1 Linear forest model

1.1 No storm (deterministic)

We consider a discrete time system with a linear dynamic of the form

$$s_{t+1} = As_t + Ba_t$$

where s_t is the state at time t and a_t is the action of an agent at step t.
Let n be the number of trees, a state at time t is the vector $s_t = (s_t^1, ..., s_t^n, H_t)^\intercal \in \mathbb{R}^{n+1}$. The action at time t is the vector $a_t = (a_t^1, ..., a_t^n)^\intercal \in \mathbb{R}^n$.
A is the transition matrix of the system. It depends on the graph of interaction between the trees G and two additional parameters α and β. α can be thought as a growth parameter which will influence the rate of growth of the trees, β is an interaction parameter that dictate how strong the interaction between the trees is (in this case the main interaction is a bonus or malus to the rate of growth based on the relative size of neighboring trees).

1.1.1 Dynamic of the model

We denote as $\mathcal{V}_i$ the set of vertices connected to i in G. The dynamic of the system without exterior action will be the following

$$\forall i \in [\![1, n]\!], s_{t+1}^i = s_t^i + \alpha(H_t - s_t^i) + \frac{\beta}{|\mathcal{V}_i|} \sum_{j \in \mathcal{V}_i} (s_t^i - s_t^j)$$
$$H_{t+1} = H_t$$

We can notice that with this definition, the last coefficient of state vector remains constant and equals to its initial value that we will denote as H. With this model, the growth is allowed by the term $\alpha(H_t - s_t^i)$ which encourages the growth of the tree up to an asymptotic value H and penalizes overgrowth. The last term $\frac{\beta}{|\mathcal{V}_i|} \sum_{j \in \mathcal{V}_i} (s_t^i - s_t^j)$ models the interaction between the trees (having high neighbors is detrimental to the growth because they partially hide the sum, absorb more nutrient from the ground ...).
We can now define the corresponding transition matrix A. For all $i \in [\![1, n]\!]$ and $j \in [\![1, n]\!]$,

$$A_{i,j} = \begin{cases} 1 - \alpha + \beta\, \mathbb{I}\{\mathcal{V}_i \neq \varnothing\} & \text{if } i = j \\ -\frac{\beta}{|\mathcal{V}_i|} & \text{if } j \in \mathcal{V}_i \\ 0 & \text{otherwise} \end{cases}$$

In addition $\forall j \in [\![A, n]\!], A_{j,n+1} = \alpha$, $A_{n+1,j} = 0$ and $A_{n+1,n+1} = 1$.

For instance, with two trees in interaction, we obtain

$$s_{t+1} = \begin{pmatrix} 1 - \alpha + \beta & -\beta & \alpha \\ -\beta & 1 - \alpha + \beta & \alpha \\ 0 & 0 & 1 \end{pmatrix} s_t + Ba_t$$

1.1.2 Actions

For each tree described by the model the agent can either let it grow or cut it and directly plant a new one. Multiple trees can be cut at once.

$$s_{t+1} = As_t + Ba_t$$

We define $B = -A \times \begin{pmatrix} I_n \\ 0...0 \end{pmatrix}$ and restrict the actions $a_t \in \{K_t s_t | K_t = \text{diag}(a_t^1, ..., a_t^n), (a_t^1, ..., a_t^n) \in \{0,1\}^n\}$

At step t and for each tree i the agent can either choose $a_t^i = 0$ to let the tree grow or $a_t^i = 1$ to harvest and replant the tree.

1.1.3 Rewards/cost

We consider a quadratic reward at time t of the form

$$r_t = s_t^\intercal M s_t + a_t^\intercal N a_t$$

Given the form of a_t defined previously, if N is diagonal, $a_t^\intercal N a_t$ amounts to get a reward quadratic in the size of the tree. The first term $s_t^\intercal M s_t$ although not used in the experiments, can allow to model the value attached to other ecosystem services that requires a grown forest such that carbon storage, biodiversity or other recreational uses.

1.2 Storm risks

At each time step, the forest experiences a storm with fixed probability p_{storm}. In case of a storm, each tree is removed independently with a probability that depends on the height of its neighbors (the higher the neighbors, the less likely to be destroyed). To model this, we add an extra term to the dynamic of the system.

$$\begin{aligned} s_{t+1} &= As_t + Ba_t - AL'_t(s_t - \begin{pmatrix} I_n \\ 0...0 \end{pmatrix} a_t) \\ &= A(I - \begin{pmatrix} I_n \\ 0...0 \end{pmatrix} K_t - L'_t + L'_t \begin{pmatrix} I_n \\ 0...0 \end{pmatrix} K_t) s_t \\ &= A(I - K'_t - L'_t + L'_t K'_t) s_t \\ &= A(I - L'_t)(I - K'_t) s_t \end{aligned}$$

where $K'_t = \begin{pmatrix} I_n \\ 0...0 \end{pmatrix} K_t$, $L'_t = \begin{pmatrix} I_n \\ 0...0 \end{pmatrix} L_t$ and

$$L_t = Y_t^{storm} \text{diag}(Z_t^1, ..., Z_t^n),$$

with $Y_{storm} \sim \mathcal{B}(p_{storm})$ and $Z_t^i \sim \mathcal{B}(p_t^i)$,

$$p_t^i = \exp\left(- \sum_{j \in \mathcal{V}_i} s_t(i)/(HD) \right)$$

D is a parameter of the model linked to the destructive power of the storm (it could be replaced by a random variable defining the power of a given storm, its distribution could even be changed through time). p_{storm} is also a parameter of the model, however we could imagine that it does not stay constant overtime to take into account changes in the environment (increase in storm risks due to global warming for example).

1.3 Risk measure

1.4 Extensions

A first interesting extension of the model is to allow multiple types/species of trees. One way to implement this is to explicitly extend the state by specifying the type of each tree from a predefined set $\mathcal{C}$ of predefined types (eg. oak, pine or the absence of tree if the agent does not want to replant). The state becomes the couple (s_t, c_t) where $c_t = (c_t^1, ..., c_t^n) \in \mathcal{C}^n$. This formulation already allows defining risks linked to the diversity of trees, such as decease or behavior in the face of forest fire. One may also want to also change the growth dynamic of the tree (α, β or H), which implies replacing the constant transition matrix A by a state dependent matrix $A(c_t)$.
A less expressive alternative for this diversification of trees is to simply extend the representation of the state s_t (without having the additional c_t) by associating a value for H to each individual tree: $s_t = (s_t^1, ..., s_t^n, H_t^1, ..., H_t^n$. The value of H_t^i can then be set at harvest time to change the dynamic of the new tree (both growth speed and final height. The type of tree would correspond exactly to a value of H.

2 Discrete forest model

We describe the forest as a Markov decision process (MDP) with state space $\mathcal{S} = [\![0, H]\!]^n$ describing the height of n distinct trees, where H is a parameter corresponding to the maximal height of the trees. The action space is $\mathcal{A} = \{\text{grow}, \text{harvest}\}^n$. The trees influence each other during their growth and their relations are represented by an interaction graph G. We denote the set of trees adjacent to the tree i in G as $\mathcal{V}_i$.

2.1 No storm

At each time step, each tree can either grow by one unit or keep the same height. The probability of growth at time t of the tree i is $p_t^g(i)$ define as

$$p_t^g(i) = clip\left(\frac{H - s_t^i}{H} \times \exp\left((\sum_{j \in \mathcal{V}_i} s_t^i - s_t^j)/n\right) \times \frac{1}{Z(s_t)}, [0, 1]\right)$$

#todo

2.2 Storm risk

#todo

3 Experiments with the linear forest model

3.1 Without storms

Unless specified, the experiments were conducted using a graph structure describing a forest with trees placed in a grid pattern. Each tree interacts with its 4 nearest neighbors in the grid. The objective was to maximize the total reward over 100 steps. Each plot display the average over 100 runs with different initial state. The parameters of the system are $\alpha = 0.2$ and $\beta = 0.1$.
The first policy evaluated is to set a cutting age T in advance for all the trees and harvest periodically according to this cutting age. The policy can either be performed by synchronizing the harvest (all the trees are cut down at the same time) or offsetting the harvest (each year about $\frac{n}{T}$ are cut down).

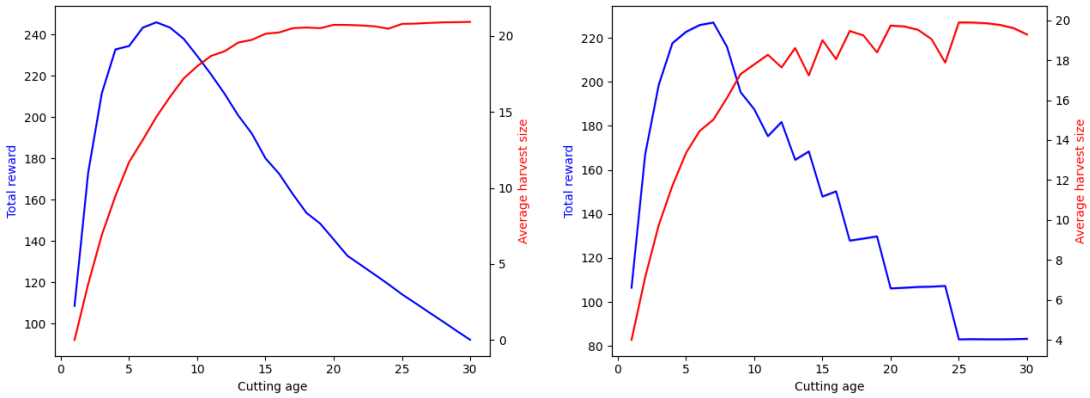

Figure 1: Periodic harvest without storms (Left: Periodic harvest with an offset, Right: Synchronized periodic harvest)

The second policy evaluated consists in setting a threshold on the height H_{cut} in advance and harvesting the trees as soon as they reach this height.

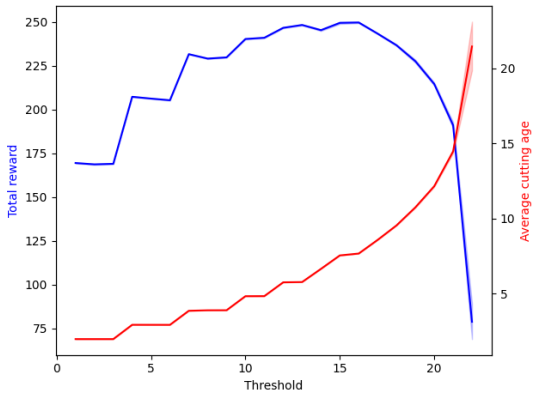

Figure 2: Harvest when the height reaches a certain threshold, without storms

We last experimented with a learning algorithm. In this case, PPO is used. We train the network for 100 steps then evaluate the total reward obtained by the learned policy over 100 steps.
We observe that the average total reward of the learned policy is around 260 while it was 250 for the optimal threshold policy, 245 for the optimal cutting age policy and 225 for the synchronized cutting age policy

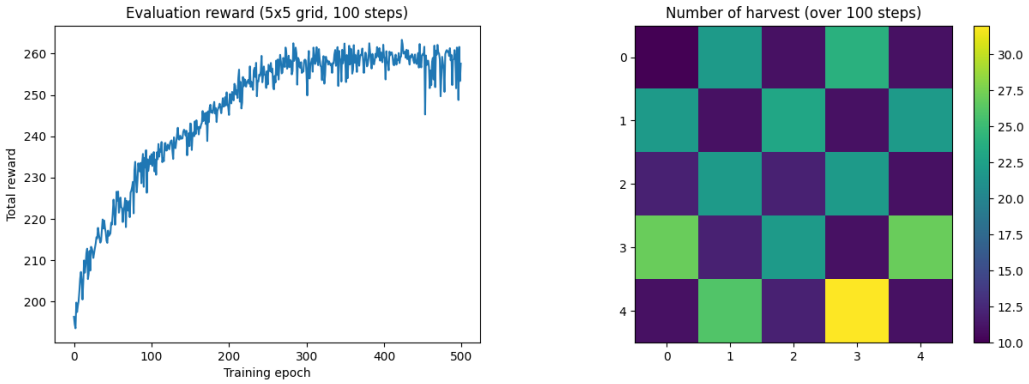

Figure 3: Learning a good policy using PPO for a forest without storms

We can notice a periodic pattern. Some trees are prevented from growing by cutting them frequently so they stay small, allowing other trees to grow faster and higher. We observe the same kind of learned behavior when the trees interact with their 8 nearest neighbors instead of 4.

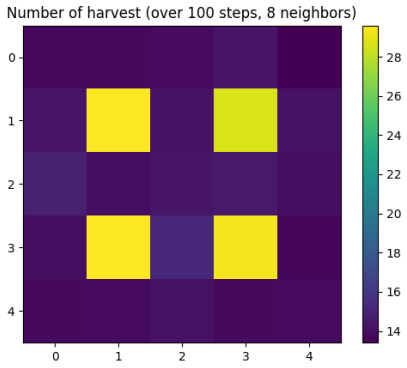

Figure 4: Number of harvest for a policy learned using PPO on a grid environment in which each tree influence its eight nearest neighbors

3.2 With storms

The same experiments were repeated for the model with storms. The parameters are $p_{storm} = 0.05$ and $D = 4$. Overall, the total reward is lower than in the deterministic setting due to the loss of some

trees during the storms. We observe that the optimal cutting age decreases and threshold marginally decrease. The frequency of harvest of each tree for the learned policy is hard to interpret, however it remains almost constant over multiple runs (so it does not depend strongly on the storms).

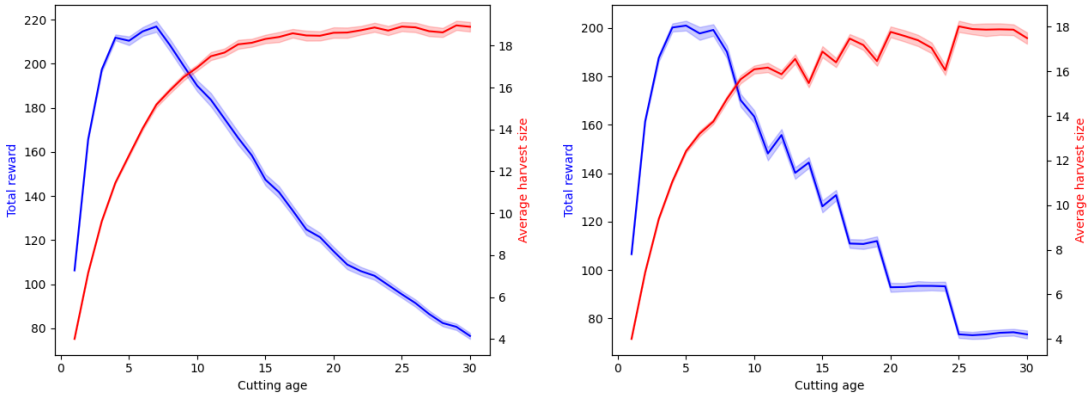

Figure 5: Left: Periodic harvest with an offset, Right: Synchronized periodic harvest

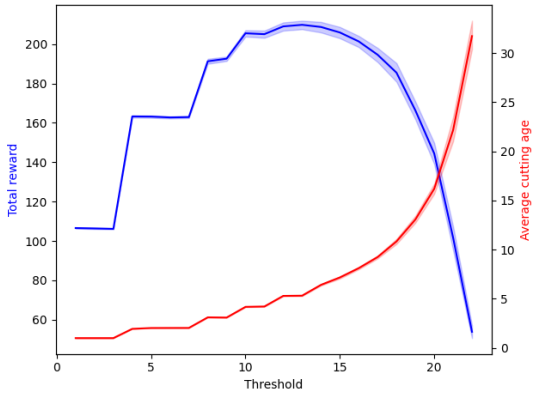

Figure 6: Harvest when the height reaches a certain threshold

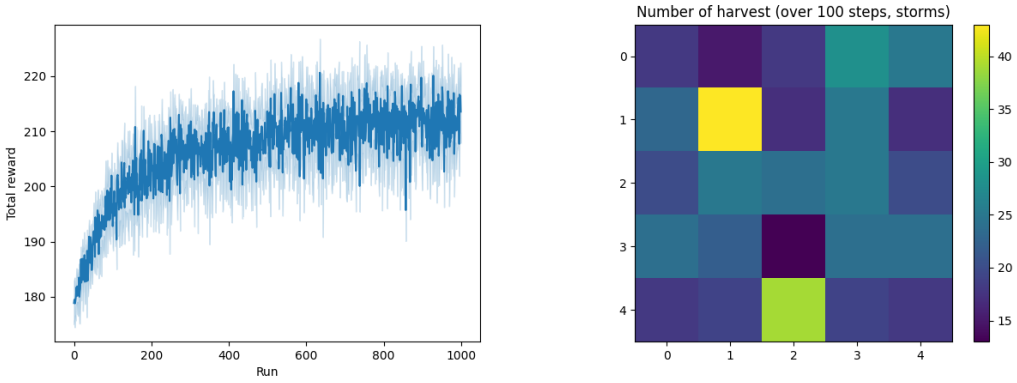

Figure 7: Learning a good policy using PPO