Using a Markov Decision Process to Inform Prairie Management Decision Making

Christopher Craig
christopher.craig@colorado.edu
CU Boulder

ACM Reference Format:
Christopher Craig. 2021. Using a Markov Decision Process to Inform Prairie Management Decision Making. In *Spring 2021, Boulder, CO.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 BACKGROUND

1.1 The Grassland Monitoring Team

Government employed grassland managers in the United States are individuals or groups who are in charge of the management and upkeep of government protected grasslands. Until recently, grassland managers have had little to no evaluation of their management's effects, and as a result there exists considerable uncertainty about the most appropriate management tools and prescriptions to use for their grasslands [3].

The Grassland Monitoring team is a multi-agency group of grassland managers and scientists comprised of a partnership between the US Fish and Wildlife Service, The National Conservancy, and The Minnesota Department of Natural Resources. Their goal is ultimately to improve the management of native prairies using standardized protocols and pooling data in an adaptive management framework. To help achieve this goal, over 13,000 acres of native prairie across 152 management units in Minnesota, North Dakota, and South Dakota have been collectively surveyed since 2008 [3].

The following are shared goals that are desired across all entities of grassland management working with the Grassland Monitoring Team:

(1) Maintain or increase the percent cover of native prairie vegetation.
(2) Maintain the floristic diversity of native grassland ecosystems.
(3) Minimize the percent cover of invasive/exotic vegetation.
(4) Maintain the structural diversity of native grassland ecosystems.

[3]

1.2 LISA

LISA (Laboratory for Interdisciplinary Statistical Analysis) is a statistical collaboration laboratory with a goal to increase the quantity and quality of statistics and data science applied to advance high-impact research at CU Boulder and expert decision making in the community [10].

This project was a collaboration request by Aaron Schwartz, a post doctorate student in Ecology at CU Boulder. Aaron is currently doing research with Grassland Monitoring Team, and came to LISA for help in building an adaptive management framework to aid decision making in the management of tall-grass prairie ecosystems. Ultimately, the objective of this project is to use the tall-grass prairie monitoring data to begin the development of a state transition model to provide decision support to grassland managers.

2 DATA

The data for this project came from over 13 years of monitoring prairie lands across the northern US. The data for these prairies was collected via observational systematic sampling using plots and transects.

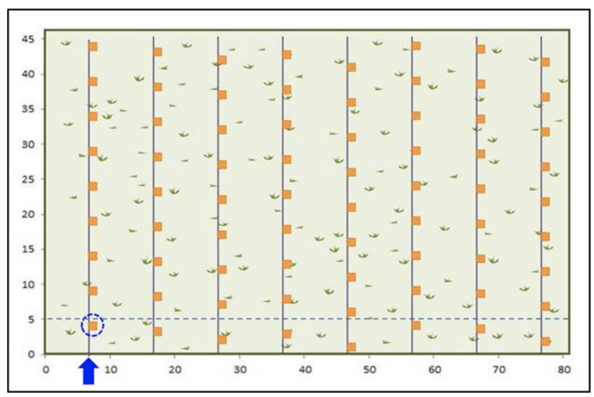

Figure 1: Example of an area being sampled via plots and transects

[9]

Shown in figure 1, a prairie is split up into parallel lines called transects, and plots are areas of land that are sampled along each transect. The Grassland Monitoring Team measure several variables within each plot which are then generalized to the transect level. For this project, there are 3 main variables of interest within the data set. First, native cover measures how much of each plot is covered by plant-life that is considered native to that prairie. Each plot is then assigned to one of four categories, <25% native cover, 25-50% native cover, 50%-75% native cover, and >75% native cover. To generalize to the transect level, we used the mode of native cover for the plots within each transect.

APPM, Spring 2021, Boulder, CO

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

Next, the plant community variable assigns each plot as mostly herbaceous or mostly shrubbery. Shrubbery can be thought of as plants that are woody, such as trees and bushes. Both native and non-native woody species can spread rapidly in prairie, making it difficult to maintain the open grassy habitat that most prairie species depend upon [4]. Thus, limiting the amount of shrubbery within a prairie is very desirable for a grassland manager. Again, the mode was used to generalize plant community from the plot to the transect level.

Third, the team measures native indicators, which counts the numbers of 'Tier 1' native species that appear across the entire transect. Indicator species are plants that, by their presence, abundance, or chemical composition, demonstrate some distinctive aspect of the character or quality of an environment [11]. Each transect is assigned to one of three categories of proportions, <0.1, 0.1 - 0.2, >0.2. Native cover, plant community, and native indicators are all variables that are designed to help quantify the shared goals that grassland management working with the Grassland Monitoring Team have.

Finally, around every year the grassland managers decides whether to leave the prairies as they are, graze them with cattle, or burn them via controlled burning. Grazing animals play an important role in maintaining the prairie ecosystem by stimulating plants to grow. This triggers biological activity and nutrient exchanges [8]. Also, over the course of a year a prairie can accumulate a lot of dead plant mass. By controlled burning, these dead plants burn which makes room for new plants to grow. Additionally, fire helps prairies to grow by stimulating grass and wildflowers to reproduce, reducing competition from weeds, and discouraging the encroachment of shrubs and trees [11].

Unfortunately the data for this project is still being accumulated and we still don't have the decisions that were made for each transect in between measurements. For the analysis in this paper we simulated decisions from a uniform distribution with each decision having an equal probability of being assigned to a transect. These decisions are read by the model as the decision that was made after the transect was surveyed.

3 DOMAIN PROBLEM

The Grassland Monitoring Team recently used the data described above to construct of a decision model to support how grassland managers will treat the prairies in between surveys. Using the three variables described above; native cover, plant community, and native indicators, they created a state space with twenty states to characterize the prairies for use in a decision model. Figure 2 shows how these state spaces were determined from the data. State 1 is considered to be the worst state for a transect to be in, while each successive stage is considered to be a better state for a transect with state 20 being the best.

After the defining their state space they created a state transition model. However, their model was like a black box in that they didn't quite understand how their model worked. Due to this issue, they didn't trust their model results which rendered the overall model uninformative.

We determined that one of the main issues with their model was that their state space was much too large which causes a few issues.

		Proportion of Native Indicators		
% Native Cover	Plant Community	>0.2	0.1 - 0.2	<0.1
>75	Herbaceous	20	19	18
>75	Shrubbery	17	16	15
50-75	Herbaceous	14	13	12
50-75	Shrubbery	11	10	9
25-50	Herbaceous		8	7
25-50	Shrubbery		6	5
<25	Herbaceous		4	3
<25	Shrubbery		2	1

Figure 2: Original state space chosen by Grassland Management Team employees

First, some of the states that are close to each other in number may be very difficult to distinguish between in the field. For example, states 1 and 2 only differ in the proportions of native indicator plants, which in reality could end up only being the difference of a handful of plants in a few plots. Also, the matrix of observed transition rates from one state to another was somewhat sparse, with many zeros and many very small transition rates. This could have lead to some of their model interpretation issues and is likely an indicator that their state space was too large for their data set. Figure 3 shows the matrix of observed transition frequencies in this original state space.

The specific goals of this project are to simplify their state-space and use their 13 years of prairie monitoring data to begin the creation a state transition model to help inform management decisions moving forward.

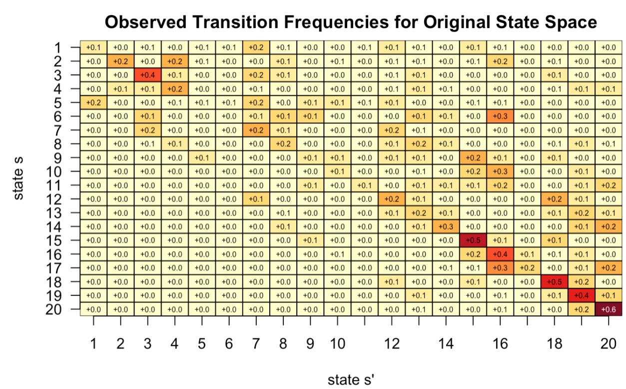

Figure 3: Matrix of observed transition frequencies from the original state space

4 METHODS/RESULTS

4.1 Choosing a New State Space

Our first task was to create a simplified state space that would make the probability transition matrix less sparse and make the final model more easily interpreted. There is not much literature about defining state spaces for state transition models, so we propose 2 alternative state-spaces using a similar methodology that the Grassland Monitoring Team initially used, as well as a set of state spaces that keep each variable separate. The first simplified state-space shrinks the number of states to 12 by combining some of the states with relatively low numbers of observations. The second simplified state space further shrinks the number of states to 8 by trying to remove all instances of where the number of observations of transitioning from one state to another is 0. Our final solution to this problem is keep each variable separate, meaning each variable has its own state space defined by its own categories. For example, the plant community state space has two states, one for the herbaceous category and another for shrubbery. We think that this makes the most sense, at least to begin with, since the meaning of transitioning from one state to another is very interpretable. Figure 4 shows how the two alternative state-spaces were chosen based on the collected data, and figures 5 and 6 show the matrices of observed transition frequencies for these state spaces.

1st Proposed State Space				
		Proportion of Native Indicators		
% Native Cover	Plant Community	>0.2	0.1 - 0.2	<0.1
>75	Herbaceous	12	11	10
>75	Shrubbery		10	9
50-75	Herbaceous		8	7
50-75	Shrubbery			6
25-50	Herbaceous			5
25-50	Shrubbery		4	3
<25	Herbaceous			2
<25	Shrubbery			1

2nd Proposed State Space				
		Proportion of Native Indicators		
% Native Cover	Plant Community	>0.2	0.1 - 0.2	<0.1
>75	Herbaceous	8		7
>75	Shrubbery	6		5
50-75	Herbaceous			4
50-75	Shrubbery			3
25-50	Herbaceous			
25-50	Shrubbery			2
<25	Herbaceous			
<25	Shrubbery			1

Figure 4: The two proposed state spaces that use the same methodology as the Grassland Management Team's original state space

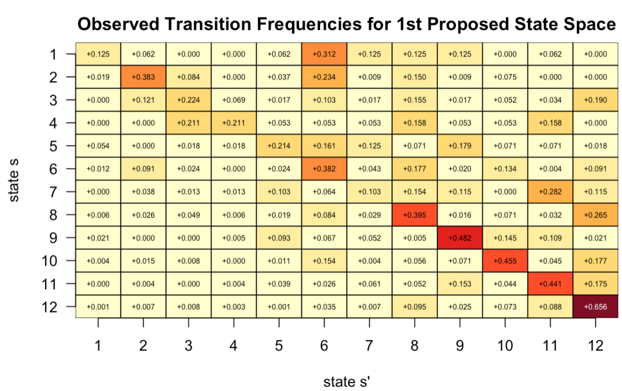

Figure 5: Matrix of observed transition frequencies for the first proposed state space with 12 states

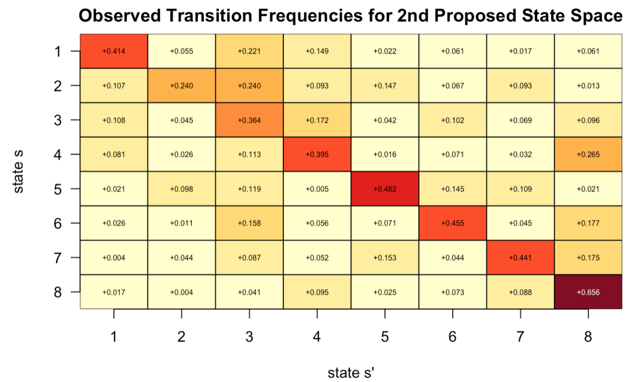

Figure 6: Matrix of observed transition frequencies for the second proposed state space with 8 states. This state space gets rid of instances where transitioning from one state to another is zero

4.2 State Transition Model

The basic idea of a state transition model is that there is an agent that is in some state S_t at time t. This agent can take one of multiple actions A_t, after which the it will stay in its state with some probability or transition to another state with some probability. Then, at time $t+1$, the process repeats for state S_{t+1}. Figure 7 shows an example of a simple Markov Decision Process, where there are three states in green, and two actions which are in red. The probabilities are shown written above the black lines. To create our state transition model we decided to use a Markov Decision Process (MDP). A MDP is a mathematical framework used for modeling decision-making problems where the outcomes are partly random and partly controllable [1].

The use of a MDP relies on a few important assumptions. First, the actions in a MDP need to be completely controlled and the states must be completely observable. In our case, this first assumption definitely holds, however the second may not. Sampling prairies is expensive in terms of time, money, and manpower, so some of the transects in the data set were sampled with large periods of time in between them. It is possible that in between these measurements, the transect could have been in another, unmeasured state. Also, once the data for decisions made is aggregated, we expect that many of the decisions of leaving/burning/grazing the prairies were made on off years when the prairies were not measured. There are workarounds to this which will be discussed later, but for now we assume that both of these assumptions hold. Finally, the states in the system need to follow the Markov Property. The Markov property states that the conditional probability of being in any future state, such as the physical state of a prairie transect, depends only on the present state of the system [2]. This means that we are assuming that future states of a transect are only dependant on the current state and not past states. We agreed that this assumption fits the process of the prairie ecosystem.

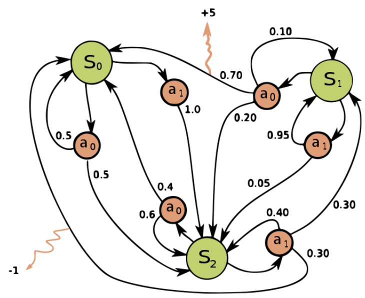

Figure 7: An example of a Markov Decision Process with states in green and actions in red

4.3 How Does a Markov Decision Process Work?

To solve a MDP we need 5 things:

(1) **P**, a probability transition matrix. We determined our probability transition matrix by non parametric bootstrap sampling the observed transitions from the data.
(2) **S**, a set of states. These are the state spaces for the prairie transects
(3) **A**, a set of actions. In our case these are Leave, Burn, and Graze.
(4) **R**, a reward function, which assigns some value given for reaching state $S_{t+1} = s'$ from state $S_t = s$
(5) γ, a discount for future awards. This is a chosen value between 0 and 1.

The goal of a MDP is to come up with a policy $\pi(a|s)$ which will determine the optimal action a to take at state s.Choosing the best possible policy depends upon finding the optimal state-value function and the optimal action value function. To define these functions, we first need to define the pieces.

When The Grassland Monitoring team takes an action on a prairie under a policy $\pi(a|s)$, the transition probability matrix **P** determines the subsequent state s'. The probability of transitioning from state s to s' given an action a is:

$$\mathbf{P}^a_{ss'} = P(S_{t+1} = s'|S_t = s, A_t = a)$$

After a transect transitions from state s to s' and a policy $\pi(a|s)$ was followed, it gets a reward based on the reward function **R** as feedback. Given that we are in state s and take action a, our expected reward is:

$$\mathbf{R}^a_s = E[R_{t+1}|S_t = s, A_t = a]$$

This is a short term reward that is received after transitioning from state s to state s' given the action a. Summing all future rewards and discounting them with the discount factor γ gives us our return **G**. γ is a chosen value that ranges from 0 to 1. The closer to 1 γ is, the more weight is put on future rewards. Choosing γ lets you balance between short and long term rewards.(6)

$$\mathbf{G}_t = R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + ...\gamma^k R_{t+k+1} = \Sigma_{i=0}^N \gamma^k R_{t+i+1}$$

With the return $\mathbf{G}_t$ defined, we can define the state-value and action-value functions. Following some policy π, the state value function $V_\pi(s)$ tells us how good it is to be in state s. In a similar vein, the action-value function $Q_\pi(s)$ tells how good it is to take that action.

$$V_\pi(s) = E_\pi[\mathbf{G}_t|S_t = s]$$

$$Q_\pi(s, a) = E_\pi[\mathbf{G}_t|S_t = s, A_t = a]$$

4.4 Solving a MDP

An optimal policy, π_*, would give us the optimal state-value and action-value equations which in turn maximizes the return $\mathbf{G}_t$.

$$\pi_* = \arg\max_\pi V_\pi(s) = \arg\max_\pi Q_\pi(s, a)$$

To calculate π_*, we rely on the Bellman Equation, Bellman Expectation Equation, and the Bellman Optimality Equation. The Bellman equation breaks the state-value and action-value functions into their immediate reward and the future value function [7].

$$\begin{aligned} V_\pi(s) &= E_\pi[\mathbf{G}_t|S_t = s] \\ &= E_\pi[R_{t+1} + \gamma(R_{t+2} + \gamma R_{t+3} + ...)|S_t = s] \\ &= E_\pi[R_{t+1} + \gamma \mathbf{G}_{t+1}|S_t = s] \\ &= E_\pi[R_{t+1} + \gamma V_\pi(s+1)|S_t = s] \end{aligned}$$

$$\begin{aligned} Q_\pi(s, a) &= E_\pi[\mathbf{G}_t|S_t = s, A_t = a] \\ &= E_\pi[R_{t+1} + \gamma(R_{t+2} + \gamma R_{t+3} + ...)|S_t = s, A_t = a] \\ &= E_\pi[R_{t+1} + \gamma \mathbf{G}_{t+1}|S_t = s, A_t = a] \\ &= E_\pi[R_{t+1} + \gamma V_\pi(s+1, a+1)|S_t = s, A_t = a] \end{aligned}$$

The Bellman Expectation Equation is another way to view the state and action value equations. The state value equation can be formulated by summing the policy determining actions $\pi(a|s)$ multiplied by the actions respective action value $Q_\pi(s, a)$.

$$V_\pi(s) = \Sigma_{a \in A} \pi(a|s) Q_\pi(s, a)$$

Similarly, the action value function can be formulates by taking the reward and summing the transition probability of landing in each state $\mathbf{P}^a_{ss'}$ multiplied by the value of that state $V_\pi(s')$.

$$Q_\pi(s, a) = \mathbf{R}^a_s + \gamma \Sigma_{s' \in S} \mathbf{P}^a_{ss'} V_\pi(s')$$

Now we can substitute the action value function into the state value function, and vice versa

$$V_\pi(s) = \Sigma_{a \in A} \pi(a|s)(\mathbf{R}^a_s + \gamma \Sigma_{s' \in S} \mathbf{P}^a_{ss'} V_\pi(s'))$$

$$Q_\pi(s, a) = \mathbf{R}^a_s + \gamma \Sigma_{s' \in S} \mathbf{P}^a_{ss'} (\Sigma_{a' \in A} \pi(a'|s') Q_\pi(s', a'))$$

With these defined, we can now define the optimal state and action value functions.

$$\arg\max_\pi V_\pi(s) = \max_{a \in A}(\mathbf{R}^a_s + \gamma \Sigma_{s' \in S} \mathbf{P}^a_{ss'} V_\pi(s')))$$

$$\arg\max_\pi Q_\pi(s, a) = \mathbf{R}^a_s + \gamma \Sigma_{s' \in S} \mathbf{P}^a_{ss'} \max_{a' \in A}(Q_\pi(s', a'))$$

Under our assumptions, these optimality equations are what is used to solve the MDP. Now using one of several possible algorithms that utilize dynamic programming, we can obtain the optimal policy [7].

4.5 Implementing a MDP

While the main idea behind solving a Markov Decision Process may seem complex, the implementation of a MDP is relatively simple. The first step in solving the MDP is coming up with the probability transition matrices split by action. This is where we obtain the values for the probabilities defined above:

$$\mathbf{P}^a_{ss'} = P(S_{t+1} = s' | S_t = s, A_t = a)$$

This means that given a state space, we end up with the same number of probability transition matrices as actions. In our data set, we have three actions and therefore three probability transition matrices. One of the difficulties of solving a MDP is determining these probability transition matrices. This is due to the fact that once observations are split up into states and then further split into actions, there can be relatively few observations for certain changes of states. Bootstrap sampling mitigates this issue by using random sampling with replacement to estimate the sampling distribution of our transitions [6]. To implement this in R, we created a function that once given a state space creates three matrices of observed transition frequencies. These matrices are then non parametrically bootstrap sampled to obtain the probability transition matrices.

The next step in solving a MDP is to come up with the rewards. We proposed two different reward functions, however note that there are many more potential reward functions that could be suitable to use. The idea behind coming up with a reward is try to quantify how good or bad it is to transition from state s to state s'. Remember that in our state spaces, lower numbered states are worse states for a transect to be in, while higher numbered state spaces are better. This means that transitioning from a state to a higher numbered state is good, and transitioning to a lower numbered state is bad. Our first proposed reward function gives a reward of +1 if the state transitions to a higher numbered state, and gives a reward of -1 if the state transitions to a lower numbered state. If a state stays the same the reward is 0. Figure 8 shows a matrix of this reward function. Our second proposed reward function is similar, with the reward being proportional to the difference between the numbers of states. Figure 9 shows a matrix of this reward function.

Choosing a discount factor γ is the next step in implementing a MDP. γ weighs how much the immediate rewards are valued as opposed to later rewards for the return function $\mathbf{G}$. Since we want the conditions for a transect to improve and then continue to improve, we would choose a γ value of close to 1. More about this choice is discussed in the model results.

The final step is to decide upon an algorithm to find the optimal policy. The R package 'MDPToolbox' proposes functions related to the resolution of discrete-time Markov Decision Processes [**mdptoolbox**]. This package offers several algorithms that can be used to solve an MDP for an optimal policy. These algorithms include finite-horizon, value iteration, policy iteration, and linear programming. For this project the value iteration, policy iteration, and linear programming algorithms are appropriate to use.

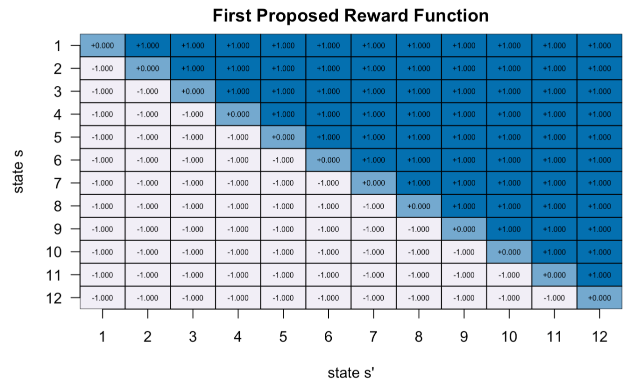

Figure 8: The first proposed reward function shown on the transition matrix of the first proposed state space

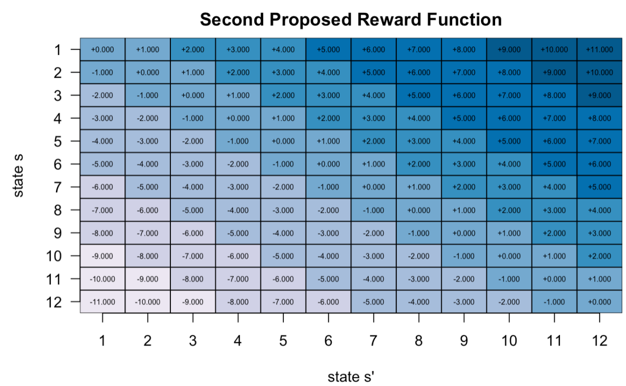

Figure 9: The second proposed reward function shown on the transition matrix of the first proposed state space

4.6 Model Results

Since we don't have all the data for the actions that were taken in between the prairie surveys, there isn't much we can show in terms of results on which actions should be taken given that a transect is in a certain state. We can, however, show the optimal policies given with our randomly generated actions. Figures 10 - 13 show how a policy can change when using different algorithms and discount factors. For these figures, we used three different algorithms with 11 different discount factors each ranging from .5 to 1 with increments of size .05. This results in 33 different optimal policy outputs. These policies were run and grouped by state for our first proposed state space and the three state spaces that are made out of our separated categorical variables.

Since our actions were randomly sampled, we can't make any inferences about the MDP and the actions we should take given that our transect is in a certain state. The aggregated policies for these plots vary much more than we would expect in reality using real data. These plots do, however, show that carefully selecting an algorithm and a discount factor to solve a MDP is very important, and that the MDP may not give a clear cut answer to which decision should be made for a prairie given that its in a certain state. This

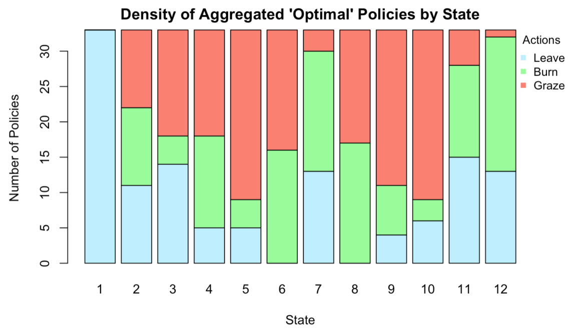

Figure 10: Density of 33 optimal policies by state for the 1st proposed state space

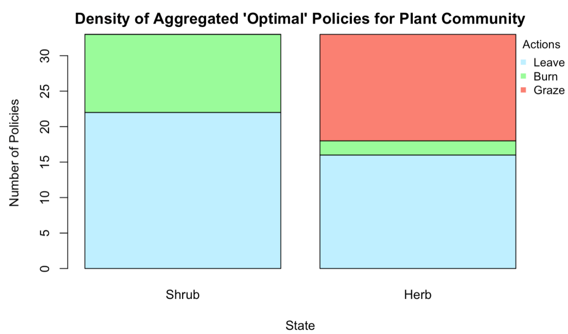

Figure 11: Density of 33 optimal policies for the Plant Community variable

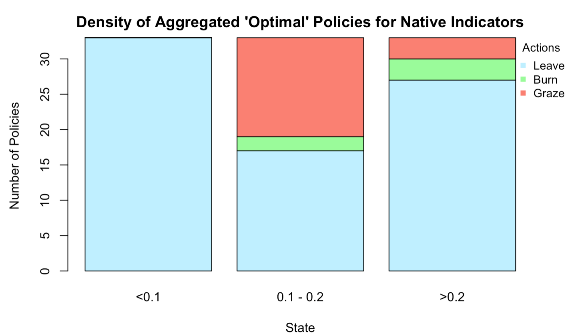

Figure 12: Density of 33 optimal policies for the Native Indicator variable

illustrates why this model will only be a piece of what will drive grassland managers to make a decision on how to deal with a prairie.

5 FUTURE WORK/CONCLUSION

There is still a lot of work to be done with this decision model. We believe that this basic MDP model that we have created is a good jumping off point for the Grassland Management Team to go off of.

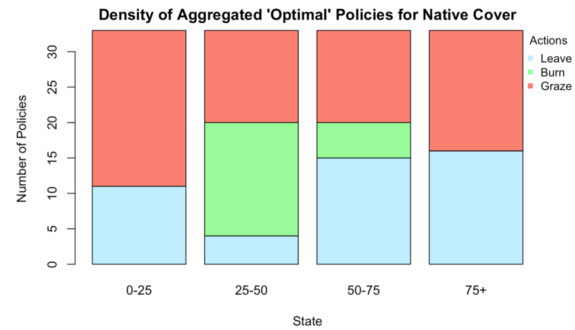

Figure 13: Density of 33 optimal policies for the Native Cover variable

Some of the decisions that were made for this model could be improved upon with a more intimate knowledge of the prairie ecosystems. For example, someone who knows more about the process of how a prairie transitions from one state to another could come up with more appropriate reward functions. Under our first proposed state space, both of our proposed reward functions value transitioning from state 1 to state 3 the same as transitioning from state 10 to state 12. In reality, it is likely that these two transitions would have different values to a grassland manager. Also, since the reward functions are separated by actions, something like the financial cost of grazing/burning an area could be incorporated into the rewards.

The next step in creating this decision model would be to turn the Markov Decision Process into a partially observed Markov Decision Process (POMDP). When a grassland manager makes a decision about how to treat a transect, they know what the state of a transect was the last time it was measured, but they have to infer what the state is currently. A POMDP takes the discrete time process we described above and turns it into a continuous time process. This relaxes the assumption of full observability that we made in section 4.2. The main difference between a MDP and POMDP is that since the state is unknown at the time that a decision is being made, the model uses observations to create belief states. A belief state is the belief of what state the system is currently in, which is represented by a probability distribution over the states. For a POMDP the optimal policy prescribes which action is optimal for each belief state [5].

Finally, another important step that needs to be taken is to generalize the process from the transect level to the management level. When a grassland manager decides to burn or graze some land, they do it to an area of land that is equivalent to multiple transects. So, when deciding which actions to perform, they need to account for the fact that different transects within a management level may have different actions suggested by the optimal policy. The Grassland Management Team could decide to reconcile this issue within the transition model, or deal with it outside of the model.

While there is still work to be done, we hope that this project removes some of the ambiguity surrounding state transition models for the Grassland Monitoring Team, and eventually results in a model that is informative and helpful to grassland managers.

REFERENCES

[1] Schwartz A Feinberg F. “Handbook of Markov decision processes: methods and applications”. In: *Discounting Ecological Criteria in Uneven-Aged Management* (2001), p. 576.

[2] GERALD J. LIEBERMAN FREDERICK S. HILLIER. *Introduction to Operations Research*. Boston: Mcgraw-Hill, 2005, pp. 1056–1062.

[3] Conservation Gateway. *Grassland Monitoring Team*. URL: https://www.conservationgateway.org/ConservationByGeography/NorthAmerica/UnitedStates/minnesota/Pages/grasslandmonitoringteam.aspx. (accessed: 03.09.2021).

[4] Chris Helzer. *Konza Prairie Trip Part 2 – Tree and Shrub Encroachment*. 2015. URL: https://prairieecologist.com/tag/shrubs/. (accessed: 03.09.2021).

[5] Vikram Krishnamurthy. *Partially Observed Markov Decision Processes: From Filtering to Controlled Sensing*. Cambridge, MA: Cambridge University Press, 2016.

[6] Jingjing Liang Mo Zhou Joseph Buongiorno. “Bootstrap Simulation, Markov Decision Process Models, and Role of Discounting in the Valuation of Ecological Criteria in Uneven-Aged Forest Management”. In: *Discounting Ecological Criteria in Uneven-Aged Management* (2010), pp. 243–271.

[7] Andrew G. Barto Richard S. Sutton. *Reinforcement Learning:An Introduction*. Cambridge, MA: The MIT Press, 2015, pp. 54–102.

[8] National Park Service. *Fire and Grazing in the Prairie*. 2020. URL: https://www.nps.gov/tapr/learn/nature/fire-and-grazing-in-the-prairie.htm. (accessed: 04.19.2021).

[9] Landscape Toolbox. *Vegetation Management and Monitoring*. URL: https://learn.landscapetoolbox.org/course/vegetation-measurement-and-monitoring/lessons/3-2-positioning-sampling-units-across-landscapes/. (accessed: 04.21.2021).

[10] Boulder University of Colorado. *Laboratory for Interdisciplinary Statistical Analysis*. URL: https://www.colorado.edu/lab/lisa/. (accessed: 05.01.2021).

[11] Private Land Partnerships Working Together for Wildlife. *INTRODUCTION TO GRASSLAND MANAGEMENT*. URL: https://www2.dnr.state.mi.us/PUBLICATIONS/PDFS/HUNTINGWILDLIFEHABITAT/LANDOWNERS_GUIDE/Resource_Dir/Acrobat/Grassland_Mgmt.pdf. (accessed: 04.19.2021).