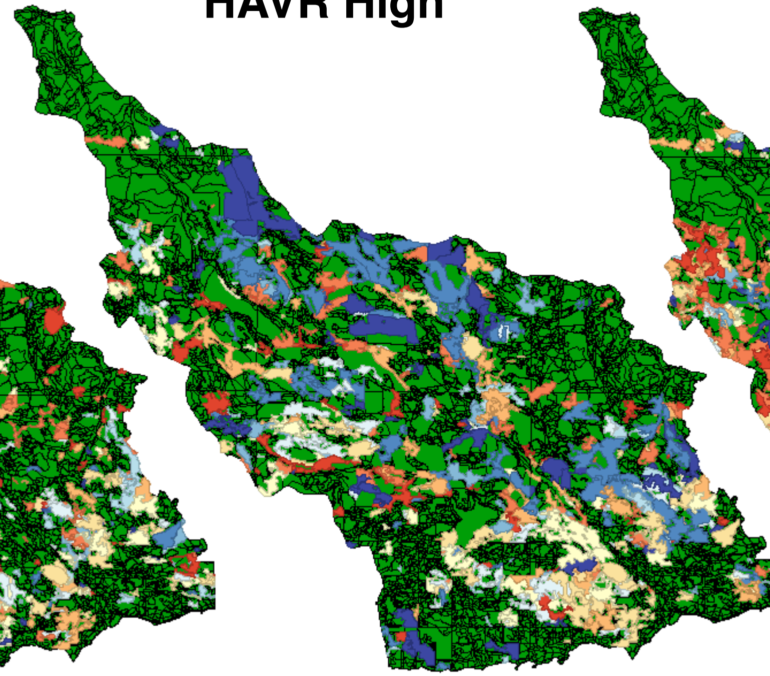
Three Example Plans

HAVR



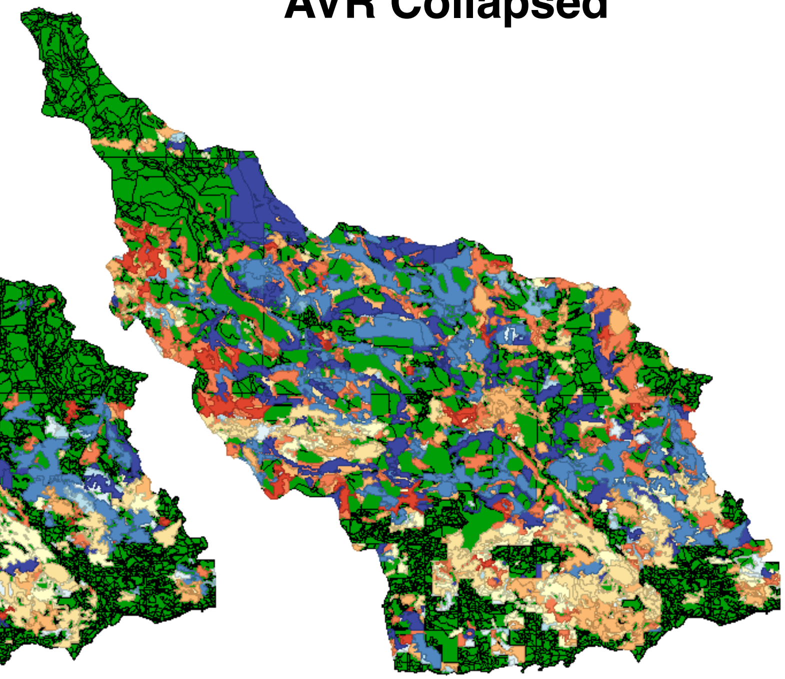
Typical results from HAVR reward model. Sustainable, low cut plan.

HAVR High



Common local minima from another run of HAVR. More aggressive plan, still sustainable over 100 years.

AVR Collapsed



Unsustainable plan coming from an AVR run. Forest population collapses completely.

Decade in which cell was harvested

0-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91-100

Conclusion

- Spatiotemporal planning can be usefully represented as a factored MDP
- Equilibrium Landscape Policy:
  - parametrized in terms of local states and actions
  - provides a consistent distribution over landscape actions
- Policy Gradients algorithm can be used to learn an abstract policy that applies at all locations.

Current Work

- Forest Fire
  - Let burn vs suppress fires
  - spatial policy for treatment/suppression actions using local landscape features
  - Clair Montgomery, Rachel Houtman, Aaron Gagnon
- Invasive Species on River Networks
  - compact policy representation of optimal solution
  - H.J. Albers, Kim Hall

Future Work

- Higher order spatial structures:
  - connected corridors for migration, firebreaks, viewsheds
- Visualization of spatial policies
- Applying to other domains:
  - urban planning, forest fires, invasive species management

Thank You