Problem Chosen	2022 MCM/ICM Summary Sheet	Team Control Number
E		2218360

DeepForest: A Reinforcement Learning Approach to Forest Management for Carbon Sequestration

Summary

Carbon dioxide in the atmosphere is one of our most pernicious greenhouse gases, contributing to global warming and all of its downstream impacts – sea level rise, desertification, and food and water scarcity, among others. Because forests are capable of capturing and storing significant quantities of carbon dioxide not only in living plant tissue, but also in the soil, proper management can mitigate a great deal of these harms. Forest managers face the nontrivial problem of having to decide between preserving trees and allowing them to grow and reproduce, or harvesting them and sequestering their carbon into a potentially more long-lived products.

We take a reinforcement learning approach to modeling forests with **DeepForest**. By simulating the ordinary differential equations that govern forest carbon pools and fluxes in an OpenAI Gym environment, and deploying a Deep Q neural network to learn the long-term returns of state-action pairs, DeepForest is capable of recommending a management strategy for any arbitrary forest in order to maximize carbon capture and other potential objectives.

With DeepForest, we find that in systems like temperate coniferous Scots pine forest, managers ought to harvest conservative numbers of mature trees at regular periods, allowing the trees to grow and recover in cycles. In systems with higher litter rates like temperate deciduous silver birch forests, managers may prefer to more aggressively cut trees down before they decompose. Conversely, tropical palm rainforests with excellent growth and slow soil decomposition ought to be left alone to absorb carbon.

Forest managers may not only wish to prioritize carbon sequestration. DeepForest can also take into account biodiversity, forest size, or the economic productivity of harvested products. It can accordingly harvest overly-abundant tree species, or time harvests for maximum profit.

Thus, we develop a flexible, interpretable machine learning approach to evaluating complex forest ecosystems and finding the optimal management strategies to maximize carbon sequestration.

Contents

1 Introduction

1.1 Problem Background

With increasing levels of atmospheric carbon dioxide (CO_2) threatening to exacerbate global warming and disrupt climate patterns, policymakers are looking to sequester carbon away in both natural and artificial reservoirs, thereby slowing the accumulation of our most significant long-lived greenhouse gas.

Forests play a vital role in the carbon cycle. The trees that make them up fix carbon from the air via photosynthesis, store them in their tissues, and contribute litterfall to soil- and water-bound pools of carbon. Hence a forest's capacity to store carbon can be evaluated from 1) the carbon fixed from net primary production (NPP); 2) the carbon stored in various ecosystem pools; and 3) long-term storage in forest products – that is, sequestering carbon in the form of furniture, lumber, plywood, paper, and other potentially long-lived wood products. Depending on the environmental conditions and forest management strategies employed, forests can be potent carbon sources or sinks.

Moreover, forests can be valued for their biodiversity, cultural import, recreational use, and economic productivity. Forest managers have the responsibility of weighing all of these factors and making the right choices to preserve, or harvest, the trees in their care.

1.2 Problem Restatement

As part of the International Carbon Management (ICM) Collaboration, we:

1. Develop a model to determine how much CO_2 a forest sequesters over time.
2. Determine the management plan which maximizes CO_2 sequestration.
3. Explore management plans which balance carbon sequestration with other value considerations, including biodiversity, forest size, and economic productivity.
4. Examine the CO_2 sequestration and optimal management plans for forests with different initial conditions, including age and type of trees, geography, topography, and benefits and lifespan of forest products.

1.3 Our Approach

A forest is a dynamically changing system where carbon enters, exits, and flows between distinct reservoirs, and every initial condition and action taken disturbs the system in different ways. Making the "right" decision for the long-term future at every time point is a non-trivial task. Given the infeasibility of a one-size-fits-all solution, we treat the ICM Collaboration as a *reinforcement learning (RL)* problem, that is, a machine learning problem concerned with how an intelligent agent – the forest manager – ought to take actions – preserving or harvesting trees – to maximize a cumulative reward – carbon sequestered, for example. By learning through trial and error, the agent can find a good solution when no definite answers may exist. We can treat the environment as a Markov decision process (MDP) because its next state depends only on the observed current state, known physical and chemical laws, and the manager's action. **Enter *DeepForest*.**

2 Notation

We first introduce all the variables and parameters used throughout the rest of the paper in order to assist the reader's understanding.

	Description
C	A vector of different carbon pools in the ecosystem (Mg)
C_S	The carbon pool representing soil carbon (Mg)
C_P	The carbon pool representing product carbon (Mg)
$C_{T,i}$	The carbon pool representing the amount of tree carbon for tree i (Mg)
T	A vector representing the various counts of trees in the ecosystem
$T_{O,i}$	The number of old trees of type i
$T_{Y,i}$	The number of young trees of type i
t	The time state in the simulation (years)
t_f	The final time in the simulation
m_i	The maturation rate of tree i
K_X	The death/decomposition rate of variable X
N	The carrying capacity (max number of trees) in the forest
α_i	The tree to carbon conversion rate

	Description
β_i	The policy to tree harvest scaling factor
$\nu_{O,i}$	The tree death to carbon loss scaling factor in old trees
$\nu_{Y,i}$	The tree death to carbon loss scaling factor in young trees
$\omega_{O,i}$	An old tree growth parameter
$\omega_{Y,i}$	A young tree growth parameter
π	A valid policy that the agent may use
π^*	The optimal policy solving the control problem
$\mathcal{P}$	The state space of valid policies
$R(C,T,\pi)$	The control problem's reward function
γ	The control problem's discount factor
θ	Network parameters

3 Assumptions

Ecosystems are highly complex, and assumptions must be made for the sake of simplification and interpretability.

1. **The atmosphere is an infinite source and sink of CO_2.** The total mass of carbon in the forest is negligible compared to the carbon in the atmosphere [7], [1].

2. **The forest's only relevant pools of carbon are the trees and soil.** Animal carbon is negligible, and microorganism and fungal carbon is generally captured by soil and root measurements, respectively [1].

3. **Carbon can only be fixed from the atmosphere and enter the forest via plant photosynthesis.** Nearly all the carbon entering the system is through the primary production of plants, since algal and bacterial photo-/chemo-synthesis is negligible.

4. **Carbon can only be released to the atmosphere and exit the forest via decomposition.** Carbon returns to the atmosphere as carbon dioxide largely via the cellular respiration of decomposers.

5. **The only relevant plants are trees.** Non-woody plants are negligible and cannot be converted to forest products.

6. **Trees of each species can be divided into age classes, within which there is uniform growth, death, and reproductive behavior.** See canonical ecosystem models such as [11].

7. **Trees follow logistic growth and exponential decay of tree count.** The growth of trees is limited by a carrying capacity that is determined by resource limitations and intra/interspecies competition. On the other hand, death and maturation of trees are subject to exponential decay.

8. **Litterfall rate follows exponential decay of tree carbon mass.** See [12].

9. **Decomposition, whether of soil carbon or forest product carbon, follows exponential decay of mass.** See [5].

10. **A manager can only choose inaction, or to harvest a portion of existing mature trees.** This is implied by the problem statement.

11. **Harvesting kills a tree, preventing further growth and reproduction.** This is implied by the problem statement, "forest managers must find a balance between [harvesting] and [...] allowing the forest to continue growing [...] as living trees."

12. **50% of harvested carbon is lost during manufacturing.** See [18].

13. **The death or harvesting of a tree from a particular species and age class entails the removal or transfer of, on average, a fixed mass of carbon.** For the sake of simplicity, or else we would need to sample from an ill-defined distribution. See [14].

14. **The effects of environmental variations are implicitly encoded in user-defined hyperparameters.** Nutrient limitation, intra/interspecies competition, geography, topography, and weather only matter to the extent that they affect carbon flux via growth, reproduction, maturation, litterfall, decay, and death rates, which are controlled by hyperparameters.

15. **The forest has a constant density of trees as it grows in size.** Reasonable per [16]. Necessary for unit conversion from measurements like megagrams C/hectare/year to megagrams C/tree/yr.

4 The Model

DeepForest follows an RL paradigm and consists of three components: the environment, the agent, and the optimizer.

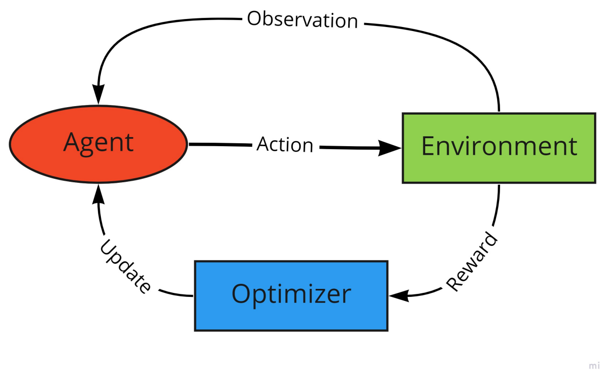

Figure 1: The RL paradigm. The **environment** describes the state of the forest and its evolution, producing an *observation* of that state and a *reward*, depending on carbon sequestered, biodiversity increased, or so on. The **agent**, or forest manager, must take an *action* based on the observation. The **optimizer** is the part of the algorithm which takes the reward and informs the agent's behavior.

4.1 The Environment: States and Observations

We simulate our DeepForest in a highly explainable fashion by building a custom OpenAI Gym environment from scratch.

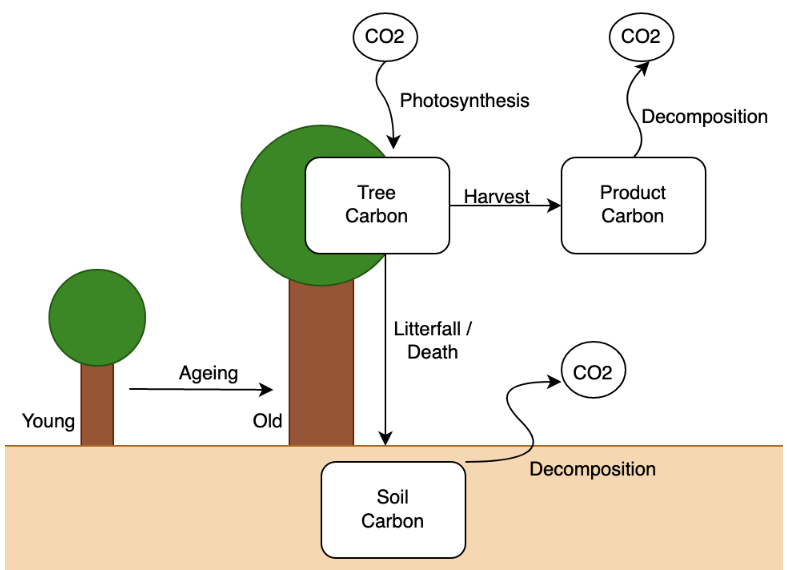

Figure 2: Carbon pools and fluxes in our simplified forest. Tree carbon grows via photosynthesis, and can either be lost to soil carbon or harvested to forest product carbon, the two of which have different decomposition rates. N.B. The system of tree counts, which tracks the reproduction and death of trees, is not depicted.

Complex carbon dynamics in ecosystems are often modeled with a system of ordinary differential equations [17]. In particular we draw inspiration from the Terrestrial Ecosystem (TECO) model [9].

$$\frac{dC(t)}{dt} = Bu(t) - A\xi(t)KC(t) \tag{1}$$

Here $Bu(t)$ is a vector which represents extrinsic factors expanding forest carbon pools, A a matrix which represents transfer rates, $\xi(t)$ a function which represents extrinsic factors shrinking carbon pools, and K a diagonal matrix which represents decay rates.

Our model differs in two key ways from Luo's implementation. First, instead of dividing carbon into 8 different plant, litter, and soil-based pools, we simplify with only $n+2$, where n is the number of tree species: $C = (C_{T,i}, C_S, C_P)$, or carbon in each tree species i, carbon in the soil, and carbon in forest products, respectively. We no longer treat forest litter as a distinct pool, but rather include it as part of soil carbon. While some models are interested in the soluble nitrogen and phosphorus compounds which rapidly decompose in litter, we are only interested in carbon, which remains in the "skeleton" of the leaf and decays together with resident soil [1].

Second, we introduce a system of $2n$ tree counts, given by $T = (T_{Y,i}, T_{O,i})$, or young and old trees for each species i, respectively. This is because younger trees have different growth, reproduction, and death rates from older trees, and some ecosystem models have set a precedent of separating the set of trees into age/size classes [11]. This also allows us to specify that new trees (seedlings) enter the class of young trees, while forest managers tend to only harvest from the class of old, mature trees, reflecting status quo reality.

Hence we define the state of DeepForest as the following vector: $(C_S, C_{T_i}, C_P, T_Oi, T_Oy)$.

The dynamics for the state are given as

$$\frac{dC_S}{dt} = -K_S C_S + \sum_{j=1}^{n} (K_{O,j}\nu_{O,j}T_{O,j} + K_{Y,j}\nu_{Y,j}T_{Y,j} + K_{T,i}C_{T,i}) \tag{2}$$

$$\frac{dC_{T,i}}{dt} = -K_{T,i}C_{T,i} + f_i(T_{O,i}, T_{Y,i}) - \alpha_i T_{O,i}\pi_i(C,T,t) - K_{O,j}\nu_{O,j}T_{O,j} - K_{Y,j}\nu_{Y,j}T_{Y,j} \tag{3}$$

$$\frac{dC_P}{dt} = -K_P C_P + \sum_{j=1}^{n} \frac{\alpha_i}{2} T_{O,i}\pi_i(C,T,t) \tag{4}$$

$$\frac{dT_{O,i}}{dt} = m_i T_{Y,i} - (\beta_i \pi_i(C,T,t) + K_{O,i})T_{O,i} \tag{5}$$

$$\frac{dT_{Y,i}}{dt} = -(m_i + K_{Y,i})T_{Y,i} + (\omega_{O,i}T_{O,i} + \omega_{Y,i}T_{Y,i})(1 - (\sum_{j=1}^{n} T_{O,j} + T_{Y,j})/N) \tag{6}$$

where π_i represents the forest manager's actions. Before we explain the rest of the formulae in detail, notice that the effects of agent actions and natural carbon dynamics occur concurrently. This is reflected in our reinforcement learning model environment as well. Additionally, notice that the transition between states as a result of the dynamics and the agent actions is deterministic.

4.1.1 Carbon Dynamics

Though the system is complicated, we can break down each equation into explainable parts.

Tree carbon grows via photosynthesis by $f_i(T_{O,i}, T_{Y,i})$ amount, a linear combination of the young and old tree counts [20]. The coefficients are scaled to reflect that young trees tend to grow faster than old trees.

Tree carbon simultaneously loses carbon to the soil via litterfall by $K_{T,i}C_{T,i}$ amount, where $K_{T,i}$ is the first-order rate constant [12]; and via death rates by $K_{O,j}\nu_{O,j}T_{O,j} + K_{Y,j}\nu_{Y,j}T_{Y,j}$ amount, also a linear combination of young and old tree counts [13]. Young trees have higher mortality rates than old trees.

Alternately, tree carbon can also be harvested by the agent and transferred to product carbon by $T_{O,i}\pi_i(C, T, t)$ amount (see 4.3).

Both soil carbon and product carbon decompose by $K_S C_S$ and $K_P C_P$ amount, also a first-order reaction [5], [18].

4.1.2 Tree Count Dynamics

The number of young trees increases logistically by $(\omega_{O,i}T_{O,i} + \omega_{Y,i}T_{Y,i})(1 - (\sum_{j=1}^{n} T_{O,j} + T_{Y,j})/N)$ amount, with a carrying capacity of N, through fruiting [6]. Old trees tend to produce more viable seeds, more often, on account of their size.

Young trees mature into old trees at a rate of m_i.

4.2 The Environment: Rewards

We also wish to define a metric for the value of a forest manager's decision, or more specifically, the marginal payoff the agent receives from a specific action. In RL, this is denoted as the reward function $R(C, T, \pi)$. Choosing the right policy entails solving the optimal control problem

$$\pi^*(C, T) = \text{argmax}_{\pi \in \mathcal{P}} \int_0^{t_f} e^{-\gamma t} R(C, T, \pi)\, dt \tag{7}$$

$$\frac{dC}{dt} = F(C, T, \pi) \qquad C(0) = C_{\text{init}}$$

$$\frac{dT}{dt} = G(C, T, \pi) \qquad T(0) = T_{\text{init}}$$

where $e^{-\gamma t}$ represents a discount rate (with $\gamma > 0$), C_{init} and T_{init} represent initial conditions, and $\pi = (\pi_1, ..., \pi_n) \in \mathcal{P}$ represents the state space of valid policies. We include the discount rate to encourage the agent to prioritize immediate rewards over far-off ones, since we tend to have time preferences for most goods – for example, many present-day targets like endangered species or at-risk coastlines may be irrecoverably lost to climate change unless we act to mitigate carbon soon.

We value the forest in 4 different ways: carbon sequestered, forest size, economic payoff, and biodiversity (calculated by Shannon entropy [15]). Thus, we represent R as a weighted sum of four different reward functions

$$R(C, T, \pi) = \sum_{i=1}^{4} w_i R_i(C, T, \pi)$$

where

$$R_1(C, T, \pi) = \nabla \cdot C \qquad R_2(C, T, \pi) = \nabla \cdot T$$

$$R_3(C, T, \pi) = \sum_{i=1}^{n} \frac{\alpha_i}{2} \pi_i(C, T) \qquad R_4(C, T, \pi) = \frac{d}{dt} \exp\{-\sum_{i=1}^{n} \big(\hat{p}_i \log(\hat{p}_i)\}$$

each of which corresponds to the marginal change in each of the four axes, respectively. ∇ is the divergence operator, and $\hat{p}_i$ is the proportion of trees of species i.

Because computers cannot solve the continuous time optimal control problem described above, we create the discrete analog which DeepForest learns to approximate the solution to.

$$\pi^*(C,T) = \text{argmax}_{\pi \in \mathcal{P}} \sum_{i=1}^{t_f} \gamma^i R(C,T,\pi) \tag{8}$$

In the above equation, note that $\gamma < 1$.

4.3 The Agent

For a system with n trees, we define the action space of the agent as a vector $\in [0,1]^n$, where each entry represents the marginal proportion of trees of a specific species to harvest at a given time. A policy $\pi \in \mathcal{P}$ maps from the state to the action space, and the set of valid policies $\mathcal{P}$ is defined as

$$\mathcal{P} = \{\pi : (C,T) \to [0,1]^n | \pi \text{ is fixed within each year}\}$$

We desire our agent, the forest manager, to find the optimal policy $\pi^*(C,T) = (\pi_1^*, ..., \pi_n^*)$ which satisfies Equation 8. The fact that π is fixed over a year reflects the fact that the agent can only take one new action each year, rather than continuously varying its policy. As $\mathcal{P}$ is easily seen to define a closed set and we are optimizing over a continuous function, finding the argmax of Equation 8 is a feasible problem.

4.4 The Optimizer

The rewards the agent is attempting to maximize are determined by the environment – the carbon cycle and other real-world payoffs in our case – and not explicitly known to the agent. Therefore, it must learn by interacting with the environment and observing the rewards it garners.

We employ a model-free approach where the agent does not attempt to explicitly model the environment but instead directly learns the utility of states and action. This is suitable as our model of the environment is deterministic, so the agent can effectually rely on samples from the environment. Furthermore, our action space is finite, so a value-based approach – understanding the value of specific actions in specific states – is more effective than policy-based approaches – directly learning policies.

4.4.1 The Q-Function

During learning and decision making, immediate rewards may be sparse or volatile, and foresight is important for learning the true value of taking specific actions in specific states. Therefore, the action-value function determines the true long-term utility of states and actions for the agent. The action-value function, or the Q-function, denotes the expected return of taking an action a in the state s following a policy π. It follows the Bellman optimality equation and takes the following form where s' and a' denotes the state and action in the next time step.

$$Q^\pi(s,a) = \mathbb{E}[R(s,a) + \gamma Q^\pi(s',a')|s,a]$$

The optimal action-value function is recursively defined as

$$Q^*(s,a) = \mathbb{E}[R(s,a) + \gamma \mathrm{max}_{a'} Q^*(s',a')|s,a]$$

The optimal policy is defined as

$$\pi^*(s) = \mathrm{argmax}_a Q^*(s,a) \tag{9}$$

Hence, with a good approximation of the Q-function, we can extract a close-to-optimal policy.

4.4.2 Deep Q Network

The task facing our DeepForest agent is particularly challenging because the state space is large and the environment is constantly changing. Additionally, there may be similar rewards for different actions, and rewards for the same action may change over time as the environment evolves. Therefore, we hope to leverage deep neural networks to approximate $Q^*(s,a)$ by learning parameters θ for a $Q_\theta(s,a)$. Specifically, we employ a neural network of 2 fully connected layers with a hyperbolic tangent activation function in between. Two layers is generally sufficient for capturing linear relationships as described in the DeepForest environment.

The right policy is attained through the connection between Q and π: our Deep Q Network (DQN) agent takes actions given by

$$a(s) = \mathrm{argmax}_a Q_\theta(s,a)$$

4.4.3 Network Design

DQN employs a mechanism called Experience Replay to generate training data from an agent's interaction with the environment. It stores information about the experiences needed for training and uses mini batches to update the neural network. This mechanism helps combat overfitting to current episodes, accelerates learning with mini-batches, and effectively reuses past interactions.

The DQN is trained by minimizing the following loss at each iteration:

$$L_i(\theta_i) = \mathbb{E}_{s,a}[(y_i - Q(s,a;\theta_i))^2]$$

where $y_i = \mathbb{E}_{s'}[(R + \gamma \mathrm{max}_{a'} Q(s',a';\theta_{i-1})|s,a]$. In other words, for each data sample, we compare the predicted Q value from current state s to the target Q value y_i, using network weights from the current ith iteration. This target Q value estimates the value of state s by utilizing the max Q value of the next state s' estimated by network weights from the $(i-1)$th iteration [8].

The network from the $(i-1)$th iteration is known as the Target Network, whose weights are periodically replaced after completing all training steps in an iteration. Fixing the Target Network's parameters when minimizing the loss helps improve stability of learning. Also note that the target Q value is dependent on the network weights, which is different from the definite ground truth labels available in supervised learning.

Additionally, note that the target Q value is calculated with the assumption that a greedy policy is followed, but the actual agent may not necessarily do so, a property called off-policy. We implement a linearly decreasing exploration schedule during training so that the agent not only exploits, but also explores the environment.

5 Results

All figures below are inference results of the policy derived from a trained neural network Q-function approximator according to 9. The initial conditions and hyperparameter values are kept unchanged from training to inference to stay consistent with real world values. During the inference run, in addition to the actions taken by the DeepForest manager, the environment continuously acts upon the state of the model, changing carbon and tree count according to 2 - 6.

5.1 Scenario 1: *Pinus sylvestris* temperate coniferous forest, no management

Pinus sylvestris, or the Scots pine, is an extremely widespread evergreen coniferous tree representative of temperate coniferous forests across Eurasia. Moreover, it often forms pure forests in northern Europe, making it a suitable subject for a single-species forest model.

We begin by examining how much carbon such a forest accumulates over the course of 100 years when there is no manager, and therefore no harvesting. The result is shown in Figure 3.

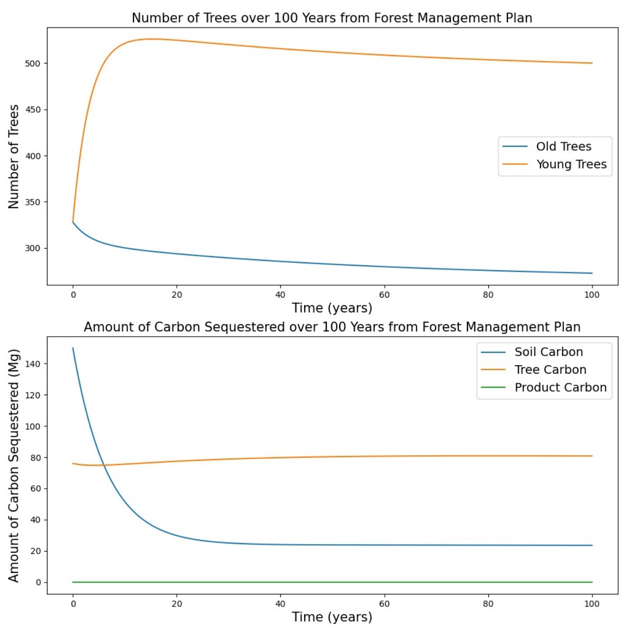

Figure 3: Scenario 1: *Pinus sylvestris* temperate coniferous forest, no management. Logistic growth.

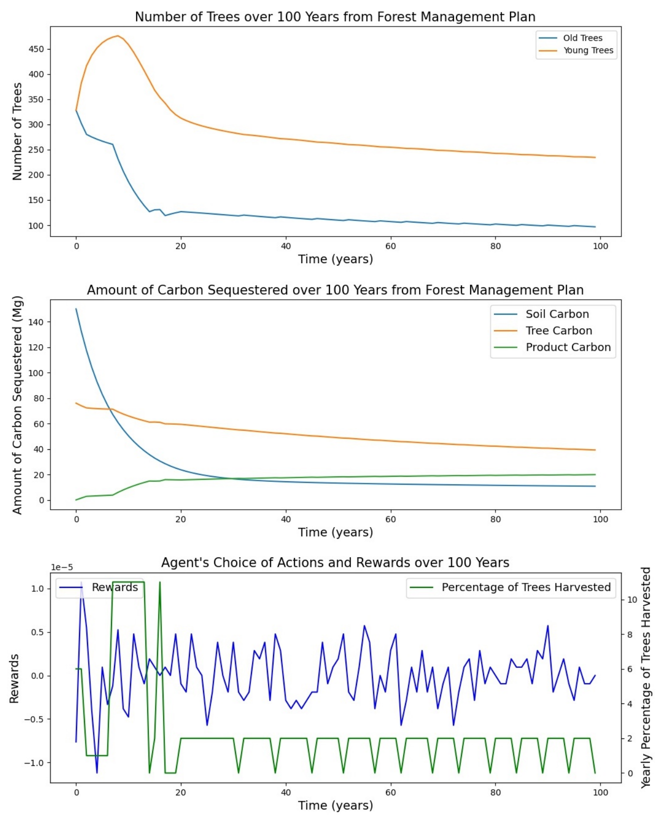

Figure 4: Scenario 2: *Pinus sylvestris* temperate coniferous forest, maximizing R_1. Periodic and modest harvests.

In an unconstrained forest, the number of trees initially increases exponentially, as we can see in the the young tree count curve for the first ten years of the simulation. However, the number of young trees levels out and slowly declines due to the carrying capacity, and so too does the number of old trees, since the maturation rate cannot keep up with the natural death rate. The soil carbon follows a first-order decay, as expected. The tree carbon remains relatively constant, balanced by the rapid growth rate of a great many young trees, against the litterfall and death rate.

5.2 Scenario 2: *Pinus sylvestris* temperate coniferous forest, maximizing R_1

We now allow the manager to harvest nonzero quantities of mature trees from the same forest, with the goal of sequestering as much carbon as possible. DeepForest yields the policy shown in Figure 4.

The agent learns a strategy which involves alternating between harvesting a consistent proportion of trees, and holding off on any harvesting to allow the forest to recover and sequester additional carbon. This periodic behavior mirrors that of the "shelterwood harvesting" strategy in real life, and the habit of postponing much of lethal harvesting is validated by [3]. Despite fluctuations in the immediate rewards, DeepForest appears to have learned to look beyond the instantaneous benefits to aim for more long-term carbon sequestration.

5.3 Scenario 3: *Pinus sylvestris* temperate coniferous forest, maximizing R_3

As discussed earlier, forests are not solely valued for their carbon sequestration potential. When we instead seek to maximize economic profit from the Scots pine forest of Scenarios 1 and 2, we get the policy shown in Figure 5.

The agent harvests in a conservative fashion for the first half of the century, before beginning to cut over 60% of mature trees for the remainder of the century. In other words, the agent treats the forest as a long-term investment with interest: instead of harvesting early for immediate reward, the agent holds off and allows the forest to grow and reproduce unmolested, before quickly ramping up a consistently high harvest rate for a much greater profit.

We conjecture that the agent decides to transition to substantial harvesting when the number of trees falls beyond a certain threshold, beyond which the benefits from natural reproduction can no longer offset the economic interests of harvesting. Although product carbon does increase substantially, poor tree counts and forest carbon sequestration indicate the ecosystem is probably severely damaged by the end of the century. This management strategy mirrors that of real-life "clear-cutting," which is indeed employed frequently on pine forests. However, "clear-cutting" is only sustainable with old, unhealthy stands, and clearly ruins healthy forests.

5.4 Scenario 4: *Pinus sylvestris* temperate coniferous forest, maximizing $R_1 + R_3$

Now that we've examined the two ways a forest can be valued – carbon fixation and economic productivity – in isolation, we now choose a reward function that combines both for good measure, with equal weightings of $0.5, 0.5$, resulting in Figure 5.

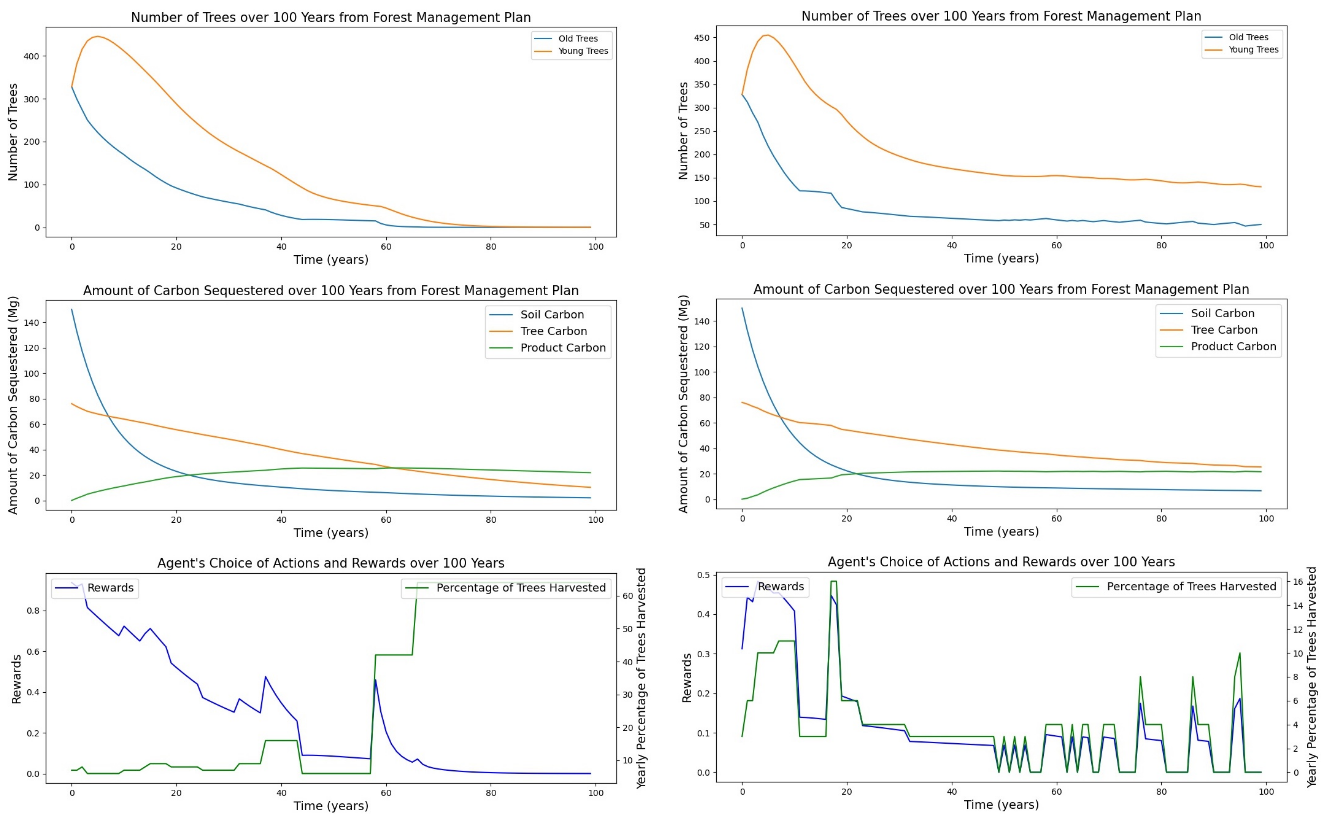

Figure 5: (Left) Scenario 3: *Pinus sylvestris* temperate coniferous forest, maximizing R_3. Late and greedy harvests. (Right) Scenario 4: *Pinus sylvestris* temperate coniferous forest, maximizing $R_1 + R_3$. Late and periodic harvests.

We expect a management strategy that combines elements of the strategies from Scenario 2 and 3, and this is indeed what happens. With the exception of a brief aggressive harvesting round at the beginning of the century reminiscent of Strategy 2, the agent mirrors Strategy 3 by harvesting more conservatively until the half-century. However, when it does begin to cut down trees more aggressively, it does so in the periodic cycles characteristic of Strategy 2.

This distinct combination – harvesting trees and allowing the forest to recover in cycles, after a period of limited activity – shows that DeepForest is capable of attaining a balance between two selected values in a reasonable and transparent fashion. Meanwhile, it is also important to note that DeepForest does not merely combine the policies for the two objectives of carbon sequestration and economic payoff directly – it adapts and adjusts its strategy based on the ever changing environmental conditions and the experience it has gained through training.

5.5 Scenario 5: *Pinus sylvestris* temperate coniferous forest, left uncut

In all of the scenarios we've encountered so far, the manager has chosen to cut down a nonzero number of trees. We return to the simple forest of Scenario 1 with the goal of finding a set of conditions that might induce a manager to leave the forest entirely uncut. For the sake of simplicity our reward function is R_1 only.

Indeed, when growth, reproduction, or product decay rates increase sufficiently; or when carbon return per harvested tree, litterfall, soil decay, or tree death rates decrease sufficiently, the agent no longer makes any harvests. Making each of these aforementioned hyperparameter changes either in isolation or in combination yields a policy similar or identical to Figure 6.

DeepForest recognizes that there is no scenario in which transforming a tree into product carbon is worthwhile, either because that tree sequesters more carbon over the course of its remaining lifetime, or because its carbon remains sequestered for longer as part of the forest. Hence, it converges at an optimal policy of leaving the forest uncut.

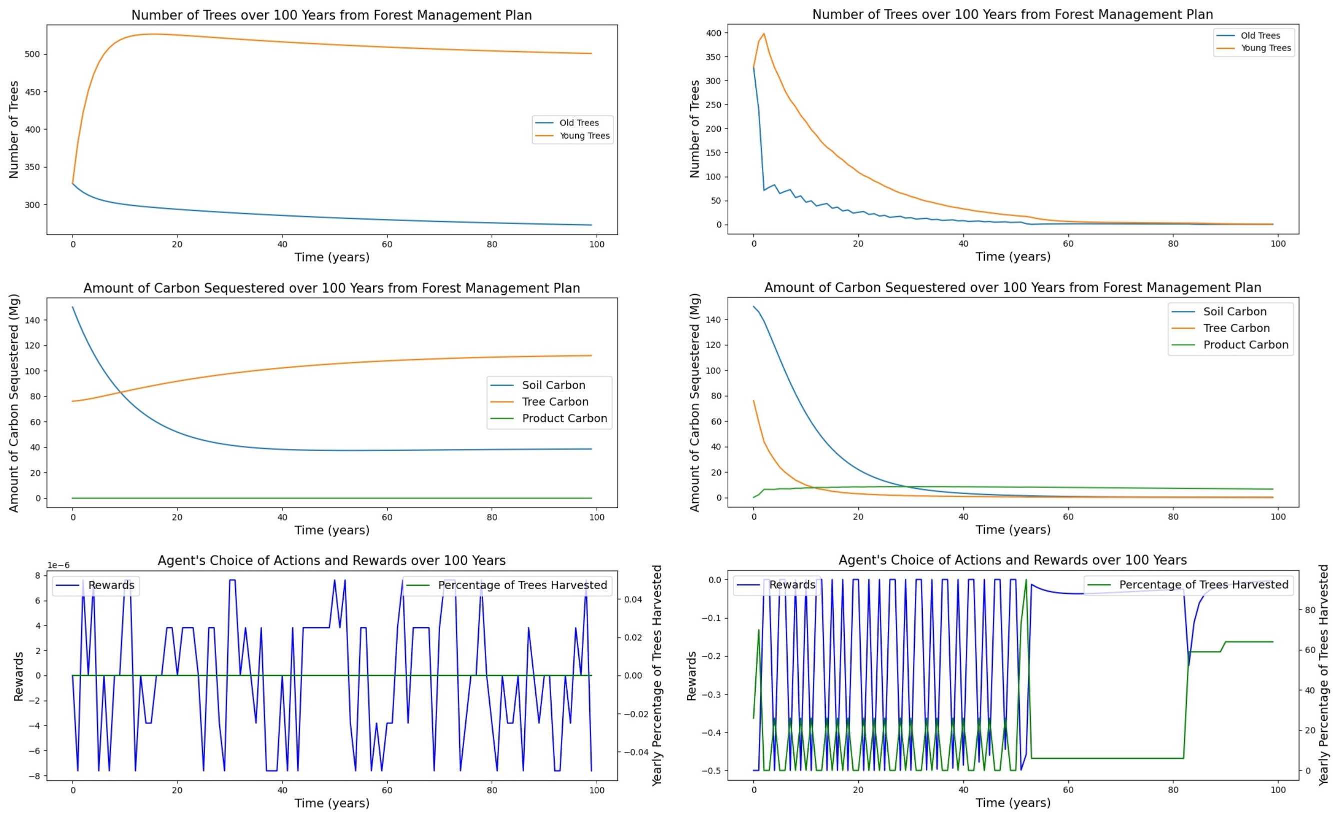

Figure 6: (Left) Scenario 5: *Pinus sylvestris* temperate coniferous forest, left uncut. Harvesting isn't worth it. (Right) Scenario 6: *Betula pendula* temperate deciduous forest, maximizing R_1. Early, aggressive harvests.

5.6 Scenario 6: *Betula pendula* temperate deciduous forest, maximizing R_1

In order to demonstrate that DeepForest is capable of operating on a variety of different forests, we now simulate a temperate deciduous forest composed of *Betula pendula*, the silver birch, a deciduous tree with similarly widespread distribution throughout Eurasia. Deciduous forests are characterized by the fact that a great deal of tree carbon is lost as shed leaves during the autumn and winter, so the annual litterfall rate is much higher than in Scenario 1. Moreover, the silver birch is a smaller tree than the Scots pine, so the average carbon mass of a mature species is smaller. Consequently, DeepForest produces the policy shown in Figure 7.

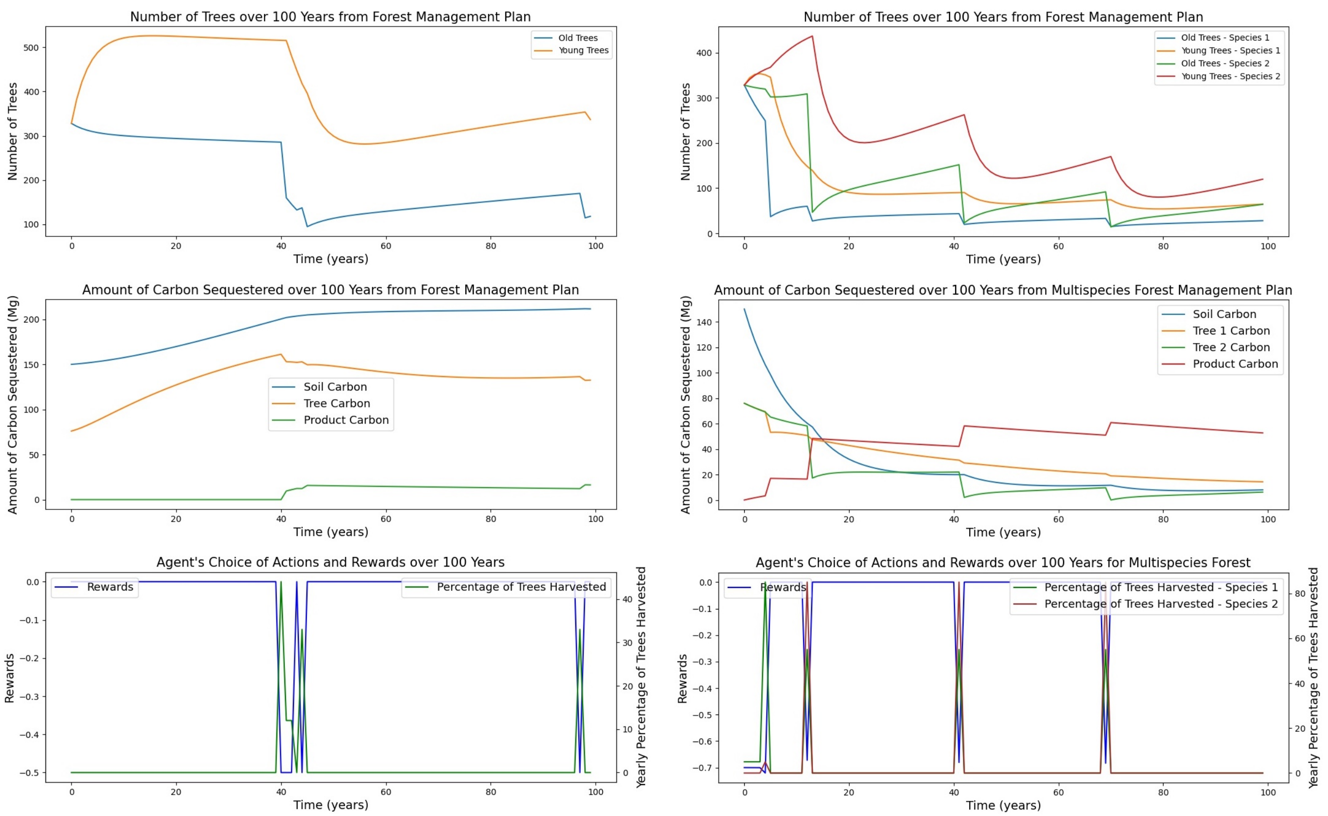

Figure 7: (Left) Scenario 7: *Iriartea deltoidea* tropical rainforest, maximizing R_1. Few and moderate harvests. (Right) Scenario 8: *Pinus sylvestris* & *Picea abies* temperate coniferous forest, maximizing R_4. Periodic harvests of more prolific species.

Here the agent must balance between the need to preserve tree carbon lest it be lost to soil carbon, and the reduced payoff of cutting down smaller trees. It seems that the former won out, and the optimal policy is to cut down a greater proportion of trees early on.

5.7 Scenario 7: *Iriartea deltoidea* tropical rainforest, maximizing R_1

We also explore the environment of a tropical rainforest composed of *Iriartea deltoidea*, a very common species of canopy palm in Nicaragua, Bolivia, and a great portion of the Western Amazonian basin. Tropical rainforests have higher NPP in general than temperate ones on account of their year-round warm, wet, sunny equatorial climate. And in sufficiently waterlogged soils, anaerobic conditions inhibit decomposition, leading to Figure 7.

The combination of excellent growth by living trees and the relatively long soil residence of litterfall and dead matter makes harvesting a relatively unappealing prospect, at least for the purpose of carbon sequestration.

5.8 Scenario 8: *Pinus sylvestris* & *Picea abies* temperate coniferous forest, maximizing R_4

Forests are also valued for their biodiversity – a diverse forest has greater genetic resources and ecological stability. Rich forests are often considered to have cultural and recreational values, perhaps as parks and reserves, not to mention aesthetic beauty.

The Scots pine can also form mixed forests with, notably, *Picea abies*, or the Norway spruce. When we simulate a forest with both species in abundance and seek to maximize biodiversity, we get the policy shown in Figure 7. We use Shannon entropy to measure biodiversity, and rewards get better when each species has similar relative counts. We also reward nonzero tree counts to discourage DeepForest from cutting down all the trees as a "cheat."

The DeepForest appears to successfully maintain a relatively similar number of trees of the 2 species throughout the century. It does so through intermittent harvesting seasons that act as checks on the relative and absolute abundance of the trees. Notably, the harvesting of the 2 species occur at the same time step, but the magnitude of the percentage of trees to harvest vary between the 2 species. The agent starts off harvesting more of species 1, but when species 2 begins dominating in numbers, the agent transitions to harvest a greater proportion of species 2 than species 1. With this strategy, the agent is able to maintain biodiversity and forest size.

Without DeepForest's management, the Norway spruce tends to outcompete the Scots pine, just like in real Northern European forests. The fact that our environment arrives at such accurate conclusions when primed with real-life initial values speaks well of its performance.

6 Sensitivity Analysis

Because DeepForest incorporates a very large number of parameters, it may not be feasible to test its sensitivity numerically. However, we may analyze the sensitivity of these parameters with the theory of dynamical systems. For simplicity, we only provide analysis of an environment with only one tree species, though this theory may be easily extended to systems with more. In extending the system, however, the theory becomes considerably more complex, so we make no claims about the parameter sensitivity of diverse forests.

We consider two different types of sensitivity. The first is sensitivity of the optimal control problem defined in equation 7. That is, we ask if we start with initial conditions $C_{\text{init}} + \epsilon_C C_1$ and $T_{\text{init}} + \epsilon_T T_1$ for arbitrary vectors C_1 and T_1 and small $\epsilon_C, \epsilon_T > 0$, how much will π^*_ϵ, the new optimal policy, deviate from π^*? Dontchev and Hager provide sufficient conditions to guarantee that the deviation of π^*_ϵ is minimal for small enough $\epsilon > 0$, so we do not elaborate on the conditions here [2].

The second type of sensitivity we consider is parameter sensitivity. That is, given a policy and a set of initial conditions, we analyze evolution of the ecosystem given small perturbations to the parameters. To analyze this type of sensitivity, we calculate the Jacobian of ODEs described in equations 2 - 6. See [19] for an explanation about the importance of the Jacobian in analyzing parameter sensitivity.

The system is observed to have at most 2 fixed points, with one at 0, and another which cannot be found analytically. For ease of analysis, we only linearize about 0, observing that

$$J_T(0) = \begin{bmatrix} -K_S & K_{T,1} & 0 & K_{O,1}\nu_{O,1} & K_{O,1}\nu_{O,1} \\ 0 & -K_{T,1} & 0 & -K_{O,1}\nu_{O,1} + f_{T_{O,1}}(0) & -K_{Y,1}\nu_{Y,1} + f_{T_{Y,1}}(0) \\ 0 & 0 & -K_P & \pi_{T_{O,1}}(0) & \pi_{T_{Y,1}}(0) \\ 0 & 0 & 0 & -\beta\pi_{T_{O,1}}(0) - K_{O,1} & m_1 \\ 0 & 0 & 0 & \omega_{O,1} & \omega_{Y,1} - m_1 K_{Y,1} \end{bmatrix}$$

The sensitivity of the parameters for ODE states close to 0 will be given by the eigenvalues λ_i of $J_T(0)$. If we can show that $\Re(\lambda_i) \neq 0$, then we can conclude that perturbing the parameters slightly will not vastly change the behavior of the system. The eigenvalues are calculated to be

$$\begin{aligned} \lambda_i = {} & -K_S, -K_{T,1}, -K_P, \\ & -\frac{1}{2}(\omega_{Y,1} m_1 K_{Y,1} - \beta\pi_{T_{O,1}}(0) - K_{O,1}) \pm \frac{1}{2}((\omega_{Y,1} - m_1 K_{Y,1} - \beta\pi_{T_{O,1}}(0) - K_{O,1})^2 \\ & \qquad - 4((\omega_{Y,1} - m_1 K_{Y,1})(\beta\pi_{T_{Y,1}}(0) + K_{O,1}) + \omega_{O,1} m_1))^{1/2} \end{aligned}$$

In the cases above, we observe that so long as $(\omega_{Y,1} - m_1 K_{Y,1})(\beta\pi_{T_{Y,1}}(0) + K_{O,1}) \neq 0$, the system is resilient to minor parameter perturbations. Given that the parameters in each of the scenarios above avoided this region, we are justified in using computer systems to solve the differential equations and the optimization problem.

7 Discussion

7.1 Model Effectiveness

DeepForest is a highly explainable model. Each component of the ordinary differential equations that govern its environment is derived from mechanisms in real-world forests. It represents carbon and tree count dynamics in a clear, concise, and widely accepted [9] fashion. Because Deepforest can easily be broken down for an in-depth understanding of how carbon is flowing through the system and why, it is particularly constructive for policymakers, who rely on transparent causal inferences to guide their understanding and effect change.

Characterizing the forest manager as DeepForest's agent further enhances its explainability. At every time step, DeepForest directly computes how many trees of each species to harvest, rendering decision-making on the user's part much simpler. Our agent employs both past experience and long-term foresight to learn but only focus on the present and future when making decision render it a great guidance for human decision making.

Another notable strength of our model is its capability to adapt to drastically different initial conditions. DeepForest is exceptionally flexible for applications to alternate circumstance. Parameter values can be easily changed to match any specific real-world situations, and direct analysis of the effects of isolated parameter changes can be easily studied. It is also remarkably easy to introduce new components, whether it be new flows between sources of carbon, additional factors impacting the number of trees, or entirely new reservoirs of carbon.

The decision model is also highly flexible. The rewards and their corresponding weighting can be easily altered to reflect various goals and needs of the policy makers. The capability of DQN to compactly represent both the high-dimensional observation space and the Q-function with a relatively simple architecture also aids with the ease of extension of our model. As have been discussed, DeepForest is capable of suggesting a wide spectrum of management plans in response to different objectives and scenarios. Its high adaptability allows us to apply the pattern recognition strength, data extrapolation capability, and computation power of machine learning to a wide array of circumstances.

Additionally, our model-free value-based RL approach does not require any training data, which is difficult to obtain in the real world. It also enables quick training even when computation power is limited. It is important to note, however, that the lack of ground truth data makes it impossible to confirm the "correctness" of the decisions our agent makes and can sometimes lead to high variance when agent learns through exploration.

7.2 Management Strategies

As have shown in 5, our model suggests diverse set of management strategies based on the circumstance. There are a few common and overarching paradigms.

1. **Alternating periods of harvests and forest recovery.** For systems similar in carbon balance to our Scenario 1 baseline, DeepForest often suggests alternating between periods of preservation and harvest, allowing the forest can grow and recover before the next culling. The exact periods are typically relatively short, on the scale of a few years. This idea of sustainable harvesting appears to be relevant across a wide variety of initial conditions and configurations.

2. **Stepwise progression.** When DeepForest aggressively initiates or halts sustained harvests, it tends to suggest stepwise escalation or reduction. This typically occurs across longer periods of time, on the scale of decades, and appears when DeepForest is more focused on a specific reward, such as carbon sequestration or economic payoff, or when the carbon balance is tipped far towards accumulation or loss.

3. **Transition across plans** For almost all scenarios, our agent shifts across different management plans instead of staying fixated on a set course of action throughout. The transition points often match changes in the trajectory of the state, which encompasses carbon amount and tree count. Specifically, transition points between management plans appear to happen around transformation of the convexity of the trajectories of carbon amount and tree count as well as around intersection of the curves of carbon amount.

   Redirecting and changing management plans appears to be especially relevant when the agent is informed to pursue composite objectives, for which we rarely observe the agent to maintain a solitary plan throughout the entire century. This is coherent with the intuitive idea that agility and openness to change is particularly important when there is the need to balance multiple values.

7.3 A More Effective Forest Transition Policy in Russia

As was observed with *Pinus sylvestris* in Scenarios 1-4, it was often more desirable to include tree harvesting as part of the forest management plan, even when valuing sequestered carbon more highly than economic benefits. But even so, it is possible that many countries are over harvesting their supply of specific trees.

Consequently, International Carbon Management Collaboration (ICMC) has observed this in Russia and proposed a carbon management plan in Great Britain that consists of scaling back the annual forest harvests of *Pinus sylvestris* (Scots pine) to one harvest every 10 years. Though this may produce better environmental benefits in the long term, Lutz et al. demonstrate the popularity of harvesting from forests in Russia, as otherwise it would need to rely on China for its wood [10]. Any immediate implementation of ICMC's policy would result in a mass disruption in multiple industries, and as Garcia et al. point out, it would be unlikely to stick. The industries benefiting from the harvesting of Scots pine are unlikely to adhere to such a long term policy without any concrete benefits in the short term [4].

To counter this blowback, we propose the gradual introduction our plans to the target forests in Russia. There are multiple actors are affected by the management of Scots pines, namely the loggers (both legal and illegal) and the forest managers [10]. Though designing a specific transition policy is beyond the scope of this paper, we still briefly outline a transition plan here. Rather than immediately scaling back logging, this can be done over the process of multiple years. This may not produce the optimal net benefit in respect to carbon sequestration, but prevents any potential shock to the economy that might come as a result of increased logging intervals. To make sure loggers adhere the increased intervals, the forest manager (and by extension the government) should incentivize the decrease in logging, possibly through subsidies or tax benefits. To prevent potentially illegal logging, forest managers should increase forest protection and attack the root causes of logging.

In terms of our model, we can train DeepForest to learn such a policy by using a reward function in the form

$$R(C, T, \pi) = tR_1 + (t_f - t)R_2$$

Thus, closer to the beginning of the simulation, DeepForest will prioritize maximizing R_2 to preserve the economic payoff, but as it draws closer to the end of the simulation (for large t), it will be more beneficial to maximize R_1, i.e, carbon sequestration. Though we do not implement this scenario due to time constraints, we are confident training a model with the above scenario would effectively figure

8 Limitations and Future Work

Below we list a few drawbacks to our model and explore the potential for improvement.

1. **Simplifying Assumptions.** As ecosystems are extremely complex, we are forced to make simplifying assumptions to study the carbon and tree dynamics in a forest. For example, nitrogen and phosphorus cycles are neglected, with the assumption that they are encoded in growth and carrying capacity. Inevitably this causes vulnerability to overlooking information that may potentially be important down the line. Future work might include working to introduce more carbon pools or more nuanced dynamics to better capture the effects of the shifting ecosystem.

2. **Model Complexity.** The dynamics of both carbon and tree count follow a complex system of ODEs. By virtue of the nonlinearities introduced by the agent's policy and the logisitic

growth in Equation 6, the model likely exhibits chaotic behavior and may be inaccurate over very long periods of time. Introducing more variables such as carbon pools or different tree age classes would also increase the sensitivity of the model and potentially create more erractic behavior. Future steps include creating models which can measure more changes while still being resilient to chaotic behavior.

3. **Explainability of DeepForest's Policy.** As is the issue with many neural networks, DeepForest exhibits a black box issue. Though the overall architecture of DeepForest is easily interpreted– taking actions which maximize $R(C, T, \pi)$– the specifics of its calculations (such as assigning values to each state-action pair) are too complex to interpret. Working to understand why DeepForest works with an ecological motivation will help to make its management plans more robust.

4. **Scalability for Very Diverse Forests.** With the addition of each new tree species to the environment, the agent's action space increases by an extra dimension, and for very diverse forests, the becomes quite computationally expensive. Furthermore, because there are exponentially more actions the agent can take, the algorithm will take significantly longer to train. Thus, our model works most optimally for a relatively small number of types of trees, but we would benefit from learning to code a more efficient and optimized model.

5. **Stochasticity.** We do not include any stochasticity in the model, which could help account both for differences in measurements (such as tree age and size) and the occurrence of rare, but catastrophic events such as epidemics and wildfires. Fortunately, the optimal control framework lends its self well to studying continuous time stochastic processes, and as such, it would be beneficial to let DeepForest learn to solve stochastic optimal control problems instead.

9 Conclusion

To recap, this paper introduces DeepForest, a general RL-based framework for forest management for carbon sequestration and other purposes. By designing a system of ODEs to govern carbon cycle and tree growth dynamics, we create an environment in which DeepForest can use Deep Q Network to learn close-to-optimal policy for forest management without training data and in limited computational power. By considering different ecosystems and different values such as carbon sequestration, economic payoff, biodiversity, and forest size, the model learns policy plans for a wide variety of circumstances and objectives. Moreover, DeepForest simulations are consistent with our expectations of real-world forest management plans and offer guiding insight for human decision making. Finally, we discuss a general framework to introduce these plans to actual forests with disruption, using Russia as a case study.

References

[1] Matson Pamela A Chapin, Stuart F. and Peter M. Vitousek. *Principles of Terrestrial Ecosystem Ecology*. Springer, London, 2021.

[2] A. L. Dontchev and W. W. Hager. Lipschitzian stability for state constrained nonlinear optimal control. *SIAM Journal on Control and Optimization*, 36(2):698–718, 1998.

[3] Jiang J. Ding W. Gaoue, O.G. Optimal harvesting strategies for timber and non-timber forest products in tropical ecosystems. *Theor Ecol*, page 287–297, 2016.

[4] Claude A. Garcia, Sini Savilaakso, René W. Verburg, Victoria Gutierrez, Sarah J. Wilson, Cornelia B. Krug, Marieke Sassen, Brian E. Robinson, Hannah Moersberger, Babak Naimi, Jeanine M. Rhemtulla, Hélène Dessard, Valéry Gond, Cédric Vermeulen, Franck Trolliet, Johan Oszwald, Fabien Quétier, Stephan A. Pietsch, Jean-François Bastin, Anne Dray, Miguel B. Araújo, Jaboury Ghazoul, and Patrick O. Waeber. The global forest transition as a human affair. *One Earth*, 2(5):417–428, 2020.

[5] Ryan M. Giardina, C. Evidence that decomposition rates of organic carbon in mineral soil do not vary with temperature. *Nature*, page 858–861, 2000.

[6] Gwendal Restoux Annabelle Amm Christian Pichot Bruno Fady Hendrik Davi, Maxime Cailleret. Disentangling the factors driving tree reproduction. *Ecosphere*, 2016.

[7] Michael D. Korzukhin, Michael T. Ter-Mikaelian, and Robert G. Wagner. Process versus empirical models: which approach for forest ecosystem management? *Canadian Journal of Forest Research*, 26(5):879–887, 1996.

[8] Longxin Lin. Reinforcement learning for robots using neural networks. 1992.

[9] Y. Luo, Z. Shi, X. Lu, J. Xia, J. Liang, J. Jiang, Y. Wang, M. J. Smith, L. Jiang, A. Ahlström, B. Chen, O. Hararuk, A. Hastings, F. Hoffman, B. Medlyn, S. Niu, M. Rasmussen, K. Todd-Brown, and Y.-P. Wang. Transient dynamics of terrestrial carbon storage: mathematical foundation and its applications. *Biogeosciences*, 14(1):145–161, 2017.

[10] David A. Lutz, Herman H. Shugart, and Mark A. White. Sensitivity of Russian forest timber harvest and carbon storage to temperature increase. *Forestry: An International Journal of Forest Research*, 86(2):283–293, 01 2013.

[11] Pekka Malo, Olli Tahvonen, Antti Suominen, Philipp Back, and Lauri Viitasaari. Reinforcement learning in optimizing forest management. *Canadian Journal of Forest Research*, 51, 03 2021.

[12] Kutjavin Ivan N. Osipov, Andrey F. and Kapitolina S. Bobkova. Ratios between aboveground net primary production, litterfall and carbon stocks in scots pine stands (russia). *CERNE [online]*, 2021.

[13] Timo Pukkala, Erkki Lähde, and Olavi Laiho. Species interactions in the dynamics of even- and uneven-aged boreal forests. *Journal of Sustainable Forestry*, 32(4):371–403, 2013.

[14] Anna Repo, Tuomas Rajala, Helena M. Henttonen, Aleksi Lehtonen, Mikko Peltoniemi, and Juha Heikkinen. Age-dependence of stand biomass in managed boreal forests based on the finnish national forest inventory data. *Forest Ecology and Management*, 498:119507, 2021.

[15] Carlo Ricotta. Bridging the gap between ecological diversity indices and measures of biodiversity with shannon's entropy: comment to izsák and papp. *Ecological Modelling*, 152(1):1–3, 2002.

[16] Max Rietkerk, Stefan Dekker, Peter Ruiter, and Johan van de Koppel. Self-organized patchiness and catastrophic shifts in ecosystems. *Science (New York, N.Y.)*, 305:1926–9, 10 2004.

[17] M.J. Schelhaas, Esch, Thomas Groen, Hans Jong, Bernardus Jong, Markku Kanninen, Jari Liski, Omar Masera, G.M.J. Mohren, Gert-Jan Nabuurs, Taru Palosuo, L. Pedroni, A. Vallejo, and Terhi Vilén. Co2fix v 3.1 a modelling framework for quantifying carbon sequestration in forest ecosystems. 01 2004.

[18] Brad Seely, Clive Welham, and Hamish Kimmins. Carbon sequestration in a boreal forest ecosystem: results from the ecosystem simulation model, forecast. *Forest Ecology and Management*, 169(1):123–135, 2002. The Role of Boreal Forests and Forestry in the Global Carbon Budget.

[19] Steven H. Strogatz. *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry and Engineering*. Westview Press, 2000.

[20] Guan J. H. Deng L. Zhang J. G. Li G. Du S. Yue, J. W. Allocation pattern and accumulation potential of carbon stock in natural spruce forests in northwest china. *PeerJ*, 01 2018.

Appendix: Baseline Simulation Parameters

The values of the parameters used in the baseline experiment (Scenario 1 and 2) are shown below:

Parameter	Description	Value
C_S	The carbon pool representing soil carbon (Mg)	Initial value = 150
C_P	The carbon pool representing product carbon (Mg)	Initial value = 0
$C_{T,i}$	The carbon pool representing the amount of tree carbon for tree i (Mg)	Initial value = 76
$T_{O,i}$	The number of old trees of type i	Initial value = 328
$T_{Y,i}$	The number of young trees of type i	Initial value = 328
m_i	The maturation rate of tree i	1/37
K_T	The decomposition rate of trees (litterfall rate)	0.025
$K_{O,i}$	The death rate of old trees of type i	0.05
$K_{Y,i}$	The death rate of young trees of type i	0.25
K_S	The decay rate of soil	0.15
K_P	The decay rate of product	0.005
N	The carrying capacity (max number of trees) in the forest	3000
α_i	The tree to carbon conversion rate	0.076
β_i	The policy to tree harvest scaling factor	1
$\nu_{O,i}$	The tree death to carbon loss scaling factor in old trees	0.076
$\nu_{Y,i}$	The tree death to carbon loss scaling factor in young trees	0.0038
$\omega_{O,i}$	An old tree growth parameter	0.5
$\omega_{Y,i}$	A young tree growth parameter	0.1

The Math Modeler

21 February, 2021 Volume 1 Moscow, Russia

So cutting down trees might be good?

21 February, 2021

A novel study by the International Carbon Management Collective (ICMC) using machine learning has found surprising results about the effectiveness of incorporating tree culling into forest management as a way to address climate change.

One of the most pressing issues in climate change concerns the release of carbon dioxide (CO_2) into the air. If an excessive amount of CO_2 remains in the air, this can disrupt climate patterns and damage societies in the long run. In fact, this is already happening right now in many parts of the world, with increased annual floods and hotter winters being a direct effect of the CO_2 which has accumulated in the sky

Controlling the release of CO_2 is a problem in ecology commonly known as *carbon sequestration*, a field in which organic matter's ability to store carbon is studied. Consider the plants we eat and the trees we cut down– all of it contains carbon directly absorbed from the CO_2 in the air in a process known as *photosynthesis*. One of the most common methods of fighting climate change is to plan a great number of trees in order to take CO_2 out of atmosphere.

So all we need to do then to save the world is plant a bunch of trees then, right? Wrong. New evidence shows that it may actually be beneficial to cut down a number of trees each year. Take the *Pinus sylvestris*, more commonly known as Scots pine. This coniferous tree is common throughout Europe and Russia and is known for its great amount of litterfall, or the amount of branches and leaves falling from the tree.

Living trees are better at retaining their carbon than the dead branches and leaves, which quickly get absorbed into the soil. From here, microbes in the soil will break the organic compounds into CO_2, releasing it back into the atmosphere. So when a tree such as a Scots pine exhibits a great amount of litterfall, it may actually be *losing* carbon if it loses the organic material in its leaves and branches faster than it can absorb CO_2 from the atmosphere.

In such a scenario, as the ICMC's machine learning algorithm pointed out, it may actually be beneficial to cut the Scots pine down in order to better sequester the carbon. The wood in furniture, fences, and other processed goods will decompose slowly, at a much lower rate than the organic matter falling from trees.

With this realization, forest managers are making new policies which include harvesting a small proportion of trees such as Scots pine each year. In doing so, society reaps the potential economic benefits of the wood while keeping CO_2 out of the atmosphere. In other words, a win-win.