When clouds go dry: an integrated model of deforestation, rainfall, and agriculture

Rafael Araujo*

July 27, 2023

Abstract

Deforestation of tropical forests affects rainfall, changing the productivity of the agricultural sector, the main driver of deforestation. This deforestation-rainfall mechanism is in effect even in regions that are thousands of kilometers away from the forest, but does it result in a sizable externality? I develop an integrated climate and land-use model to measure the externality impact that land use decisions have on agricultural productivity through changes in rainfall. As an application, I estimate the model with pixel level climate data for the entire Amazon Rainforest and pixel level land use data for the Brazilian state of Mato Grosso, one of the most important agricultural hubs in the world. I then consider a counterfactual where farmers are allowed to deforest protected areas. I find that, due to the precipitation decrease resulting from deforestation, the returns of crop production decrease by 2% with some regions losing up to 8%. Tail events increase this loss of crops production in some regions to 12%.

JEL: *Q23, Q56, Q15, C35, L73*

Keywords: *Deforestation, Amazon, Climate, Land Use, Discrete Choice Model*

1 Introduction

Tropical forests provide crucial ecosystem services that affect the carbon cycle (Baccini et al., 2012), temperature (Lawrence et al., 2022), biodiversity (Haddad et al., 2015), and rainfall (Spracklen et al., 2012). Tropical deforestation has the potential to alter rainfall patterns in a continental scale via a rainfall-transpiration mechanism: trees are responsible for increasing air's humidity contributing to rainfall downwind. While tropical deforestation is predominantly driven by agriculture (Pendrill et al., 2022), less trees lead to less water, a fundamental input for agriculture. This human-nature interaction can generate a sizable externality, since some agents can gain from the deforestation in one region, by having more available land, while negatively affecting the production of regions that are hundreds or even thousands of kilometers away from them, by modifying the climate.

Despite the importance of this deforestation-rainfall mechanism, known for decades in natural sciences (Spracklen et al., 2018), the economics literature overlooks the feedback between land use decisions and climate (Balboni et al., 2022). The literature that focuses on causes and consequences of rainfall has not addressed the interaction of rainfall changes and human behavior with an integrated model. There are studies on the effects of rainfall and water availability on agricultural production (Ortiz-Bobea et al., 2019, 2021; Mérel and Gammans, 2021; Skidmore, 2022; Rattis et al., 2021; Rafey, 2020) that take climate as given and a literature on natural sciences studying the effects of deforestation on rainfall (Spracklen et al., 2012; Lawrence and Vandecar, 2015; Wright et al., 2017; Jasechko et al., 2013) that disregards the adaptation of agents to the changing climate. The economics literature that models deforestation of tropical forests as a discrete choice problem (Domínguez-Iino, 2021; Souza-Rodrigues, 2019; Araujo et al., 2020; Hsiao, 2021) does not consider the climate externalities of the agents' land use choices. In a broader literature of human-nature interaction, to the best of my knowledge, the only topic studied in an integrated model is related to greenhouse gas emissions and economic production (Nordhaus, 2019).

To fill this gap, I build an integrated model of land use choice and climate with two components. First, I develop a discrete choice model of land use, which accounts for human adaptation behavior, a feature that is not present in climate models from natural sciences. Second, I develop a climate model of precipitation, which accounts for climate externalities among agents, a feature not present in models from economics. In my model, farmers decide whether to deforest their plot of land. If the decision is to deforest, they choose which agricultural product to produce, considering that each choice of agricultural product is subject to a different soil suitability, transportation cost, and responds differently to changes in rainfall. As rainfall affects the return of agricultural activity it also affects the

1

deforestation decision.

Concomitantly, the deforestation decision of one farmer affects the rainfall supply of other farmers. The climate model of precipitation provides the means through which agricultural expansion/deforestation affects the natural supply of water inputs to agriculture. This effect of deforestation on rainfall is captured by a complex network of humidity transport, a phenomenon known in the literature as flying rivers (Marengo et al., 2004; Nobre, 2014). With a newly developed approach to model atmospheric circulation along thousands of kilometers, I create a measure of how much exposure to upwind forest each farmer faces. My model requires data of land use and wind patterns. Wind patterns compose the circulation of the atmosphere and can be used to draw the trajectories of air parcels over land. Trajectories that travel over areas with larger forest coverage have their humidity recharged by the forest's transpiration and are able to deliver more rainfall downwind. I show that the upwind exposure to the forest has a causal effect of significantly increasing the farmer's rainfall supply, a mechanism that is in effect even in places that are located hundreds or thousands of kilometers away from the forest. This climate model allows me to compute a causal decrease in rainfall for any possible scenario of deforestation.

The next step is to integrate both land use and climate models into a single consistent framework. Agricultural productivity depends on rainfall and consequently deforestation decisions depend on rainfall. In turn, rainfall depends on deforestation decisions. Integrating both models into a single model consists in properly defining and solving a fixed point problem that finds the land use scenario consistent with the climate generated by it. I derive conditions for this fixed point problem to be a contraction. Under these conditions, I develop an algorithm that, starting from any scenario of land use, finds a convergence path of land use to the unique equilibrium. In this setting, the production function of agricultural products has an endogenous total factor productivity (TFP) that depends on the land use of every agent via climate feedback.

As an application, I study the effect that the deforestation of the Amazon Rainforest has on the agricultural production of the Brazilian state of Mato Grosso, one of the most important agricultural hubs in the world. I estimate the climate model using three-dimensional pixel level data for the entire Amazon. I estimate the land use model with a rich pixel level data for Mato Grosso that allows me to identify what agricultural product each pixel is producing. With the estimated model, I generate policy changes to study counterfactual land use allocations. In this paper, I focus on a particular policy relevant counterfactual concerning the protection of Indigenous Territories. The demarcation of Indigenous Territories was an important part of a series of policies that proved to be effective in combating deforestation in the early 2000s (Burgess et al., 2019; Assunção et al., 2013). Nonetheless, this

2

demarcation process suffers continuous pressure for rollbacks (Golden Kroner et al., 2019).

As I focus on the agricultural sector located in Mato Grosso, I present a counterfactual where the policy change is to allow agricultural production inside the Indigenous Territories of the Xingu River Basin, hereby Xingu (see Figure 2B). The Xingu is located close to the most productive regions in the state, which make extensive use of double cropping, that is, two harvest seasons in the same agricultural year[1]. The adoption of double cropping relies on a steady and abundant rainfall supply and therefore is a choice prone to be affected by changes in climate.

How deforestation affects rainfall is determined by the atmospheric circulation built in the climate model. In the Amazon region, wind moves from the ocean to the Xingu and then to the most productive regions in Mato Grosso. Thus, the Xingu has a direct impact on double cropping ability, supplying water through the transpiration-rainfall mechanism.

In the counterfactual scenario where farmers are allowed into the Xingu, deforestation would amount to more than 43% of the region. In the rest of the state of Mato Grosso, precipitation would decrease on average by 2% with some regions losing up to 4%. The returns of crops production would decrease by 2% with some regions losing up to 8%. Tail climate events increases this loss of crops in some regions to 12%.

The relevance of the deforestation-rainfall mechanism, and thus of this paper, goes beyond the Amazon and the agricultural sector. Deforestation in the Amazon affects rainfall in Paraguay, Argentina, and the Brazilian states of South and South-East regions, located thousands of kilometers away from the forest (Spracklen et al., 2012). This deforestation-rainfall mechanism is a characteristic of tropical forests, and thus is also present in the Congo and the Indonesian Rainforests. This paper uncovers an additional source of variation in rainfall which can be composed with a large literature studying the impact of changes in rainfall in a variety of topics, such as, energy generation by hydropower plants (Araujo, 2022; Barrios et al., 2010), water supply for human consumption and infant mortality (Rocha and Soares, 2015), civil conflict (Anderson et al., 2021; Harari and Ferrara, 2018), and migration (Albert et al., 2021).

This paper also contributes to the literature of deforestation in general (Burgess et al., 2012; Souza-Rodrigues, 2019; Asher et al., 2020). Government policies to curb deforestation (Burgess et al., 2019; Assunção et al., 2019) can be more effectively designed once externality effects can be better accounted for, especially on credit market policies (Assunção et al., 2020) that can benefit from having a better assessment of climate risks. On the effects of changes in agricultural productivity on deforestation (Jayachandran, 2021), the debate of

---

[1]The popularization of this technology has been possible after the introduction of genetic modified soybeans seeds, which has been explored in other papers (Bustos et al., 2016; Dias et al., 2020)

the Boserup Hypothesis – greater productivity leads to more deforestation (Heß et al., 2021) – versus the Borlaug Hypothesis – greater productivity leads to less deforestation (Abman and Carney, 2020; Assunçao et al., 2017) – has not considered that deforestation affects agricultural productivity.

Finally, this paper adds to the literature of discrete choice models studying land use and deforestation. Domínguez-Iino (2021) builds a model of land use for Brazil and Argentina to investigate the impact of trade policies on deforestation; Souza-Rodrigues (2019) estimates a demand for deforestation in the Brazilian Amazon; Araujo et al. (2020) studies the long-run efficient forestation – from a carbon sink perspective – in the Brazilian Amazon; Hsiao (2021) uses a dynamic model to study commitment issues and effects of trade on Indonesia's deforestation. This paper departs from those by accounting for the ecosystem service of rainfall provided by tropical forests.

The remainder of this article is organized as follows: in Section 2 I present the land use and climate models and discuss how to integrate and estimate both; in Section 3 I present an application for the model and the data; in Section 4 I show the estimated results; in Section 5 I discuss how to run a counterfactual and show the results of the proposed counterfactual of allowing deforestation in Xingu; in Section 6 I discuss caveats; in Section 7 I conclude.

2 Model

This section is divided into three parts. First, I develop a discrete choice model where farmers choose what to produce in their plot of land. When making decisions, farmers take precipitation as given. Second, I describe the empirical climate model that links land use with precipitation patterns. Third, I discuss how to integrate both models.

2.1 Land use depends on precipitation: a discrete choice model

A farmer owns a plot of land i and chooses land use from a discrete choice set. The farmer has the option of keeping the plot of land as forest or to deforest it and convert its use to agricultural activities. Formally, denote the choice of keeping the land as forest by F and using the land for agriculture as $c \in C$, where c denotes the choice among the possibility set C of agricultural activities.

By choosing an agricultural activity c, the farmer can transport its product to a port and sell it at the international market price, receiving as return[2]:

---

[2]This functional form can be motivated by the following return function $(r_i^c)^\alpha \exp\left[\beta_c^1 R_i + k_c + \epsilon_i\right]$. Here, precipitation affects soil productivity in a multiplicative way. Taking logs we arrive at $\tilde{\pi}_i^c$.

$$\tilde{\pi}_i^c = \alpha \log r_i^c + \beta_c^1 R_i + k_c + \varepsilon_i^c \tag{1}$$

Where, α is a normalization constant for the variance of the idiosyncratic shock; r_i^c is a measure of potential revenue of the agricultural activity c in land i; k_c is a constant that captures associated costs of production, such as wages, fertilizers, and land maintenance; ε_i^c is an idiosyncratic shock; R_i is a measure of precipitation on land i. With this specification, I can use farmers' choices as a way to recover the underlying crop yield response to water. In the literature, this is usually done by modelling a direct relation of crop yield to climate variables. These models depend on a variety of factors, such as potential and observed evapotranspiration, wind, and solar radiation. Nonetheless, as I am relying on farmers' choices, what matters is the variables that farmers are observing and accounting for when making decisions. More than a production function, Expression 1 can be interpreted as a perceived production function, as perceived by the farmers at the time of deciding the plot production.

The return that a farmer receives by keeping his plot of land as forest, is described by a constant:

$$\tilde{\pi}_i^F = k_F + \varepsilon_i^F \tag{2}$$

Where k_F is the constant and ε_i^F is an idiosyncratic shock. As in Domínguez-Iino (2021), I model the farmer's discrete choice problem as a nested logit. I create a nest for the forest choice and a nest for agricultural activities. The idiosyncratic shocks within agricultural activities are correlated, while the idiosyncratic shocks between forest and agricultural activities are not. In assumption 1 I formalize this and specify the joint distribution of the shocks to be an extreme value distribution.

Assumption 1 *The vector of land unobservables $\{\varepsilon_i^F, \{\varepsilon_i^c\}_{c \in C}\}$ has the cumulative distribution given by*

$$\exp\left[-e^{-\varepsilon_i^F} - \left(\sum_{c \in C} e^{-\frac{\varepsilon_i^c}{\lambda}}\right)^\lambda\right]$$

With the distribution of Assumption 1, the λ parameter measures the degree of correlation between the shocks in the nest of agricultural activities. It will be useful to denote by π_i^c and π_i^F Expressions 1 and 2 free of the idiosyncratic shock. Then, with Assumption 1, the conditional probability of choosing land use $c \in C$ in land i is given by:

$$P_i(c) = \frac{e^{\frac{\pi_i^c}{\lambda}} \left(\sum_{c \in C} e^{\frac{\pi_i^c}{\lambda}} \right)^{\lambda-1}}{e^{\pi_i^F} + \left(\sum_{c \in C} e^{\frac{\pi_i^c}{\lambda}} \right)^{\lambda}} \tag{3}$$

The conditional probability of choosing an element $c \in C$ given that C was chosen is given by

$$P_i(c|C) = \frac{e^{\frac{\pi_i^c}{\lambda}}}{\sum_{c \in C} e^{\frac{\pi_i^c}{\lambda}}} \tag{4}$$

And the conditional probability of choosing to keep the land as forest is given by:

$$P_i(F) = \frac{e^{\pi_i^F}}{e^{\pi_i^F} + \left(\sum_{c \in C} e^{\frac{\pi_i^c}{\lambda}} \right)^{\lambda}} \tag{5}$$

Notice that $P_i(F)$, $P_i(c)$, and $P_i(c|C)$ are functions of precipitation, that is, the land use decisions take as given the precipitation. I now turn to the other side of the problem, a model of how precipitation is affected by deforestation.

2.2 Precipitation depends on land use: an empirical climate model

The seminal work of Salati et al. (1979) opened a series of studies showing the importance of tropical forests in regulating climate systems. Among many important mechanisms through which tropical forests affect climate, an important one is its role in the water cycle. The Sun's energy evaporates the water in the oceans and creates humid air parcels that are then transported to continental lands. Along its trajectory the air parcel loses humidity via precipitation. But, on the ground, trees' transpiration recharges the air humidity contributing to rainfall downwind. Thus, trees affect rainfall. Interestingly, tropical forests can affect rainfall in locations that are hundreds and even thousand of kilometers away from the forest, creating the phenomenon know as "flying rivers" (Marengo et al., 2004; Nobre, 2014). An implication of this mechanism is that deforestation of tropical forests decreases rainfall.

The empirical climate model of how land use affects precipitation is based on Spracklen et al. (2012). To model the causal effect of tropical forests on the precipitation of a location, three steps are needed: **First**, I need to build a back trajectory of atmospheric transport for that location. That is, I need to see the trajectory of an air parcel from the Ocean to the location of interest. The dashed black lines in Figures 1A and B show two examples of back trajectories that arrive at a location of interest (the white cell). The same way that

6

Figure 1: Elements of the Climate Model
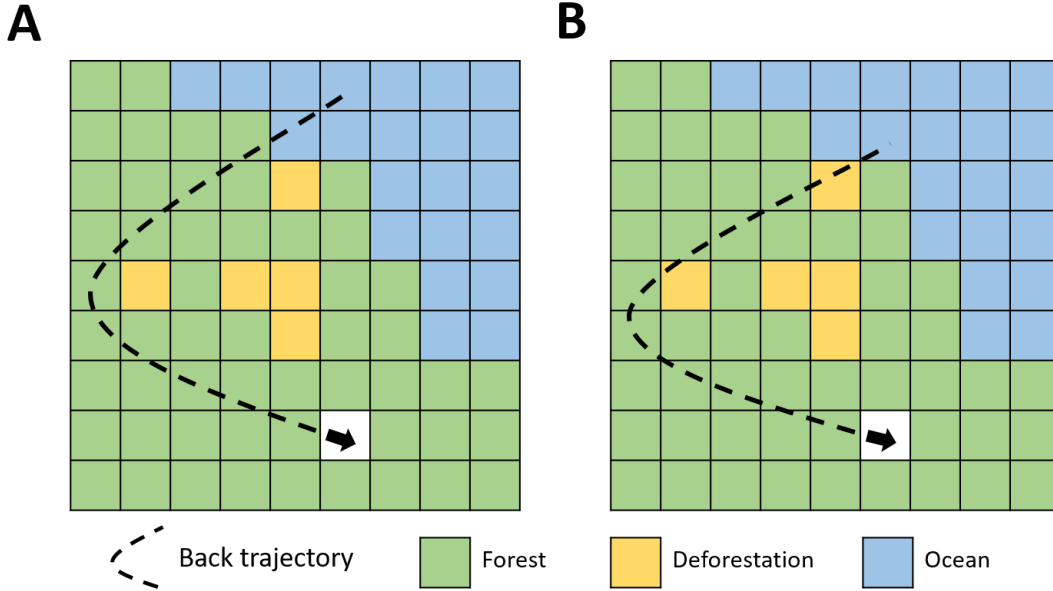
This Figure illustrates the steps needed to build the climate model. The dashed black lines represent back trajectories of atmospheric transport, starting in the white cell and walking back until the Ocean is reached. Along each trajectory, different types of terrain are covered. In A, the back trajectory is exposed to 14 pixels of forest, while in B the trajectory is exposed to 10 pixels of forest. That way, it is likely that A will deliver more rainfall for the white cell than B.

one can follow a river upstream to find its source, it is possible to follow wind patterns to find the path that the air parcel took from the ocean. Or yet, the same way that there exists a geometry and a flow direction for a surface river, there exists a pattern in the trajectories that air parcels take from the Ocean to continental lands. By building such trajectories I can determine what type of terrain an air parcel has flown over since its arrival in continental land. **Second**, for the now built back trajectory I count the number of tropical forest pixels that the air parcel has "flown" over, for example, the back trajectory in 1A was exposed to 14 pixels of forest while the back trajectory in 1B was exposed to 10 pixels of forest. That way, it is likely that the scenario in 1A will deliver more rainfall for the location of interest than the scenario in 1B; **Third**, I connect the rainfall in the location of interest (the downwind rainfall) with the count of forest pixels along the back trajectory (the upwind exposure to the forest) with a linear model.

Formally, the three steps are described below.

First, given a plot of land i in a time period q, I wish to compute the most likely trajectory of the air parcel over land i in the last Δ periods of time. Let $W_{jq} = \{x_{jq}, y_{jq}, z_{jq}\}$ denote latitude, longitude and pressure level (altitude) of an air parcel j in time q and let

V_{jq} denote the gradient of W_{jt}, that is, the air velocity vector. Then this air parcel can be traced back one step by

$$W_{j,q-1} = W_{j,q} - V_{j,q} \tag{6}$$

Across different points in space the wind changes speed and direction, which defines how a particle will move in this wind surface. I can then compound expression 6 to build a back trajectory for an arbitrary time window. Given a time window of Δ, the compounded back trajectory is given by

$$W_{j,q-\Delta} = W_{j,q} - \sum_{s=1}^{\Delta} V_{j,q-s} \tag{7}$$

Given a initial state $W_{j,0}$ and a time window Δ, I can compute a back trajectory. A plot of land i already defines a latitude and a longitude (x_{jt}, y_{jt}). By setting a pressure level (z_{it}) and a time window (Δ), I can compute the back trajectory of an air parcel over land i. For now, take the initial pressure level and the time window as given.

Second, once the back trajectory is built, I need to connect it with what is on the ground. Here, I already anticipate a practical difficulty: climate data are usually available in lower resolution than land use data. In order for me to be able to connect both types of data, some data aggregation will always be needed. Thus, define a climate region (or a climate pixel) by o, which is a set of lands and let $o(i)$ denote the climate region that contains land i; also let B_o denote the back trajectory from a climate pixel o, which is a set that contains all climate pixels that the air parcel over the climate pixel o has flied over in the previous Δ time periods; finally, let h_o be a measure of forest cover in climate pixel o. Then the upwind exposure to the forest of a back trajectory is defined as:

$$H_o = \sum_{j \in B_o} h_j \tag{8}$$

Referring again to Figure 1, along each back trajectory, the upwind exposure to the forest (H_o) is the sum of forest pixels.

Third, in the last step I model precipitation over climate pixel o as a function of the upwind exposure to the forest, parametrized by:

$$R_{om} = \theta_m H_{om} + \epsilon_{om} \tag{9}$$

Where m denotes seasons for which I want to allow heterogeneous effect of upwind expo-

8

sure to the forest on rainfall, for example, dry-wet seasons or months; R_{om} denotes precipitation in climate pixel o season m. The marginal effect of deforestation on precipitation is given by θ_m, if the deforestation happens to be on the path of the back trajectory, otherwise the effect is zero.

2.3 Land use and precipitation: equilibrium

Land use depends on precipitation and precipitation depends on land use. Thus, integrating the land use model and the climate model amounts to solve a fixed point problem, that is, finding a land use that is consistent with the precipitation generated by it.

To translate the land use decision of the discrete choice model to the deforestation scenario of the climate model, let $P(F)$ denote the vector of probabilities for all plots of land where each entry is given by $P_i(F)$, the probability of a farmer choosing forest (Expression 5). Take the measure of upwind exposure to forest to be such that $h_o = \frac{\sum_{i \in o} P_i(F)}{\sum_{i \in o} 1}$, the proportion of forest land inside climate region o. Therefore, the rainfall of climate region o, given in Expression 9, depends on the probability $P(F)$. A last translation step is to set the total precipitation of land i as the sum, across seasons of interest M, of the precipitation on the climate region $o(i)$. The set M gives flexibility to choose which seasonal precipitation is relevant for each agricultural activity, for example, the growing season of an specific crop. Then I set $R_i = \sum_{m \in M} R_{o(i)m}$.

Thus, the chain of dependence is that $P(F)$ depends on R_{om} which depends on $P(F)$. Formally, I can define a map $T : P(F) \rightarrow P(F)$. In Appendix B, I show the conditions needed for this map T to be a contraction, which are present in my application. Therefore, starting from any land use, I can iterate T until convergence to find the unique equilibrium that is consistent with both the discrete choice model and the empirical climate model. With that description of equilibrium, the production function of agricultural products is such that their total factor productivity (TFP) is endogenous, depending on the land use configuration of the entire economy via the climate feedback effect of land use and rainfall.

2.4 Estimation

Land use model. I estimate the discrete choice model by maximum likelihood in two stages (Train, 2009): first, I estimate the agricultural (lower) nest; second, using the estimated parameters of the lower nest, I build the expected value of the agricultural nest to estimate the deforestation (upper) nest. To estimate the nest of agricultural activities I solve the following maximization problem:

$$\max_{\left\{\frac{\alpha}{\lambda}, \frac{\beta_c}{\lambda}, \frac{k_c}{\lambda}\right\}} \sum_i \mathbb{I}_c \log P_i(c|C) \tag{10}$$

Where $\mathbb{I}_c$ denotes an indicator variable that equals one when c is chosen and zero otherwise. This estimation step leverages rainfall variation in the cross-section and across time. As noted in Mérel and Gammans (2021), the cross-section variation of rainfall can capture long-run effects. This is exactly what is of interest, since my counterfactual changes the climate of the region and is likely to affect land use decisions in the long run. With the estimated parameters of the agricultural nest, obtained from the optimization of Expression 10, I can estimate the upper nest by solving:

$$\max_{\{k_F, \lambda\}} = \sum_i \mathbb{I}_F \log P_i(F) + (1 - \mathbb{I}_F) \log P_i(C) \tag{11}$$

Where $\mathbb{I}_F$ denotes an indicator variable that equals one when F is chosen – the farmer chooses to keep his plot of land as forest – and zero otherwise.

Climate model. For the empirical climate model in expression 9, I use a Fixed Effects estimator. Different sets of fixed effects are tested to reduce concerns of endogeneity. More specifically, I tested fixed effects on: month; month and year; month, year, and pixel; month-pixel and year. I also control for the length of the back trajectory traveled by land, since it can mechanically change the measure of upwind exposure to the forest.

The identification hypothesis of the model comes from the exogeneity of the back trajectory paths, which are determined by exogenous wind patterns[3]. Furthermore, the mechanisms behind the link between transpiration and humidity transport (Spracklen et al., 2012) provide further evidence of the causal interpretation of Expression 9.

3 Application and Data

3.1 Tailoring the model

In this paper I focus on the effect that the conservation of Indigenous Territories in the Amazon has on the agricultural sector of the Brazilian state of Mato Grosso. Figure 2A

---

[3]Atmospheric transport is determined mainly by temperature differences between the Equator and the Poles and the rotation of the Earth (the Coriolis effect). There is a study on how the rainforest can affect atmospheric transport: The Biotic Pump Theory (Makarieva and Gorshkov, 2007) approaches this question from a theoretical perspective. As far as I know, this theory is not a consensus because it lacks empirical evidence. Nonetheless, the Biotic Pump Theory would apply in a really large deforestation context, and therefore, even if it were true, it is unlikely to have an effect on my setting.

shows the location of the Rainforest and of the state of Mato Grosso.

Mato Grosso is one of the most important agricultural hubs in the world, being responsible for 10% of the global production of soybeans. This was made possible by the introduction of new agricultural techniques that expanded the application of double cropping (Xu et al., 2021). Double cropping is when a farm has two harvest seasons in the same agricultural year. In the case of Mato Grosso, soybeans followed by a mid-season corn. The other important agricultural activity in the state is cattle ranching, corresponding to almost 40% of the land use. I work with data from 2008 to 2017, a stable period of regulatory framework (see Burgess et al. (2019)). To estimate the model I exclude observations inside protected areas, e.g., conservation units and indigenous land. These protected areas are under a different regulatory framework than other regions and have significant less deforestation, Figure 2B show the locations of protected areas. With this selection of locations, the estimation of the land use model results in parameters that govern the decision of agents on private land. Later, on my counterfactual exercise, I allow agents to act on indigenous territories as if they were in private land, effectively rolling back the protection status of these territories.

For this application, the agricultural activities set (C) includes growing cash crops and using land for pasture grazing. I define $C = \{\text{double}, \text{single}, \text{pasture}\}$ for the choices of double cropping (soybeans and corn), single cropping (soybeans), or pasture grazing. The revenue value of Expression 1 reflects the double cropping possibility with,

$$
r_i^c = \begin{cases} A_i^c(p^c - \tau_i^c) & \text{if c is single crop or pasture} \\ A_i^{c'}(p^{c'} - \tau_i^{c'}) + A_i^{c''}(p^{c''} - \tau_i^{c''}) & \text{if c is double crop w/ crops c' and c''} \end{cases} \tag{12}
$$

For crops, A_i^c is a measure of maximum attainable yield for producing crop c in land i; for pasture grazing, A_i^c is a measure of potential meat production; p^c denotes international market price for crop c or meat; τ_i^c is transportation cost of the product of c from land i to the port. Finally, R_i is taken to be the total precipitation during the growing season in Mato Grosso – September through March – on land i.

An important constraint of discrete choice models is the normalization of constants and coefficients of variables that are equal across choices. The coefficients on precipitation (β) and the constants (k) can always be rewritten to appear in differences in the conditional probability expressions, which makes it impossible to identify all of them. I then set both k_{pasture} and β_{pasture} to zero.

The normalization of one of the constants (k) is inevitable and the choice of which one is normalized does not affect the estimation nor the counterfactual. The choice of setting

β_{pasture} equal to zero, on the other hand, can affect the estimation of the upper nest and the counterfactual, since I am already considering that the return of the forest does not depend on rainfall (Expression 2). Nonetheless, in order not to make use of this restriction, I would require a separate model of cattle ranching, possibly a dynamic one, that would account for changes in water availability on livestock growth, reproduction, and productivity. In my data, simply estimating pasture returns on rainfall has yielded non significant correlations in a variety of specifications. There are a number of explanations for a zero effect of rainfall on pasture returns, for example, livestock can be more resistant to changes in rainfall than crops and ranchers can resort to confinement and changes in livestock diet (Skidmore, 2022). Since the focus of my paper is the effect on crops, I believe this restriction is a reasonable simplification.

3.2 Data

To estimate the model, I need data on land use, crop prices, transportation cost, precipitation, soil suitability, and I need to build the back trajectories of atmospheric transport.

Land use

Land use data is from Simoes et al. (2020). The data classifies annually land use from 2001 to 2017 for pixels of 250 meters resolution for the Brazilian state of Mato Grosso.

Importantly, to build this land use data, information throughout the year is used, which allows it to identify double cropping. That is, it identifies whether two crops were produced in the same pixel for the same agricultural year. For Mato Grosso a double cropping consists of soybeans followed by corn. Furthermore, the data identifies single cropping (soybeans), forest, and pasture areas. [4].

Table 1 shows the evolution of land use in Mato Grosso across the years. Throughout the years the adoption of double cropping increased. In 2017 the double cropping planted area

---

[4]I collapse some of the land uses present in the data. Soybeans and Millet is collapsed into single cropping, since Millet is a intercrop option which is generally not sold, but instead, used to avoid leaving the soil bare. I also collapse Cotton as single cropping, since it is a very small proportion of crops. Finally I drop observations of urban areas and water. Other double cropping systems in the data are soybeans-cotton and soybeans-sunflower. Nonetheless, those are a very small proportion ($< 1\%$) of the data. That way I just collapse them to double cropping together with soybeans-corn. I drop the land use classification of sugarcane. Sugarcane is a very small proportion of the planted area in the data and also a small proportion among crops. Furthermore, it is exclusively concentrated in the southeast region of Mato Grosso. Also, sugarcane is the only perennial crop in the data, which could, in principle, justify the necessity of a dynamic model. For an in depth study of the economics and dynamics of sugarcane see Sant'Anna (2021). Finally, I do not make a distinction among wetland, rainforest, cerrado and secondary vegetation for the choice set, that is, all three categories is defined by the choice F.

Table 1: Land Use Over Time

Land use	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Forest	583.7	600.3	588.3	571.6	583.4	585.3	587.5	594.6	572.3	571.2
Pasture	267.2	248.8	262.0	269.9	257.3	247.6	241.1	229.0	255.6	254.1
Double Crop	28.2	29.1	32.3	33.4	45.9	52.0	55.7	50.8	51.8	63.7
Single Crop	32.0	33.8	31.0	37.0	24.6	28.3	28.8	38.5	34.0	24.6

This table presents the total area (in 1,000 km^2) for each possible land use across the years. Pasture grazing is the predominant agricultural land use. Also noticed that double cropping became the predominant way of producing cash crops in 2010.

of soybeans and corn was greater than the single cropping of soybeans, contrary to what was seen in 2008. The pasture grazing activity is the predominant choice among agricultural activities and is the main driver of deforestation.

Transportation cost

Transportation cost is computed as in Araujo et al. (2020). It combines: (1) a transportation network that includes the location of paved and unpaved federal and state roads, waterways, and ports; (2) freight data on transportation of corn, soybeans and meat. This data is used in a Non-Linear-Least-Squares estimation, as in Donaldson (2018), to build an application where I can compute the transportation cost between every pixel in my data and the most cost-effective port with access to international markets. Table 2 shows descriptive statistics for the transportation cost. I further describe this data in Appendix A.

Agricultural prices

International prices of different crops and meat are from the Federal Reserve Economic Data (FRED) [5]. I convert the prices data to Brazilian reais - using data of exchange rate from FRED - and deflate them using the official consumer index inflation in Brazil from the The Brazilian Institute of Geography and Statistics (IBGE). Table 2 shows descriptive statistics for prices.

---

[5]Available at www.fred.stlouisfed.org

Figure 2: Atmospheric Transport, Xingu and Mato Grosso
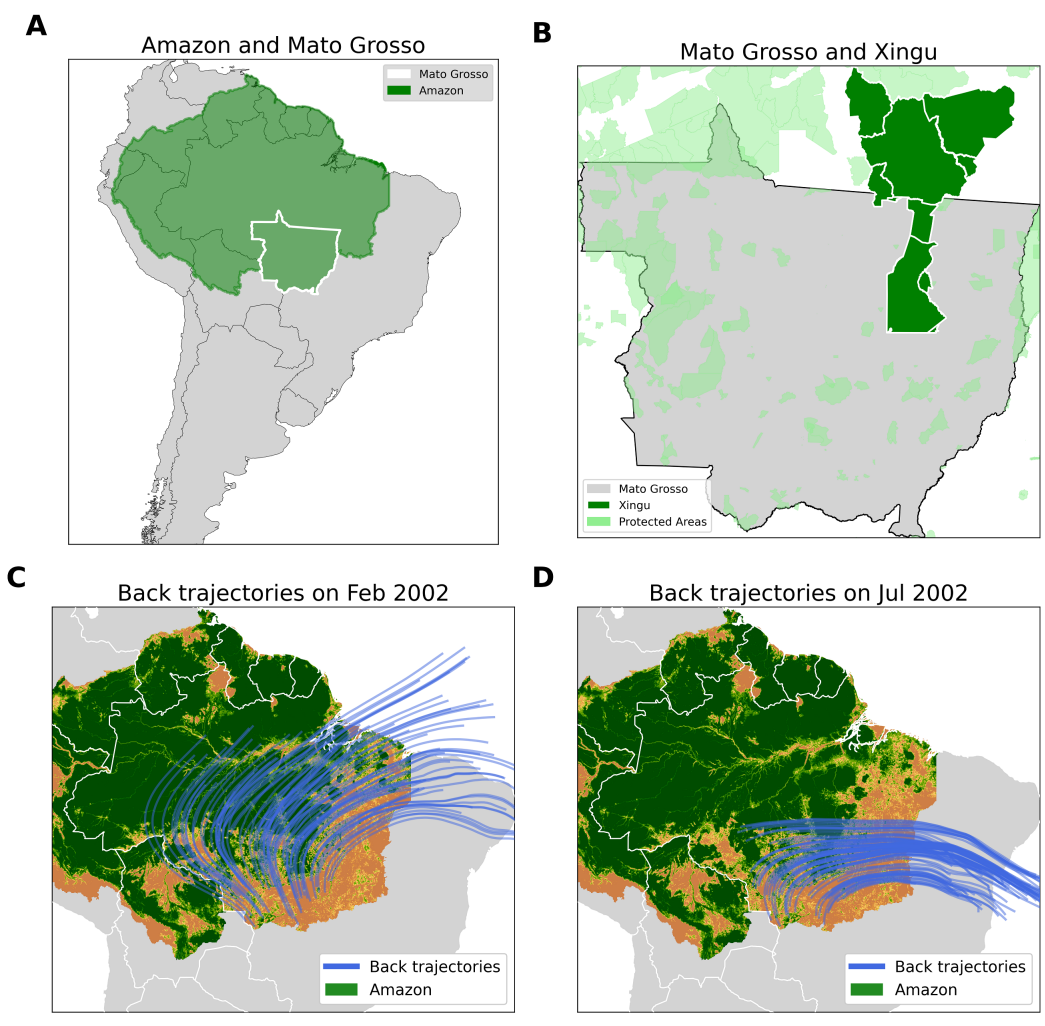
(A) shows the location of the Amazon Rainforest and of the Brazilian State of Mato Grosso. (B) shows the location of the indigenous territories of the Xingu River Basin and the locations of other protected areas in the State of Mato Grosso. I exclude the observations inside these protected areas when estimating the discrete choice model; (C) and (D) show a sample of back trajectories for the months of February and July 2002 arriving in Mato Grosso.

Table 2: Descriptive Statistics

	Soybeans	Corn	Meat
Yield (ton/ha)	2.817	3.337	0.054
	(0.478)	(1.166)	(0.039)
Transportation cost (R\$/ton)	137.84	138.34	235.55
	(18.97)	(18.6)	(24.93)
Price (R\$/ton)	754.74	366.9	5629.23
	(78.48)	(60.34)	(516.46)

This table shows descriptive statistics for: soil suitability for corn, soybeans, and meat. These are measured as tons per hectare; transportation cost for each product, measured in Brazilian reais as of 2008 per ton; price of each product in Brazilian reais as of 2008 per ton.

Soil suitability

Maximum attainable yield for each crop (A_o^c) is from the Food and Agriculture Organization's (FAO) project Global-Agroecologial Zones[6]. This data gives estimation of crop yields in kilograms per hectare at approximately 10km resolution. Those potential yields are given for a variety of scenarios, differing according to quantity of available inputs and water source. I use the yields of high inputs, which FAO describes as the market-oriented agriculture production, and rainfed cultivation, because this is the predominant form of production in the Brazilian Amazon. This data set has been extensively used in economics (see, for example, Costinot et al. (2016); Sotelo (2020); Costinot and Donaldson (2012)). To measure potential pasture suitability, I use a grass potential yield, also from FAO. To convert grass production into meat, I transform the pasture suitability to fit the average grass yield of 54kg/ha of meat – which is the average production in the state of Mato Grosso as given by the Agricultural Census of 2010 – and the maximum yield of 270kg/ha (Magalhães et al., 2019).[7] Table 2 shows descriptive statistics of the soil suitability data.

Rainfall

Total monthly rainfall data by region is from Copernicus (2017) and Funk et al. (2015). I use two different rainfall data sets because the climate model has lower resolution than the land use model. As I explain in the next subsection, the back trajectories are built with

---

[6]Available at www.fao.org/nr/gaez/en

[7]In this step I simply raise the pasture suitability to the power of 2.01 and normalize it by the mean yield of 54kg/ha.

data from Copernicus (2017), therefore it is helpful to keep consistency with data from the same source. Rainfall data from Funk et al. (2015) has higher resolution and is thus more suitable to the land use model. Nevertheless, both data are consistent with each other, in terms of magnitudes of rainfall. From Funk et al. (2015) I collect total precipitation data for the months of the growing season – September through March – for the agricultural years of 2008 to 2017. From Copernicus (2017) I collect total precipitation data by month for the years of 1985 to 2018. Figure 3 shows descriptive statistics for the region in different months for the period of 2008-2017. This figure illustrates significant variation of rainfall across months and space.

Building the Back Trajectories of Atmospheric Transport

To build the back trajectory of atmospheric transport I use monthly wind data from Copernicus (2017) from 1985 to 2018. This gives a velocity vector (u-component, v-component and vertical velocity) across time, latitude, longitude and pressures levels (altitude). As I am interested in modeling the causal impact of land use on monthly precipitation, I depart from Spracklen et al. (2012) by using the prevailing winds of each month-year as my data for the velocity vector, instead of hourly data. The resolution of the data is kept at the native resolution of 0.25° (approximately 25 km) instead of using the resolution of 1° as in Spracklen et al. (2012). Given a month, a year, and a starting point of interest, the back trajectory is computed using Expression 7. I set the initial pressure level to 800hPa and the time window of the back trajectory to 5 days, as in Spracklen et al. (2012).

The back trajectories are computed for end points in the state of Mato Grosso, since this is the region in which I want to model the rainfall effect. The back trajectories themselves can span all across the South America region. Figure 2 shows a sample of different back trajectories across different months. The variation of trajectories across months is a characteristic of the wet and dry seasons, with wind on the wet season being more exposed to the forest.

Upwind exposure to the forest

Spracklen et al. (2012) uses the Leaf Area Index (LAI) as the forestation measure, my h_o variable. I cannot use the same data for two reasons. First, there is not a way to assign a LAI to a specific land cover, since, conditional on the land cover, the LAI index may vary. Second, the LAI varies across months, and I only have annual land use data. That way, to perform a counterfactual I would need a method to assign, for each annual land cover, a LAI

measure across different months, which, for the best of my knowledge is not available.[8]

To circumvent this problem I will use only a binary forest data as a more stable proxy for vegetation. This measure has the advantage of being consistent with the outputs of the land use model. I will set my forest cover measure h_o on climate region o to be equal the proportion of pixels in o classified as forest. The empirical climate model in Equation 9 has a different coefficient for each month, which accounts for potential differences in vegetation across months.

Data on forest pixels are from Mapbiomas [9]. This data classifies land use for the entire Amazon from 1985 to 2018 at 30 meters resolution. I then build a map where pixels classified as forest receives the value of 1 and pixels classified as non forest receives zero and convert this map, with an average filter, to the same resolution of the Copernicus (2017) data to compute the h_o variable. Notice that even though I am using a binary variable of forest status (forest or deforested), the high resolution of the binary forest data gives much more granularity than available LAI data of 5km resolution (Lemieux and Luquire, 2021) [10]. Also notice that, even though the forest data is binary, the necessary aggregation procedure for the climate model generates a continuous variable of forestation status for the climate model.

In Figure 3 I show descriptive statistics of the upwind exposure to the forest. During the wet season this measure is higher, on average, and noisier than in the dry season. The figure also shows how the upwind exposure to the forest and rainfall are highly correlated.

4   Estimation Results

I will present the results of the discrete choice model and then the results of the empirical climate model. I leave the counterfactual for the next section.

4.1   Results for the Discrete Choice Model

The estimates of the model are in Table 3. The $\frac{\alpha}{\lambda}$ measures the degree of substitution within the agricultural nest. Comparing with the results of Domínguez-Iino (2021) for frontier region, I find a lower elasticity, that is a lower degree of substitution. This can be a char-

---

[8]See Figure A2 in the Appendix A. This figure shows that LAI varies across months and land use. Unfortunately the variation across land use is not sufficient to separately assign a LAI value for each land use, as evidenced by how the 95% confidence interval of the forest land cover intersects the mean values of LAI of the other two types of land cover. See also Myneni et al. (2007)

[9]Available at www.amazonia.mapbiomas.org/

[10]To better assess this difference: one pixel of 5km holds more than 27,000 pixels of 30m.

acteristic of the region, with strong persistence of double cropping adoption and pasture in some regions of Mato Grosso.

I find a deforestation elasticity (λ) higher than Domínguez-Iino (2021) finds for frontier region, probably driven by the fact that Mato Grosso is one of the most active agricultural frontiers in Brazil. The coefficient of the rainfall effect for double crop system (β_{double}) is more than twice the magnitude of the effect for single crop (β_{single}). That is, the choice of double cropping is more sensitive to rainfall than single cropping, as expected. Finally, the constant (k_{double}), which captures associated costs of double cropping, is at least 20% greater than the constant for single cropping (k_{single}).

4.2   Results for the Empirical Climate Model

Here I present the results of estimating Expression 9. Table A2 in the Appendix shows the results for different sets of fixed effects. As the unit of measurement of upwind exposure to the forest does not have an intuitive interpretation, I normalize it by their standard deviation.

I plot the results in Figure 3. The coefficients are all positive across months and different specifications of fixed effects, which means that upwind exposure to the forest along a trajectory of an air parcel increases precipitation in the destination region. In all models I include as control the length of the back trajectory traveled by land. The results are sizeable, for example, for the month of December, one standard deviation of the upwind exposure to the forest could generate an impact of 17% of the mean precipitation. Across different specifications there are some estimate variability. Nonetheless, for the period of the year that I am interested in – the growing season, which is September-March – the estimates are much more stable across specifications. Those results alone do not give an idea of the economic magnitude of these effects. It is the interaction of the effect of land use on precipitation and the effect of precipitation on land use that is of interest. I need a method to run counterfactuals, which is presented in the next section.

# 5   Counterfactual

5.1   Algorithm

In this section I discuss how to run a counterfactual with the model and assess the cost of land use policies. There is a range of policies that can be changed with the current specification of the model, such as a deforestation tax, a tax on an specific agricultural activity, the construction of a new road that alters transportation cost, or the expansion of

Table 3: Estimation Results

coefficient	estimate
Panel A - Agricultural Nest	
$\frac{\alpha}{\lambda}$	0.56
	(0.01)
$\frac{\beta_{singlecrop}}{\lambda}$	0.98
	(0.03)
$\frac{\beta_{doublecrop}}{\lambda}$	2.58
	(0.03)
$\frac{k_{singlecrop}}{\lambda}$	-3.68
	(0.02)
$\frac{k_{doublecrop}}{\lambda}$	-4.51
	(0.02)
Panel B - Deforestation Nest	
λ	0.85
	(0.03)
k_F	3.14
	(0.01)

This table shows estimation results of expressions 10 (Panel A) and 11 (Panel B). α is the normalization constant for the variance of the idiosyncratic shock measure and multiplies the measure of potential revenue; λ measures the degree of correlation between the shocks in the nest of agricultural activities; the β_c parameters measure the response of the return of land use c to rainfall; the k_c parameters captures associated costs and benefits of production or of keeping native vegetation. Standard errors in parentheses computed with block bootstrap clustered at the pixel level. All estimates have p-value < 0.01. Lower nest - number of observations: 3,185,987; number of unique pixels: 379,902. Upper nest number of observations: 7,265,252; number of unique pixels: 727,750.

Figure 3: Exposure to the Forest and Precipitation
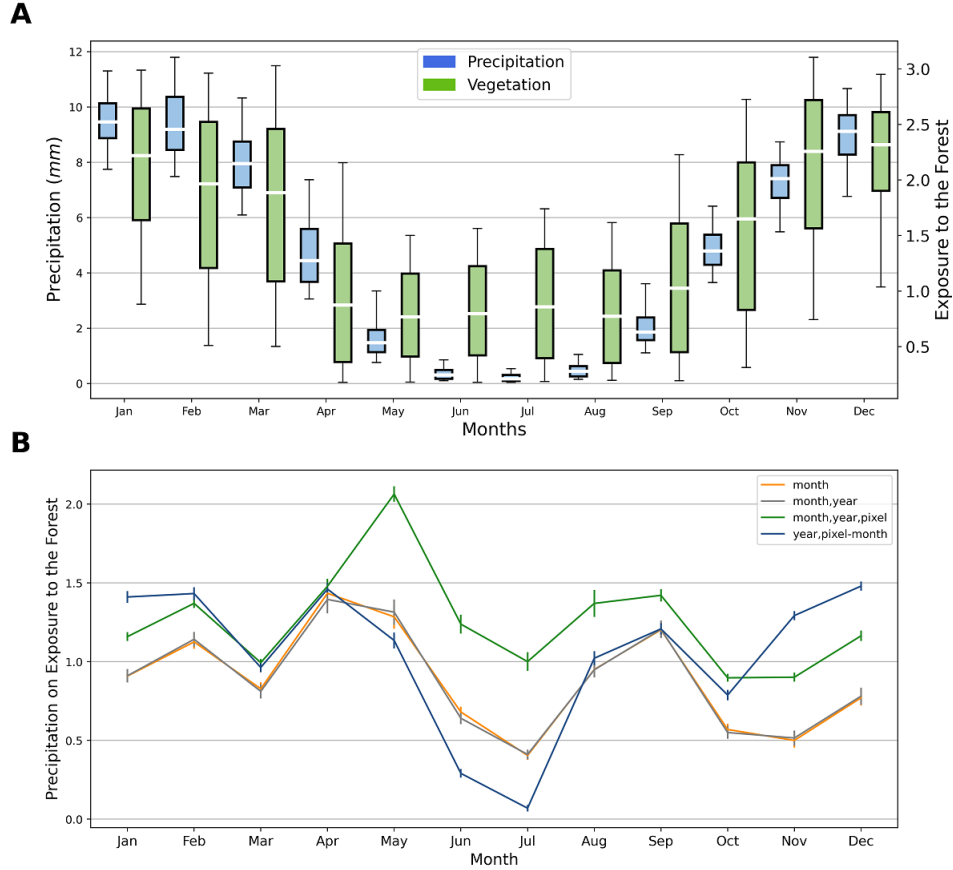
(A) the green boxes describe the upwind exposure to the forest data. The white lines inside the boxes mark the median value. The boxes boundaries and the caps shows the percentiles 5%, 25%, 75%, and 95%. The blue boxes describe the precipitation data.(B) this figure shows results for the empirical climate model 9. All specifications include as a control the length of the back trajectory traveled by land. Standard errors are clustered at the pixel level. Each line/color denotes a different specification for fixed effects.The small vertical lines show the 95% confidence interval of the estimate. Number of observations 487,560.

available land. The algorithm to generate counterfactuals presented in this section is general enough to accommodate all these possibilities. The important aspect is that the policy needs to be translated in terms of the return function of agricultural activities and/or of the forest.

In my application, I will generate a scenario where the Indigenous Territories of the Xingu River Basin – a protected area which today is almost completely covered by native vegetation, showed in Figure 2B – would be opened to deforestation and the expansion of the agricultural frontier. To do that, I apply the estimated coefficients to the Xingu region to generate a counterfactual deforestation scenario in the territory. I already introduced the notation of $P(F)$ as the vector the probabilities of choosing forest (a vector of $P_i(F)$). In a similar fashion, denote by R and H_m the vectors of rainfall (R_i) and upwind exposure to the forest (H_{om}). I then use the following algorithm:

1. Set counter t $= 0$

2. Set commodity prices and precipitation (R^t) to their temporal average.

3. With the data and estimated parameters compute the baseline probability of choosing forest ($P^t(F)$) – in my case it consists of using Expression 5 for pixels outside Xingu and setting $P_i(F) = 1$ if land i belongs to Xingu.

4. Generate a policy change – in my case, allow farmers into the Xingu region.

5. Compute $P^{t+1}(F)$ from Expression 5. If $P^t(F) = P^{t+1}(F)$ stop. Otherwise, continue.

6. Compute the change in upwind exposure to the forest (ΔH_m^t) caused by the difference between $P^{t+1}(F)$ and $P^t(F)$ using the computed back trajectories.

7. Update the precipitation to $R^{t+1} = R^t + \sum \theta_m \Delta H_m^t$

8. Update $t = t + 1$ and return to step (5).

In Step (3) I set all pixels inside Xingu to initialize as a forest. This is only because my counterfactual consists of expanding the region of possible agricultural production to an area that is covered in native vegetation. If the counterfactual does not involve allowing new pixels into the sample, no exogenous initial probabilities need to be set. In step (6), as I have multiple year data on atmospheric transport, I compute the change in upwind exposure to the forest using the wind data for all the years and then I take the average. The idea is that the 35 years of wind data is representative of the support of the distribution of wind patterns, and thus the average upwind exposure to the forest is the best predictor of expected future upwind exposure to the forest.

Notice that in this setting I do not have only the deforestation of the Xingu affecting rainfall downwind. At each iteration of the algorithm, farmers adjust their choices to the rainfall scenario. In the first iteration the Xingu deforestation will affect rainfall and land use decisions downwind. In the second iteration, downwind land use decisions (inside and outside Xingu) will affect rainfall, which will further affect land use decisions and so on until convergence.

5.2   Counterfactual Results

Allowing farmers into the Xingu region would result in the deforestation of 57,841 km^2 or 43% of the region. This deforestation would cause an average decrease in the precipitation of the growing season of 2%, with some regions losing as much as 4% of the historical average precipitation. Figure 4A shows the distribution of precipitation loss in the state, averaged by municipality for better visualization. The region west of the Xingu is the most affected because of the pattern of atmospheric transportation from east to west.

For the region outside Xingu, the consolidated regions of agricultural production, the expected return of crops – single and double cropping – decreases by 2%, with some regions losing up to 8% of the expected return. Figure 4B shows the distribution of losses in expected crop production averaged by municipality. This effect is heavily driven by changes in double cropping. Figure 5 shows the losses in expected return for single and double cropping. In general, double cropping has more than double the loss of the single crop option, since, as shown in Table 1, double cropping is 2.6 times more sensitive to changes in rainfall.

The expansion of deforestation inside the Xingu is driven by pasture grazing, responsible for 73% of the conversion of native vegetation. The implementation of double cropping is responsible for 16% and single cropping for 11% of the deforestation inside Xingu. Thus, for the agricultural sector, the distribution of gains and losses is such that farmers on Xingu region would produce at the cost of regions west of the Xingu.

As described in the algorithm section, I use the average upwind exposure to the forest – averaged across 35 years of data – to compute the counterfactual results. Thus, the counterfactual exercise so far considers an average case, with changes in rainfall interpreted as a change in its long run pattern.

Nevertheless, depending on the wind patterns of each specific year, farmers may be more exposed to the deforestation effect. If in one year more back trajectories are exposed to deforestation, then the gap between the realized and counterfactual rainfall will be larger. Exploring the year-to-year variation in the wind data, I can repeat the counterfactual exercise individually for the 35 years of data. Figure 6 shows the distribution of losses for double

## Figure 4: Change in Rainfall and Crop Production
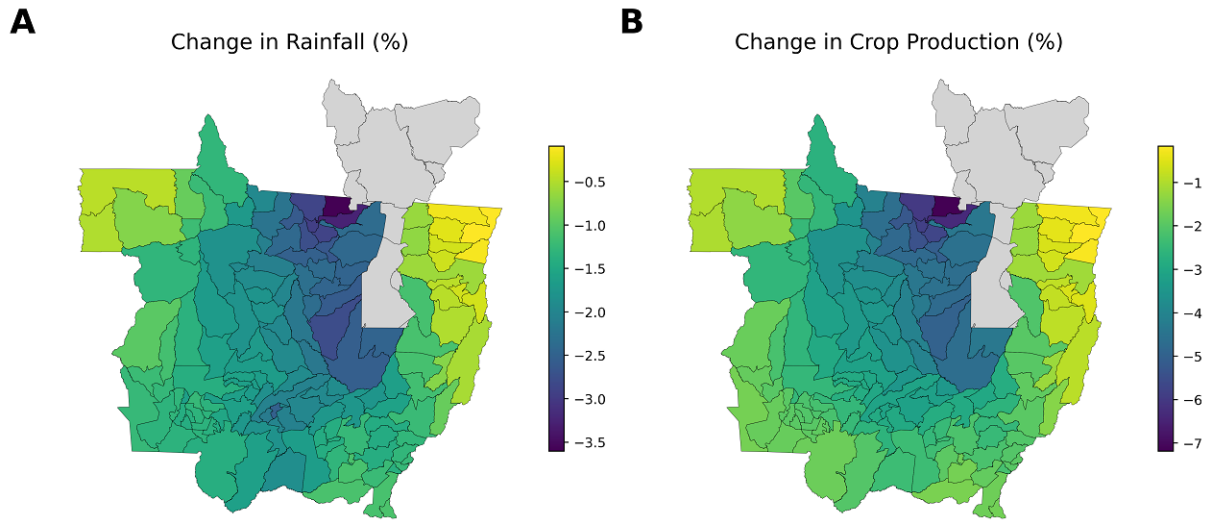
(A) Change in rainfall in the counterfactual scenario as a proportion of the baseline scenario, which is the historical average of rainfall between 2008-2017. (B) Change in crops expected returns due to the endogenous change in rainfall caused by the deforestation in the Xingu.

Figure 5: Change in Single and Double Cropping Production
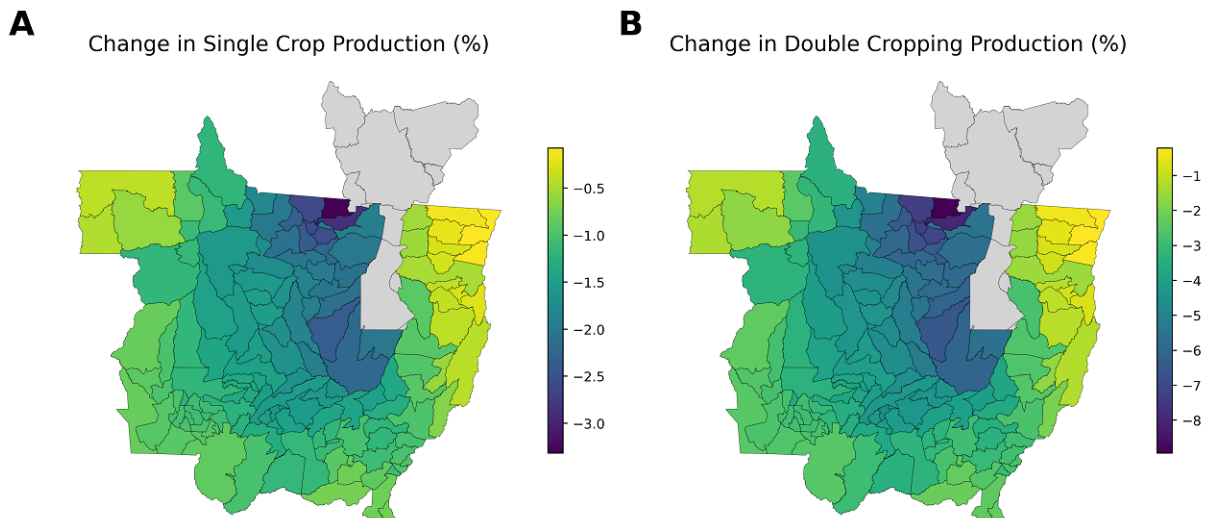
(A) Change in single crop expected return and (B) Change in double cropping expected return due to the endogenous change in rainfall caused by the deforestation in the Xingu

Figure 6: Distribution of Loss with Year-to-Year Variation
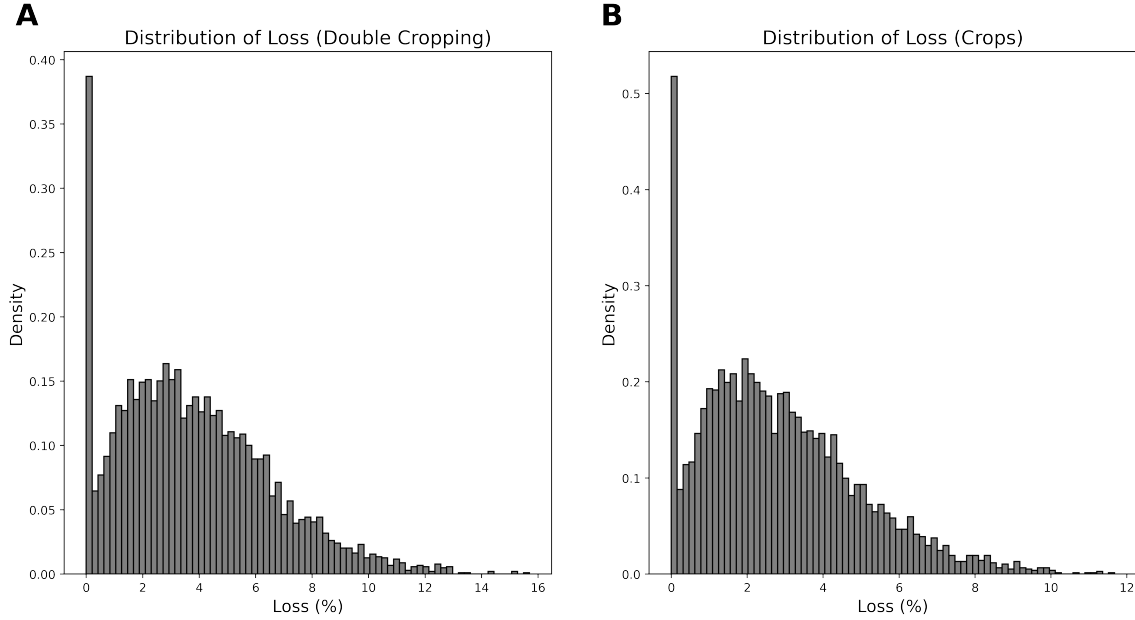
Distribution of loss (%) using year-to-year variation in the atmospheric trajectory data for double cropping (A) and crops in general (B).

cropping and crops. Double cropping has tail events of 15% of loss and crops in general has tail events of 12% of loss.

6   Caveats

In this section I discuss caveats related to assumptions, data limitations, and other relevant mechanisms of deforestation that is out of the scope of the paper.

Other effects of deforestation. First, my model does not capture other externalities of deforestation besides the precipitation mechanism and thus is not suitable to be used as a stand alone approach for cost-benefit analysis of deforestation. An important externality of deforestation is the release of greenhouse gases. The 57,841 km^2 of deforestation in the counterfactual Xingu region is equivalent to the emission of 2.8 billion tons of CO_2. With a conservative social cost of carbon of US\$5 ton/CO_2, this amounts to US\$ 14 billion of externality cost.[11] Another important limitation of my approach is that deforestation impacts water availability in a local scale by altering soil absorption capacity (Arias et al.,

---

[11]The value of 48,510 converts deforestation in km^2 to released tons of $CO2$ (Technical Note nº 2093/2018-MMA from the Amazon Fund). US\$ 5 ton/CO_2 is the carbon cost considered in the same note and used in the Amazon Fund.

2020). Although it is not likely to affect my results for the regions outside the Xingu, it can be a source of overestimation of the agricultural benefits inside the Xingu. Deforestation also affects biodiversity. The forest fragmentation in the Xingu in the counterfactual scenario could cause an abrupt decrease in biodviersity (Ferraz et al., 2003). Deforestation makes extensive use of fires and thus creates pollution that affects human health (Rangel and Vogl, 2019; Rocha and Sant'Anna, 2022). Deforestation is also connected with illegal activities, violence, and conflict (Chimeli and Soares, 2017; Fetzer and Marden, 2017). Finally, the Xingu deforestation will also affect agriculture and other sectors in other regions of South America.

Functional form of crop water response. Modelling crop yield response to water is a complex topic. The FAO model for crop yield prediction, for example, is a dynamic model which uses daily data to determine the effects of rainfall on yields (Steduto et al., 2012). Such model can generate abrupt non-linear responses in crop yields, which my model would not capture. I believe such models would be useful to quantify a large scale deforestation counterfactual, capable of generating a higher effect on rainfall than my application. Nonetheless, such models cannot be easily incorporated in equilibrium with discrete choice models and would require additional simplifying assumptions on the interaction of the land use model and the climate model in equilibrium. Still, in large scale deforestation scenarios, the linear model employed in this paper can be seen as am approximation of the crop response to water.

Benefits of cattle ranching. In my model, the constant parameter associated with each choice (the k parameter) can capture non-immediate benefits and costs of the activity. For the pasture grazing activity, the constant k_{pasture} was normalized to zero for identification, but in the model it affects the farmers decision. In the particular activity of pasture grazing, the k_{pasture} parameter can capture the fact that converting forest to pasture may be a cost-effective way of gaining and maintaining property rights over the land. That is, k_{pasture} can capture the benefit of reducing the probability of having a plot of land seized by the government or a third party and thus a significant part of pasture returns can be actually just paying for inefficiencies of the Brazilian weak institution of land titling. What this weak institutional framework may imply is that, if property rights were strengthened, than pasture activity could stop being used as a way to secure land rights and thus k_p would be smaller. Then, the overall effect on the agricultural sector of deforesting the Xingu could become more significant, since growing cash crops is the activity prominently affected by changes in rainfall.

7 Conclusion

In this paper I developed an integrated model of land use decisions and climate consequences with focus on changes in rainfall patterns. As an application, I estimated the climate model for the Amazon Rainforest and the land use model for the Brazilian state of Mato Grosso. Deforestation as the result of land use expansion causes a decrease in rainfall which in turn affects land use decision. An integrated model allows me to compute policy counterfactuals in an equilibrium setting, where the land use choices and the climate are consistent with each other. I then explore a counterfactual where farmers are allowed to produce and deforest inside the Indigenous Territories of the Xingu River Basin. Allowing farmers into the Xingu region would result in the deforestation of 57,841 km^2 or 43% of the region. In this counterfactual, precipitation in Mato Grosso would decrease on average by 2% with some regions losing up to 4% and returns of crops production would decrease by 2% with some regions losing up to 8%.

My results highlight a way to integrate the interaction of human behavior and the natural environment into a single consistent model. The mechanism of deforestation-rainfall has consequences for different regions – such as Africa, Indonesia, and other parts of South America – and for different sectors – such as agriculture, energy, and water supply for urban centers. Each one of these consequences is a topic for future research.

As with any research paper, mine has limitations. In particular, my model focus on one particular externality of deforestation and does not study effects on biodiversity, conflict and pollution. Nonetheless, I believe the approach provided in this paper make a sensible contribution in the path to better understand the value of ecosystem services.

References

Abman, R. and Carney, C. (2020). Agricultural productivity and deforestation: Evidence from input subsidies and ethnic favoritism in malawi. *Journal of Environmental Economics and Management*, 103:102342.

Albert, C., Bustos, P., and Ponticelli, J. (2021). The effects of climate change on labor and capital reallocation. Technical report, National Bureau of Economic Research.

Anderson, W., Taylor, C., McDermid, S., Ilboudo-Nébié, E., Seager, R., Schlenker, W., Cottier, F., de Sherbinin, A., Mendeloff, D., and Markey, K. (2021). Violent conflict exacerbated drought-related food insecurity between 2009 and 2019 in sub-saharan africa. *Nature Food*, 2(8):603–615.

Araujo, R. (2022). The value of tropical forests to hydropower.

Araujo, R., Costa, F., and Sant'Anna, M. (2020). Efficient forestation in the brazilian amazon: Evidence from a dynamic model.

Arias, M. E., Farinosi, F., Lee, E., Livino, A., Briscoe, J., and Moorcroft, P. R. (2020). Impacts of climate change and deforestation on hydropower planning in the brazilian amazon. *Nature Sustainability*, 3(6):430–436.

Asher, S., Garg, T., and Novosad, P. (2020). The ecological impact of transportation infrastructure. *The Economic Journal*, 130(629):1173–1199.

Assunção, J., Gandour, C., and Rocha, R. (2013). Deterring deforestation in the brazilian amazon: environmental monitoring and law enforcement. *Climate Policy Initiative*, 1:36.

Assunção, J., Gandour, C., Rocha, R., and Rocha, R. (2020). The effect of rural credit on deforestation: evidence from the brazilian amazon. *The Economic Journal*, 130(626):290–330.

Assunçao, J., Lipscomb, M., Mobarak, A. M., and Szerman, D. (2017). Agricultural productivity and deforestation in brazil.

Assunção, J., McMillan, R., Murphy, J., and Souza-Rodrigues, E. (2019). Optimal environmental targeting in the amazon rainforest. Technical report, National Bureau of Economic Research.

Baccini, A., Goetz, S., Walker, W., Laporte, N., Sun, M., Sulla-Menashe, D., Hackler, J., Beck, P., Dubayah, R., Friedl, M., et al. (2012). Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps. *Nature climate change*, 2(3):182–185.

Balboni, C., Berman, A., Burgess, R., and Olken, B. (2022). The economics of tropical deforestation. *Working paper.*

Barrios, S., Bertinelli, L., and Strobl, E. (2010). Trends in rainfall and economic growth in africa: A neglected cause of the african growth tragedy. *The Review of Economics and Statistics*, 92(2):350–366.

Burgess, R., Costa, F., and Olken, B. A. (2019). The brazilian amazon's double reversal of fortune.

Burgess, R., Hansen, M., Olken, B. A., Potapov, P., and Sieber, S. (2012). The political economy of deforestation in the tropics. *The Quarterly journal of economics*, 127(4):1707–1754.

Bustos, P., Caprettini, B., and Ponticelli, J. (2016). Agricultural productivity and structural transformation: Evidence from brazil. *American Economic Review*, 106(6):1320–65.

Chimeli, A. B. and Soares, R. R. (2017). The use of violence in illegal markets: Evidence from mahogany trade in the brazilian amazon. *American Economic Journal: Applied Economics*, 9(4):30–57.

Copernicus, C. C. S. (2017). Era5: Fifth generation of ecmwf atmospheric reanalyses of the global climate.

Costinot, A. and Donaldson, D. (2012). Ricardo's theory of comparative advantage: old idea, new evidence. *American Economic Review*, 102(3):453–58.

Costinot, A., Donaldson, D., and Smith, C. (2016). Evolving comparative advantage and the impact of climate change in agricultural markets: Evidence from 1.7 million fields around the world. *Journal of Political Economy*, 124(1):205–248.

Dias, M., Rocha, R., and Soares, R. R. (2020). Down the river: Glyphosate use in agriculture and birth outcomes of surrounding populations.

Domínguez-Iino, T. (2021). Efficiency and redistribution in environmental policy: An equilibrium analysis of agricultural supply chains.

Donaldson, D. (2018). Railroads of the raj: Estimating the impact of transportation infrastructure. *American Economic Review*, 108(4-5):899–934.

Ferraz, G., Russell, G. J., Stouffer, P. C., Bierregaard, R. O., Pimm, S. L., and Lovejoy, T. E. (2003). Rates of species loss from amazonian forest fragments. *Proceedings of the National Academy of Sciences*, 100(24):14069–14073.

Fetzer, T. and Marden, S. (2017). Take what you can: property rights, contestability and conflict. *The Economic Journal*, 127(601):757–783.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., et al. (2015). The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific data*, 2(1):1–21.

Golden Kroner, R. E., Qin, S., Cook, C. N., Krithivasan, R., Pack, S. M., Bonilla, O. D., Cort-Kansinally, K. A., Coutinho, B., Feng, M., Martínez Garcia, M. I., et al. (2019). The uncertain future of protected lands and waters. *Science*, 364(6443):881–886.

Haddad, N. M., Brudvig, L. A., Clobert, J., Davies, K. F., Gonzalez, A., Holt, R. D., Lovejoy, T. E., Sexton, J. O., Austin, M. P., Collins, C. D., et al. (2015). Habitat fragmentation and its lasting impact on earth's ecosystems. *Science advances*, 1(2):e1500052.

Harari, M. and Ferrara, E. L. (2018). Conflict, climate, and cells: a disaggregated analysis. *Review of Economics and Statistics*, 100(4):594–608.

Heß, S., Jaimovich, D., and Schündeln, M. (2021). Environmental effects of development programs: Experimental evidence from west african dryland forests. *Journal of Development Economics*, 153:102737.

Hsiao, A. (2021). Coordination and commitment in international climate action: evidence from palm oil. *Unpublished, Department of Economics, MIT.*

Jasechko, S., Sharp, Z. D., Gibson, J. J., Birks, S. J., Yi, Y., and Fawcett, P. J. (2013). Terrestrial water fluxes dominated by transpiration. *Nature*, 496(7445):347–350.

Jayachandran, S. (2021). How economic development influences the environment.

Lawrence, D., Coe, M., Walker, W., Verchot, L., and Vandecar, K. (2022). The unseen effects of deforestation: Biophysical effects on climate. *Frontiers in Forests and Global Change*, page 49.

Lawrence, D. and Vandecar, K. (2015). Effects of tropical deforestation on climate and agriculture. *Nature climate change*, 5(1):27–36.

Lemieux, P. and Luquire, K. (2021). Data stewardship maturity report for noaa climate data record (cdr) of leaf area index (lai) and fraction of absorbed photosynthetically active radiation (fapar), version 4.

Magalhães, C., Pedreira, B., Tonini, H., and Farias Neto, A. (2019). Crop, livestock and forestry performance assessment under different production systems in the north of mato grosso, brazil. *Agroforestry Systems*, 93(6):2085–2096.

Makarieva, A. M. and Gorshkov, V. G. (2007). Biotic pump of atmospheric moisture as driver of the hydrological cycle on land. *Hydrology and earth system sciences*, 11(2):1013–1033.

MapBiomas, P. (2020). Collection4.0.

Marengo, J. A., Soares, W. R., Saulo, C., and Nicolini, M. (2004). Climatology of the low-level jet east of the andes as derived from the ncep–ncar reanalyses: Characteristics and temporal variability. *Journal of climate*, 17(12):2261–2280.

Mérel, P. and Gammans, M. (2021). Climate econometrics: Can the panel approach account for long-run adaptation? *American Journal of Agricultural Economics*, 103(4):1207–1238.

Myneni, R., Knyazikhin, Y., and Park, T. (2015). Mod15a2h modis/terra leaf area index/fpar 8-day l4 global 500 m sin grid v006. *NASA EOSDIS Land Processes DAAC*.

Myneni, R. B., Yang, W., Nemani, R. R., Huete, A. R., Dickinson, R. E., Knyazikhin, Y., Didan, K., Fu, R., Juárez, R. I. N., Saatchi, S. S., et al. (2007). Large seasonal swings in leaf area of amazon rainforests. *Proceedings of the National Academy of Sciences*, 104(12):4820–4823.

Nobre, A. D. (2014). The future climate of amazonia, scientific assessment report. *Sponsored by CCST-INPE, INPA and ARA, São José dos Campos Brazil*.

Nordhaus, W. (2019). Climate change: The ultimate challenge for economics. *American Economic Review*, 109(6):1991–2014.

Ortiz-Bobea, A., Ault, T. R., Carrillo, C. M., Chambers, R. G., and Lobell, D. B. (2021). Anthropogenic climate change has slowed global agricultural productivity growth. *Nature Climate Change*, 11(4):306–312.

Ortiz-Bobea, A., Wang, H., Carrillo, C. M., and Ault, T. R. (2019). Unpacking the climatic drivers of us agricultural yields. *Environmental Research Letters*, 14(6):064003.

Pendrill, F., Gardner, T. A., Meyfroidt, P., Persson, U. M., Adams, J., Azevedo, T., Bastos Lima, M. G., Baumann, M., Curtis, P. G., De Sy, V., et al. (2022). Disentangling the numbers behind agriculture-driven tropical deforestation. *Science*, 377(6611):eabm9267.

Rafey, W. (2020). Droughts, deluges, and (river) diversions: Valuing market-based water reallocation. Technical report, Working paper.

Rangel, M. A. and Vogl, T. S. (2019). Agricultural fires and health at birth. *Review of Economics and Statistics*, 101(4):616–630.

Rattis, L., Brando, P. M., Macedo, M. N., Spera, S. A., Castanho, A. D., Marques, E. Q., Costa, N. Q., Silverio, D. V., and Coe, M. T. (2021). Climatic limit for agriculture in brazil. *Nature Climate Change*, 11(12):1098–1104.

Rocha, R. and Sant'Anna, A. A. (2022). Winds of fire and smoke: Air pollution and health in the brazilian amazon. *World Development*, 151:105722.

Rocha, R. and Soares, R. R. (2015). Water scarcity and birth outcomes in the brazilian semiarid. *Journal of Development Economics*, 112:72–91.

Salati, E., Dall'Olio, A., Matsui, E., and Gat, J. R. (1979). Recycling of water in the amazon basin: an isotopic study. *Water resources research*, 15(5):1250–1258.

Sant'Anna, M. (2021). How Green Is Sugarcane Ethanol? *The Review of Economics and Statistics*, pages 1–45.

Simoes, R., Picoli, M. C., Camara, G., Maciel, A., Santos, L., Andrade, P. R., Sánchez, A., Ferreira, K., and Carvalho, A. (2020). Land use and cover maps for mato grosso state in brazil from 2001 to 2017. *Scientific Data*, 7(1):1–10.

Skidmore, M. (2022). Out-sourcing the dry season: cattle ranchers responses to weather shocks in the brazilian amazon. *Working paper*.

Sotelo, S. (2020). Domestic trade frictions and agriculture. *Journal of Political Economy*, 128(7):2690–2738.

Souza-Rodrigues, E. (2019). Deforestation in the amazon: A unified framework for estimation and policy analysis. *The Review of Economic Studies*, 86(6):2713–2744.

Spracklen, D., Baker, J., Garcia-Carreras, L., Marsham, J., et al. (2018). The effects of tropical vegetation on rainfall. *Annual Review of Environment and Resources*, 43(1):193–218.

Spracklen, D. V., Arnold, S. R., and Taylor, C. (2012). Observations of increased tropical rainfall preceded by air passage over forests. *Nature*, 489(7415):282–285.

Steduto, P., Hsiao, T. C., Fereres, E., Raes, D., et al. (2012). *Crop yield response to water*, volume 1028. Food and Agriculture Organization of the United Nations Rome.

Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge university press.

van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., Gouillart, E., Yu, T., and the scikit-image contributors (2014). scikit-image: image processing in Python. *PeerJ*, 2:e453.

Wright, J. S., Fu, R., Worden, J. R., Chakraborty, S., Clinton, N. E., Risi, C., Sun, Y., and Yin, L. (2017). Rainforest-initiated wet season onset over the southern amazon. *Proceedings of the National Academy of Sciences*, 114(32):8481–8486.

Xu, J., Gao, J., de Holanda, H. V., Rodríguez, L. F., Caixeta-Filho, J. V., Zhong, R., Jiang, H., Li, H., Du, Z., Wang, X., et al. (2021). Double cropping and cropland expansion boost grain production in brazil. *Nature Food*, 2(4):264–273.

Table A1: Results for the Transportation Cost Model

Product	Soybeans	Corn	Meat
Cost from raster	0.07	0.07	0.09
	(0.01)	(0.01)	(0.01)
Constant	10.47	13.41	68.12
	(0.39)	(0.52)	(7.81)
# obs	4147	2557	258
R^2	0.89	0.91	0.73

This table shows regressions for each product resulted from the NLLS procedure described in the text. The unit of measure of the independent variable is R\$(2008)/ton, that is, Brazilian local currency as of 2008 for each ton of product. All estimates have p-value < 0.01.

Appendix

A   Appendix: Data

A.1   Transportation Cost

To compute the transportation cost I combine several types of data. First, I build a transportation network that includes paved and unpaved state and federal roads from the National Bureau of Infrastructure (DNIT)[12]. I convert the transportation network in a raster of 1 km resolution where each pixel represents one of the possible types of terrain: paved road, unpaved road, land with no road, and land with no road inside a protected area. Figure A1 shows a map of the roads data. The key step is to assign values to each type of pixel. To assign these values I use a second data set that gives transportation costs between pairs of Brazilian municipalities for different products from the Group of Research and Extension in Agroindustrial Logistics of the College of Agriculture Luiz de Queiroz (ESALQ)[13]. I then fit a Non-Linear-Least-Squares estimation, as in Donaldson (2018), to find the combination of pixel values that gives the highest average R^2 among different products. Table A1 shows the regressions for each product. In this specification I set the value of a pixel of paved road to 1, unpaved road to 2, land pixel to 5 and land pixel within protected area to 10.

To compute the transportation cost from each land i to international markets I collect data on waterways and ports. Waterways serve as an expressway and are not direct sub-

---

[12]Visited in 11/18/2018, http://www.dnit.gov.br/mapas-multimodais/shapefiles.
[13]Visited in 11/18/2018, https://sifreca.esalq.usp.br/

stitutes for roads, since an agent can access it only by going through a port. I make the distinction between loading ports - ports that give access to waterways - and final ports - ports that give access to international markets. By setting the cost of traversing a pixel of waterway as half the cost of using a paved road, I combine this waterways and ports data with the roads data to build a full network transportation for the Brazilian Amazon. Figure A1 shows a map of the complete transportation network. I then compute the cost for each land i to reach a final port using Dijkstra's algorithm of shortest path[14]. The final step is to use the regression for each product of table A1 to convert the computed cost from the transportation network to a monetary cost. To connect with the model, the regression for each product and the computed cost from the raster gives the transportation cost variable τ_i^c. Table 2 shows descriptive statistics for the transportation cost of different products.
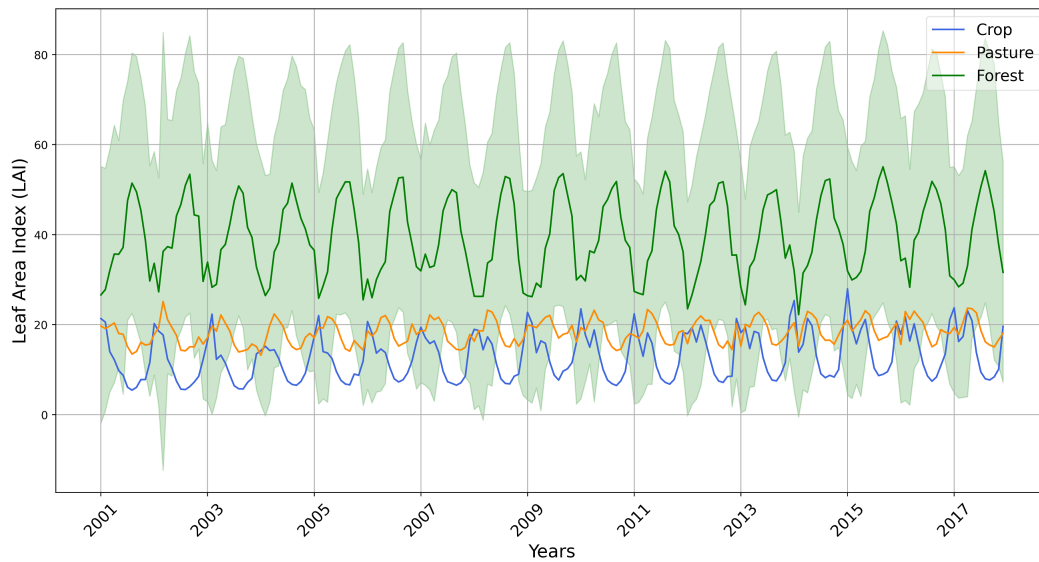
---

[14]This step was done using the graph module of van der Walt et al. (2014)

Figure A1: Transportation Network
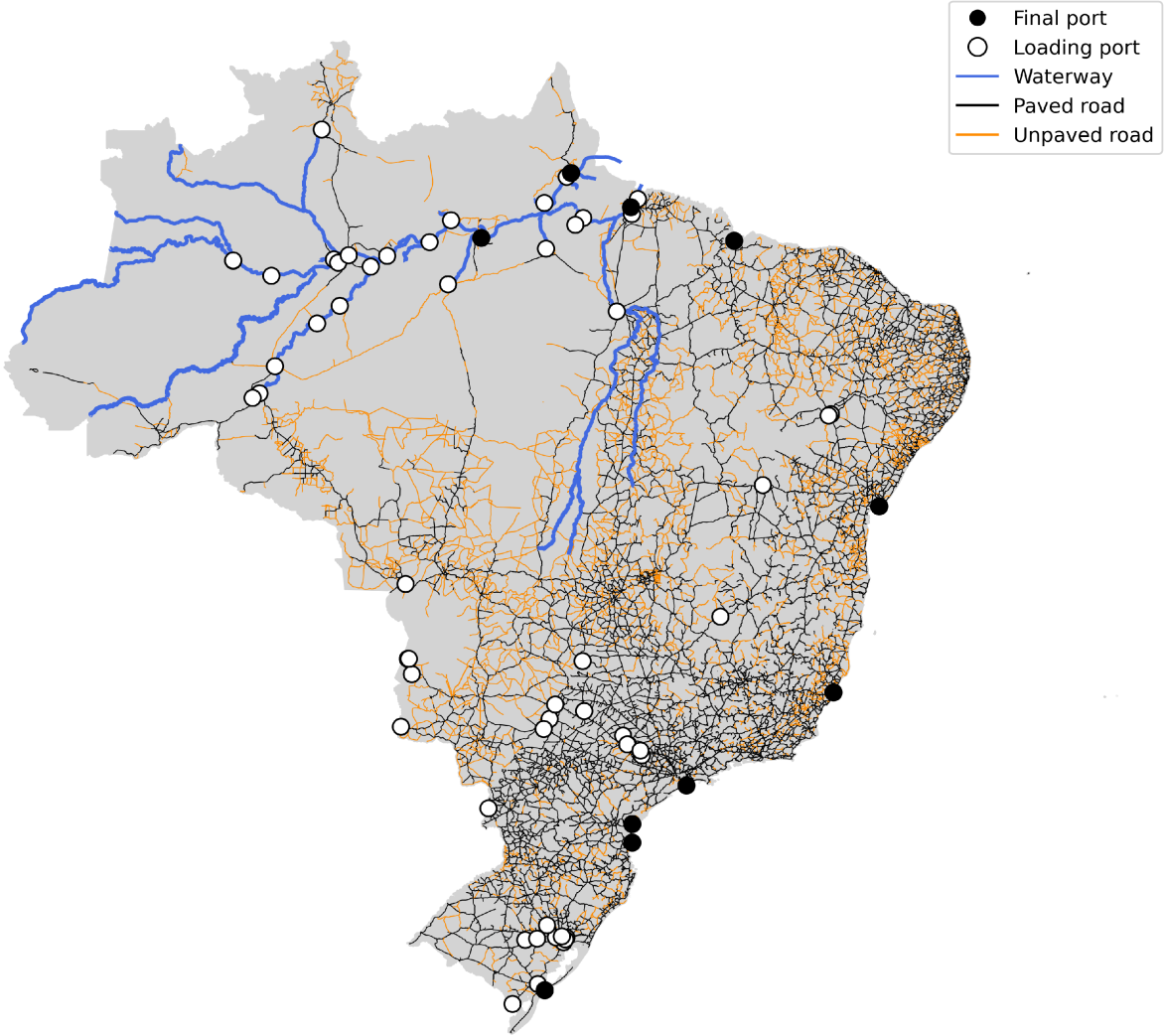
This map shows the location of paved roads, unpaved roads, waterways, and ports.

Figure A2: Land Cover and LAI

This figure shows that the average LAI varies across months and land use. The green shaded area is the 95% confidence interval for LAI inside pixels classified as forest. Unfortunately the variation across land use is not sufficient to separately assign a LAI value for each land use, as evidenced by how the 95% confidence interval of the forest land cover intersects the mean values of LAI of theother two types of land cover. To generate this data I collected average monthly LAI data from Myneni et al. (2015) and land cover data from the Mapbiomas project (MapBiomas, 2020). The mapbiomas project classifies yearly 30 meters pixels in different types of land cover, specially forest, crop, and pasture. I then compute the average and standard deviation of the LAI for each month that intersects each land use for each year.

B Appendix: Model

Contraction map of land use.

Rewrite expression 5 as

$$P_i(F) = \frac{1}{1 + \left(\sum_{c \in C} e^{\frac{\pi_o^c(P(F)) - \pi_o^F}{\lambda}} \right)^\lambda} \tag{13}$$

Define by T the map of $P(F) \rightarrow P(F)$, where $P(F)$ is, as previously defined, the vector of probabilities for all parcels of land where each entry is given by $P_i(F)$. Notice that T maps compacts sets, since each $P_i(F) \in [0, 1]$.

The partial derivative of $P_i(F)$ with respect to $P_j(F)$ is either zero (if j is not under any back trajectory that affects i) or is given by

$$\frac{\partial P_i(F)}{\partial P_j(F)} = -\frac{\left(\sum_{c \in C} e^{\frac{\pi_i^c - \pi_i^F}{\lambda}} \right)^\lambda}{1 + \left(\sum_{c \in C} e^{\frac{\pi_i^c - \pi_i^F}{\lambda}} \right)^\lambda} \frac{1}{1 + \left(\sum_{c \in C} e^{\frac{\pi_i^c - \pi_i^F}{\lambda}} \right)^\lambda} \frac{\sum_{c \in C} e^{\frac{\pi_i^c - \pi_i^F}{\lambda}} \frac{\partial \pi_i^c}{\partial P_j(F)}}{\sum_{c \in C} e^{\frac{\pi_i^c - \pi_i^F}{\lambda}}} \tag{14}$$

Where $\frac{\partial \pi_i^c}{\partial P_j(F)}$ is the marginal effect of an increase in forest probability of a single plot of land.

$$\frac{\partial \pi_i^c}{\partial P_j(F)} = \beta_c \frac{1}{\sum_{o(j)} 1} \sum_{m \in M} \theta_m \tag{15}$$

The first two ratios of Expression 14 are always between zero and one, since the sum of exponentials is a positive number. For the third ratio to be between zero and one, it is sufficient that $|\frac{\partial \pi_i^c}{\partial P_j(F)}| < 1$, that is, the increase in rainfall due to a marginal increase in the forest probability of a single pixel cannot be too high. One can then verify this condition with the estimated parameters. In my application, this condition is valid and the parameter β_c could be multiplied by up to a hundred that the condition would still be valid.

Therefore Expression 14 is bounded, such that $|\frac{\partial P_i(F)}{\partial P_j(F)}| < \omega$, for $|\omega| < 1$. By the mean value theorem, given $P(F)$ and $\tilde{P}(F)$:

$$||T(P(F)) - T(\tilde{P}(F))|| \leq \omega ||P(F) - \tilde{P}(F)|| \tag{16}$$

That is, T is a contraction.

C   Appendix: Results

Table A2: Estimation Results for the Climate Model

Month	(0)	(1)	(2)	(3)
Jan	0.908	0.910	1.158	1.410
	(0.020)	(0.021)	(0.014)	(0.019)
Feb	1.125	1.142	1.370	1.432
	(0.022)	(0.023)	(0.014)	(0.020)
Mar	0.827	0.811	0.993	0.963
	(0.021)	(0.022)	(0.011)	(0.015)
Apr	1.434	1.395	1.479	1.461
	(0.043)	(0.044)	(0.023)	(0.019)
May	1.284	1.314	2.064	1.134
	(0.038)	(0.040)	(0.024)	(0.025)
Jun	0.680	0.640	1.238	0.290
	(0.017)	(0.018)	(0.030)	(0.013)
Jul	0.404	0.410	1.000	0.069
	(0.014)	(0.015)	(0.030)	(0.010)
Aug	0.950	0.948	1.370	1.021
	(0.022)	(0.024)	(0.043)	(0.022)
Sep	1.202	1.206	1.421	1.208
	(0.026)	(0.028)	(0.020)	(0.018)
Oct	0.569	0.550	0.897	0.787
	(0.019)	(0.020)	(0.012)	(0.016)
Nov	0.499	0.515	0.901	1.292
	(0.023)	(0.024)	(0.014)	(0.014)
Dec	0.770	0.781	1.164	1.479
	(0.025)	(0.027)	(0.017)	(0.014)
# Obs	487,560	487,560	487,560	487,560
Month FE	Yes	Yes	Yes	No
Year FE	No	Yes	Yes	Yes
Pixel FE	No	No	Yes	No
Pixel-Month FE	No	No	No	Yes
Distance	Yes	Yes	Yes	Yes
R^2	0.82	0.83	0.86	0.88
R^2 (within)	0.19	0.20	0.24	0.17
R^2 (between)	-0.42	0.29	-0.40	0.47

This table shows results for the empirical climate model 9. As the unit of measurement of upwind exposure to the forest does not have an intuitive interpretation I normalize it to have the standard deviation equal to one. All specifications include as control the total distance traveled over land. Standard errors are clustered at the pixel level. All estimates have p-value < 0.01.