

ESCUELA DE
NEGOCIOS



EMPIRICAL EVIDENCE OF THE ECOSYSTEM SERVICE OF FORESTS IN CHILE, IN LOCAL CLIMATE REGULATION

TESIS PARA OPTAR AL GRADO DE MAGISTER EN ECONOMIA

07 de octubre del 2024

LEONARDO ANTONIO SEPÚLVEDA HURTADO

PROFESOR GUIA: SEBASTIÁN GALLEGOS

PROFESORES CORRECTORES: LUIS HERSKOVIC Y ESTEBAN LÓPEZ

ACCREDITATIONS

MEMBER OF



Contents

1	Introduction	4
2	Background	6
2.1	Evidence in the Amazon Rainforest	6
2.2	The Chilean Case	7
3	Data	9
4	Empirical Approach	18
4.1	Spatial-Panel Approach	18
4.2	Cross-sectional Approach	20
4.3	Tests for economic value	20
5	Results analysis	21
5.1	Results for the Spatial-Panel Approach	21
5.2	Results for the Spatial Cross-sectional Approach	25
5.2.1	Simulation	27
5.3	Results for the tryouts, for economic value	29
6	Conclusion	32
7	Appendix	37
7.1	Back trajectories average	37
7.2	Exploratory investigation	39
7.3	Land Use	40
7.4	Difference between LAI (Leaf Area Index) and NDVI (Normalized Difference Vegetation Index):	40
7.5	Example of the data and regression with the temperature levels	41
7.6	Graphics of example of the Data Set for the other variables	43
7.7	Regression by type of forest	44
7.8	Results Tables of the economic value section	45
7.9	Treatment of missing values	48

Abstract

The capacity of large forest units to influence local weather has been well studied. Evidence from the Amazon suggests that these forests can affect regions deep within the continent through a spillover effect in wind direction, which promotes vegetation growth. Are we therefore witnessing an impact of an ecosystem service provided by forests in regulating local weather? In this work, I test this hypothesis for central-south Chile, because of the presence of a large forestry industry and remnants of native forests. Using satellite remote sensing data spanning the last two decades and a Spatial-Panel model to estimate the impact of the spatial effects of forest units. A positive correlation between forest units has been found suggesting the existence of this phenomenon. Given these discoveries, when testing heterogeneous effects by municipalities of an increase of forests NDVI, suggests the existence of a positive externality, but a direct mechanism that increases the production in nearby croplands, because of the better conditions given by forests, meaning an increase in the income of the agricultural related workers, remains still unclear because ambiguous results.

1 Introduction

Forests are well known for their ecological role, yet quantifying the economic value of their ecosystem services remains challenging and must be considered in relation to the industries that they influence (Sagoff, 2011). One of them that has not been extensively studied due to the difficulty of identification is their role in the water cycle. Large forests transport humidity through the wind, and the movement of water vapor originating from forests enhances conditions for vegetation growth and influences weather patterns far beyond the moisture source (Makarieva and Gorshkov, 2007). Therefore, one would expect an intriguing positive relationship between the proximity of forests to croplands and their productivity (Teixeira et al., 2021). In the context of climate change, forests are expected to have a positive impact on regulating local weather, helping to mitigate the effects of extreme events that generate uncertainty on the agricultural industry (Depetris-Chauvin et al., 2023).

A great portion of land use is determined by landowners' choices to maximize their benefits. However, two constraints influence these decisions: on one hand, the natural land conditions, which limit production options but can offer comparative advantages under favorable circumstances; and on the other hand, incentives such as policy or market forces, which shape what can be produced (Murtazashvili et al., 2019). Evidence of this can be seen in shifts between different types of crops or transitions between forests and croplands, or vice versa. These decisions are influenced by the demand for each type of product or trade frictions, such as import tariffs. For example, closed economies tend to adjust land use to meet internal demand (Foster and Rosenzweig, 2003). In contrast, open economies tend to adjust their land use in response to international commodity prices (Farrokhi et al., 2023). If these choices are made without considering the spatial effects of forests, they could lead to the degradation of the ecosystem and a shift to less productive alternative (Staver et al., 2011) through the loss of forest services or a shift to land uses with limited reversibility, such as mining.

This thesis has two main objectives: First, to test the hypothesis that, on a large scale and in the average wind direction, the movement of water vapor from forests enhances conditions for vegetation growth and influences weather patterns far beyond the moisture source in Chile, specifically between latitudes 30° and 40° . If this hypothesis holds, we would expect to econometrically demonstrate a correlation between forests in the upstream direction of prevailing winds and vegetation patterns downstream. Second, to assess the sign and significance of this phenomenon on the income of individuals involved in agricultural activities.

For this exploratory investigation, I first use satellite remote-sensing data from the last two decades (2001-2023) for central-southern Chile (between latitudes 30° and 38° south). This data includes the Normalized Difference Vegetation Index (NDVI) and other weather controls. The novel approach involves using a Spatial-Panel model to test the average spillover impact of neighboring areas' vegetation on vegetation levels in the wind direction that is defined by the weight matrix (Araujo et al., 2023).

To measure the economic value, I analyze heterogeneous effects across municipalities (the smallest administrative units in Chile) based on the amount of vegetation measured by NDVI. Additionally, I use land use change data to differentiate between types of vegetation, such as Native and Commercial Forests, and incorporate this information as an input for household income data from the CASEN survey, which assesses the socioeconomic characteristics of Chile. The methodology involves a simple regression analysis, incorporating environmental information at the municipal level. This approach is inspired by existing literature on the effects of protected areas on local communities (Cheng et al. (2023); Kalinin et al. (2023)) and previous studies conducted in Chile about the impact of commercial forests (Bonilla-Mejía and López, 2024).

After identifying a significant average impact from the spillover effect of downwind neighbors, with increases ranging from 0.1 to 0.4, these results suggest the existence of a phenomenon where forests enhance vegetation growth conditions in the direction of the wind across the continent's interior. This pattern is consistent across all regions studied and motivates further investigation into whether this phenomenon positively impacts household income by reducing the uncertainty associated with climate change and improving plantation growth. Given this, a positive impact would be expected for farmers in municipalities with greater forest vegetation (Castle et al., 2022). The results indicate a positive correlation between increases in a municipality's average NDVI of forests and household income. However, a direct link to individuals involved in the agricultural sector was not established. Additionally, the data was insufficient to differentiate between the effects of native forests and commercial plantations, with mixed signals observed.

The understanding of the relationship of forest spillover effects in regulating local weather enriches the policy maker discussion over land use regulation and the optimal designation of new reserves or national parks (Fetzer and Marden (2017); Araujo et al. (2020)). Also this works is part of the literature of ecosystem services awareness (Laurans et al., 2013). Additionally, can works as input for more complex models as forest plantation rotation models (Piazza and Roy, 2019) or macroeconomic models that accounts for the agricultural and forestry industries (Banerjee et al., 2021). Ultimately, my research would help to address the gap in information regarding the interconnection of portions of forests through the water-recycling process in Chile.

The order of this thesis is the following (2) Background (2.1) Evidence of the the role of forests in the water cycle, (2.2) The Chilean case, why is an interesting experimental site to test this. (3) Data sources and summary statistics. (4) Empirical approaches. (5) Results analysis. (6) Conclusion. (7) Appendix.

2 Background

2.1 Evidence in the Amazon Rainforest

This phenomenon has been widely studied for tropical forests and with special attention in the Amazon forest given its international significance in influencing global weather patterns (Shukla et al., 1990). The results of these studies highlight the critical role of forests by contributing significantly to their own hydrological cycle through processes such as transpiration and cloud formation ((Staal et al., 2023); (Clara Zemp et al., 2017)), being this process of moisture recycling (or precipitation recycling) relevant for areas far away from the natural moisture of the sea (Hirota et al., 2011), an example of this is the heightened sensitivity to drought observed in the southwestern region of the Amazon (Smith et al., 2023).

The alarming issue is climate change, which is expected to prolong the duration of the dry season in the region (Boisier et al., 2015). This could trigger a vicious cycle: trees may not survive the prolonged dry season, coupled with reduced precipitation, leading the forest to produce insufficient moisture. This may lead into trespassing a tipping point, potentially causing irreversible ecological changes in the region, transitioning from tropical forest to savannah or treeless terrain (Staver et al. (2011); Verbesselt et al. (2016)).

The direct economic impact of this issue includes the loss of precipitation in nearby areas, which could jeopardize rain-dependent croplands like soy in Brazil (Teixeira et al., 2021). Additionally, the reduction in air moisture may escalate the risk of fires in the region. In summary, the main findings suggest that deforestation has spatial effects, that extend beyond the deforested region, influencing weather patterns in the direction of wind.

The origin of this phenomenon may be explained by the basin's inherent abilities to attract and capture moisture from the sea, a phenomenon supported by the controversial biotic pump theory (Makarieva and Gorshkov, 2007). Because large forested areas, characterized by higher rates of evapotranspiration compared to the sea, would attract the fluxes of air with moisture. This phenomenon suggests a relationship wherein regions with dense vegetation tend to receive more precipitation than treeless areas (Spracklen et al. (2012); Sing et al. (2023)). This theory would explain the capacity of forests to transport humidity inside the continent, with the water recycling process, because forests would keep moving moisture far away from the sea.

Hence anthropogenic disturbances, such as the clear-cutting of native forests and large fires along the coast, are likely linked to the reduction of vegetation within the continent. Among the difference between the tropical rainforests, and the temperate ones here in Chile, is necessary to take the evidence in the literature that suggests that there could be a different effect depending on the type of vegetation. A notable differentiation emerges between Native Forests and Commercial Forest Plantations. In the context of the biotic pump theory, it is that only native and biodiverse forests that evolved in the local conditions would be able to produce this phenomenon and keep

a healthy environment that moves humidity into the forest (Makarieva and Gorshkov, 2007). Furthermore, empirical evidence indicates that forest plantations exhibit reduced capacity for water retention following rainfall. The limited permeability of the soil leads to rapid drainage into surrounding basins, even more, captures less humidity than native forests, and alters the soil composition. (Balocchi et al., 2022). Having a negative impact on neighborhood areas, discouraging local farmers from establishing near forest plantations. Some novel work does this connection, using the national National Socioeconomic Characterization Survey (CASEN) in conjunction with the national forest inventory compiled by the National Forestry Corporation (CONAF) To measure the impact of Plantation Forest expansion on family income at the municipal level. Their findings suggest that there is no green growth with commercial forests because the house income decreases if there is an expansion of Forest Plantations in their municipality (Bonilla-Mejía and López, 2024). Consequently, it is imperative to introduce control variables with the type of forest or the use of soil.

2.2 The Chilean Case

The central-south of Chile, between latitudes 34° and 40°, serves as an intriguing experimental site to examine this relationship. This region has witnessed a significant expansion of the forestry industry since the 1970s, driven by efforts to augment economic growth by promoting forestry and mitigate erosion resulting from previous land use for croplands. (Balocchi et al., 2022). Incentives were introduced to encourage the conversion of native forest and eroded land into commercial forestry in the central-south region of Chile. The predominant species utilized for this purpose include *Pinus radiata* (cycles of 18 to 25 years for harvesting) and various types of *Eucalyptus* (cycles of 10 to 12 years for harvesting), representing an increase in forested areas. Nevertheless, there has been a significant conversion from native forests to plantation forests, exceeding 10% in the three main administrative regions of our study area (Maule, Biobío, and La Araucanía) between 1986 and 2011, with this trend likely dating back decades. (Heilmayr et al., 2016).

Chile also has the particularity of having two mountain ranges, running from north to south. One is along the coast, characterized by smaller peaks compared to the Andes mountain range, which boasts peaks averaging over three to four thousand meters above sea level. As with any mountain range, they serve as a natural barrier that captures the humidity that comes from the sea (Makarieva and Gorshkov, 2007).

In the coastal range are found the vast majority of the commercial forests plantations. In Chile by the year 2008, the amount of plantations was approximately 2.24 million hectares which constituted just 14% of the total forest area in continental Chile. The forestry industry accounted for nearly 7.3% of the GDP, making it the second largest industry after the mining sector.(Raga, 2009). Nevertheless, in the latest report, in 2022 the percentage represented by the industry was 1.6% of the total GDP. This reduction may be attributed to variability resulting from periods of harvesting and fluctuations in international demand. Also, the sector saw a decrease in the number of employed workers. (Poblete Hernández et al., 2023).

Obviously There is a controversial aspect regarding commercial forest plantations. On one hand, they increase water runoff by flowing on the surface instead of being absorbed, and they exhibit high rates of evapotranspiration due to their rapid growth (Little et al., 2009). Additionally, they contribute to increased fire occurrence, as evidenced near Valparaiso city (Ruiz et al., 2017), and displace agricultural lands. On the other hand, alternative land uses such as croplands with poor management could have even worse environmental impacts. This study suggests a potential positive impact from the forestry industry while acknowledging that natural native forests remain the optimal choice for environmental conservation. It is important to assess whether the positive or negative externalities from the forest industry carry more weight and to compare these findings with the remaining native forests.

The mechanism that needs to be studied is whether this phenomenon occurs in Chilean forests whether in Native or Commercial Plantations. If it does, we would expect agricultural communities to experience reduced uncertainty from weather shocks. Consequently, municipalities with higher levels of forest vegetation would, on average, have higher incomes than those in other municipalities, due to increased crop production and larger revenues. Conversely, if the negative externalities of commercial plantations outweigh the benefits, an increase in plantations could negatively impact nearby crop agriculture, potentially displacing it. This is critical because the agricultural sector employs more workers and has shorter cycles between harvests compared to the forestry sector.

3 Data

For the study of the existence of the ecosystem service, remote sensing data was essential. A summary of the data sources is provided in Table 1. The main environmental variables—NDVI (vegetation index), precipitation, evapotranspiration, humidity, land temperature surface—were sourced from Google Earth Engine (GEE) imagery, accessed via the "rgee" API (Application Programming Interface). and complemented with "rgeeExtra" (Aybar et al., 2020) Using the pre-calculated datasets, that use diverse bands of information capture monthly (Donaldson and Storeygard, 2016)

For the construction of the weight matrix, the necessary data included the back trajectories of the wind. I utilized the "SplitR" library within RStudio, which implements the methodology for analyzing and modeling these trajectories with an HYSPLIT model (Hybrid Single-Particle Lagrangian Integrated Trajectory) (Stein et al., 2015), that was developed by NOAA (National Oceanic and Atmospheric Administration) also from there comes the atmospheric data of the wind.

For the land use and socioeconomic data, I sourced information directly from Chilean institutions. The land use data is essential for distinguishing between the effects of different forest types and identifying patterns in NDVI values. This information is obtained from the Land Use Cadastre conducted by the National Forest Corporation (CONAF), sourced from the Department of Territory Information System (SIT).

The socioeconomic data comes from the CASEN survey, which includes household income and various other variables, compiled by the Ministry of Social Development and Family. I utilized the last four surveys. Although there was a survey conducted during the pandemic in 2020, I opted not to include it due that the data reflects the shocks caused by the health crisis.

The primary dataset depicted in Table 2, includes the vegetation index, in this case, the Normalized Difference Vegetation Index (NDVI) is calculated as the normalized difference between visible and near-infrared light reflection. Higher NDVI values indicate denser vegetation due to lower reflection of visible light and higher reflection of near-infrared light. Conversely, lower values suggest lower vegetation density (refer to Appendix 7.2 for further details) (Didan, 2021). The subsequent data in the table encompasses key weather variables, being precipitation and evapotranspiration both of them are measured in millimeters over a specified time period (Abatzoglou et al., 2018). Additionally, the data includes the Specific Humidity, that is the concentration of water vapor mass present in a certain mass of air in a relationship of kilograms per kilograms. (McNally, 2018).

As illustrated in Table 2, the original data frame contains 3,066,912 observations for the 92 periods contained in the variable time, which accounts for each quarter in the range of the twenty-three years from 2001 until 2023. Each period consists of 33,336 pixels at a resolution of 2000 meters. In order to be able to process the data with the computer's limitations, I divided the data frame into the regions of O´higgins, Maule,

Data	Source	Use
Environmental Data		
Wind direction	NOAA (National Oceanic and Atmospheric Administration), NCEP-NCAR Reanalysis 1	For building the average back trajectories for the spatial weight matrix "W".
Normalized Difference Vegetation Index (NDVI)	MODIS (Moderate-Resolution Imaging Spectroradiometer)	It's the Vegetation Index, for the study is the dependent variable goes from 0 to 1.
Precipitation (PP)	Terra Climate	Control variable for the study of NDVI.
Evapotranspiration (ET)	Terra Climate	Also a control variable.
Humidity (HU)	FLDAS from NASA	Also a control variable.
Land Surface Temperature (LST)	MODIS	Also a control variable.
Land Use Data		
Cadastré of Land Use per Region	Department of Territory Information System (SIT) from The National Forest Corporation (CONAF)	This information allows to distinguish the type of forest and the location.
Socioeconomic Data		
Houshold information	National socioeconomic characterization survey (CASEN)	It's use in the regresion with heterogenous effects by municipality, contains also control variables as the gender, economic status, etc...

Table 1: Summary of the data sources

Ñuble and BíoBío with parts of the Araucanía region. The primary aim was to preserve spatial coherence within the data (Because of the spatial consideration is not possible to take aleatory sub-samples of pixels).

As presented in Table 3, the difference in the average of the 92 periods between the main data frame and the regions is almost zero, except for the precipitation variable, because the precipitation declines if we move from north to south in continental Chile (Refer to Appendix 7.2).

The econometric model under examination exploits the variations originating from the clear-cutting process of forest plantations and other deforestation processes. Although when assessing the annual average of the dependent variable, the values exhibit apparent constancy over the years, as illustrated in Table 4.

For the spatial weight matrix, I needed to connect the pixels in the way of the

average back trajectory of the wind. Using the wind information data, I calculate 2 day back trajectories for a sample of pixels (it is what takes to the back trajectory to reach the ocean), in every season of the year in a sample of years in the range of study (2001-2023), using the meteorological data from NCAR/NCEP global reanalysis data. (Refer to Appendix 7.1 for further information).

After this exercise, the average back trajectory for a point in central-south Chile is depicted in Figure 3. The first step I took was to linearize the average back trajectories of wind, leading to the conclusion that the simplest approach was to account for direct left, below, and diagonal left-below contiguity of the pixels as first-degree neighbors. An example of this process is shown in Figure 4.

The weight matrix, denoted as W^1 , is the result of the Hadamard multiplication between the wind matrix, which captures only the connections influenced by wind, and the matrix representing the total first-degree neighbors for each pixel $[i,j]$. This neighbor matrix, constructed from actual data, accounts for all neighboring pixels and can be visualized similarly to the possible moves of a queen on a chessboard, since the data is organized in a grid format.

About the Land Use information, I aggregated the data into a 2000-meter pixel resolution using a counting formula. Each pixel in my dataset represents the most common land use type derived from a bigger resolution of pixels in the original cadastre. However, this approach has a notable limitation: the periodicity of dataset updates.

The type of Land Uses are: Urban and Industrial areas; Agricultural Land; Grasslands and Shrublands; Forests; Wetlands; Areas Devoid of Vegetations; Glaciers and Snow; Water bodies. And using the Land Underuse information I was able to distinguish between Native Forests or Forests Plantations (Commercial Forests).

To address the limitation of the cadastral data's infrequent updates, which occur approximately once per decade, I assume that the cadastral data for each region remains valid for an entire decade. This allows me to merge them into a single dataset covering at least half of the panel data's time span. Given the slow pace of land use changes and the 2000-meter resolution of my analysis, minor changes would likely go unnoticed, making this assumption reasonable. In Table 6, "The latest" is the append of each region and is valid from 2012 until 2023, and "The prior update" is the append of each region and is valid from 2001 until 2011.

Subsequently, Table 7 presents the changes in Land Uses between each period, detailing the alterations per municipality. These changes will be utilized to estimate the economic value. It's important to note that the municipality is the smallest official administrative subdivision in Chile. In Graphic 7, a Sankey diagram illustrates the main land use changes. It's important to note that the increase in Native Forests is not an ecological miracle, The primary native forest in the north of the range of the study area is the sclerophyllous forest, which by definition exists in the intermediate zone between tall bushes and trees. Even usually the Land Underuse of "Grasslands and shrublands" is native forests, for the same reason.

However, it is important to highlight the limitations of the datasets, which contains only data within the latitude range of 34° to 38° south due to computer limitations, In Appendix 7.3 and 7.4 are examples of how it looks the data in a map. Additionally, the results are valid only to a 2000-meter resolution due to the 'modifiable areal unit problem' (MAUP). This issue arises because the construction of the resolution involves averaging values or assigning the most common value for example for the land use categories.

Variables	Obs	Mean	Standard Deviation	Min	Max
NDVI	3,066,912	0.549	0.25	0	0.955
PP	3,066,912	70.991	71.414	0	651
ET	3,066,912	81.278	47.413	0	190.1
HU	3,066,912	0.005	0.001	0	0.009
Time	3,066,912	46.5	27.916	1	92

Table 2: Summary statistics for the entire dataset, where the acronyms stand for: NDVI: Normalized Difference Vegetation Index; PP: Precipitation; ET: Evapotranspiration; HU: Humidity or Moisture; Time: unit of time identification in quarters of years, that contains from 2001 until 2023.

Variables	Obs	Mean	Standard Deviation	Min	Max
NDVI	1,073,640	0.492	0.256	0	0.933
PP	1,073,640	57.5355	55.227	0	344
ET	1,073,640	82.836	48.288	0	160.9
HU	1,073,640	0.005	0.001	0	0.009
Time	1,073,640	46.5	26.557	1	92

Table 3: Summary statistics for the Maule region, all-time average (2001-2023) (11,670 observations * 92 periods)

Variables	Obs	Mean	Standard Deviation	Min	Max
NDVI 2001	46,680	0.48	0.255	0	0.905
NDVI 2008	46,680	0.483	0.251	0	0.904
NDVI 2016	46,680	0.52	0.253	0	0.933
NDVI 2017	46,680	0.484	0.255	0	0.928
NDVI 2023	46,680	0.481	0.257	0	0.899

Table 4: Summary statistics for the Maule region, year average (obs. = 11,670 observations * 4 periods (four quarters))

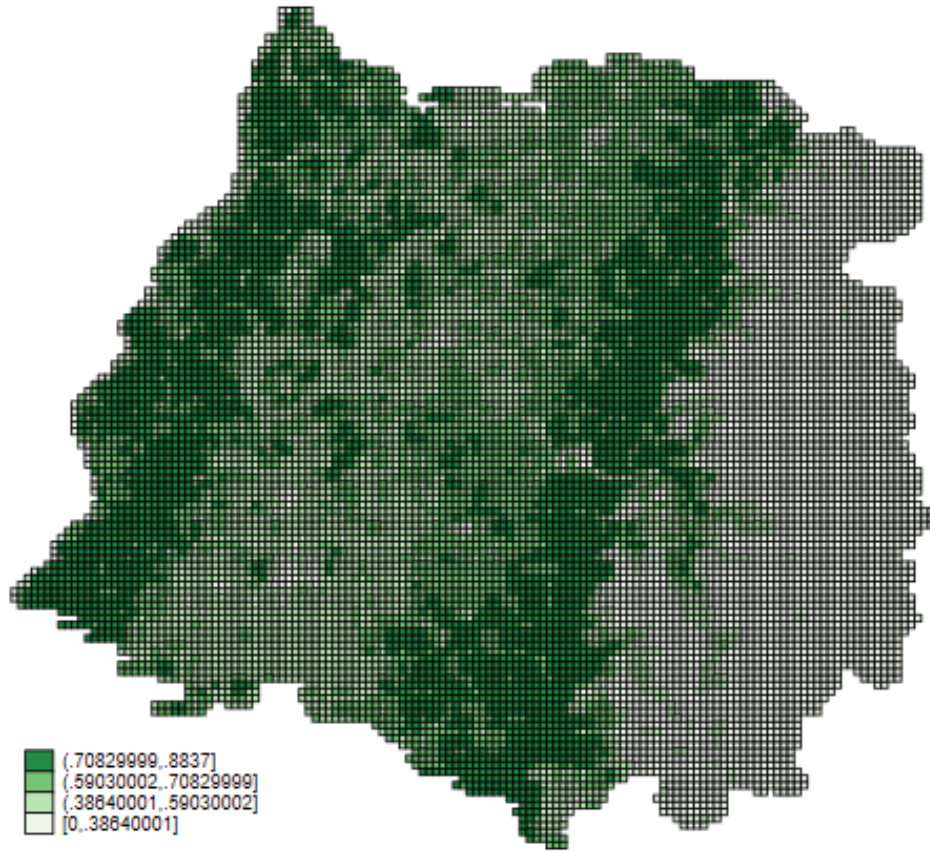


Figure 1: NDVI values in the Maule region of 11,670 pixels/observations in 2016 fourth quarter. At 2000 meter resolution.

Variables	Obs	Mean	Standard Deviation	Min	Max
NDVI 2016q4	11,670	0.525	0.236	0	0.884
NDVI 2017q4	11,670	0.506	0.242	0	0.878

Table 5: Summary of the NDVI values presented in Figures 1 and 2 respectively. In the Maule region.

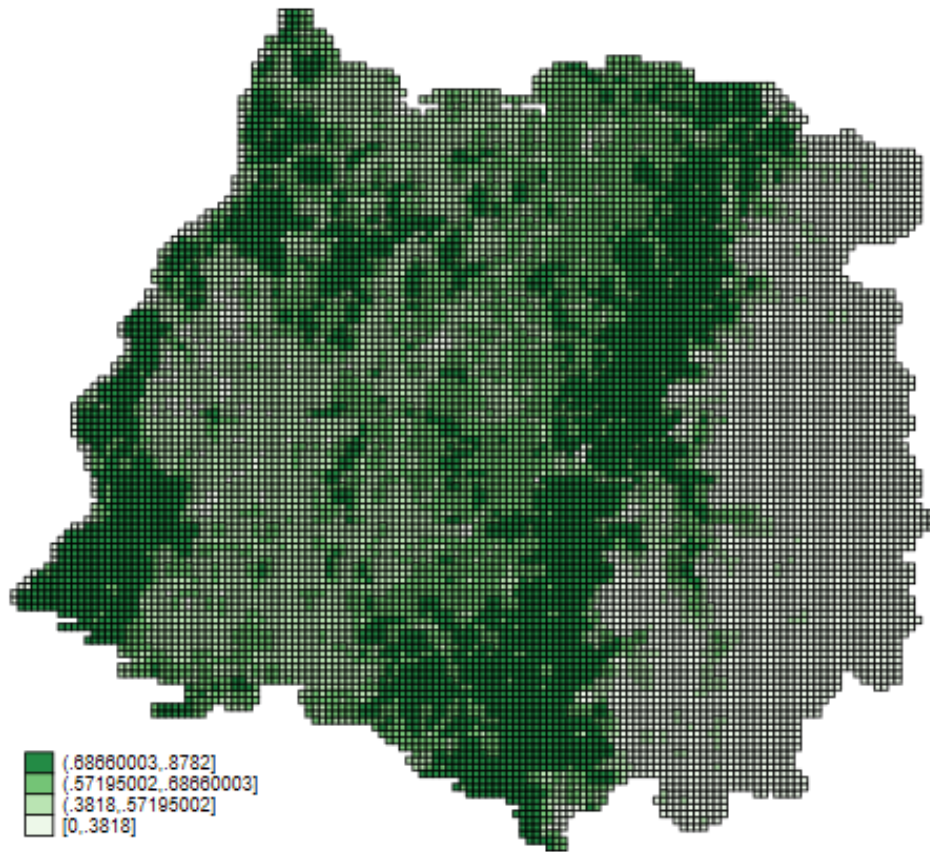


Figure 2: NDVI values in the Maule region of 11,670 pixels/observations in 2017 fourth quarter. At 2000 meter resolution.

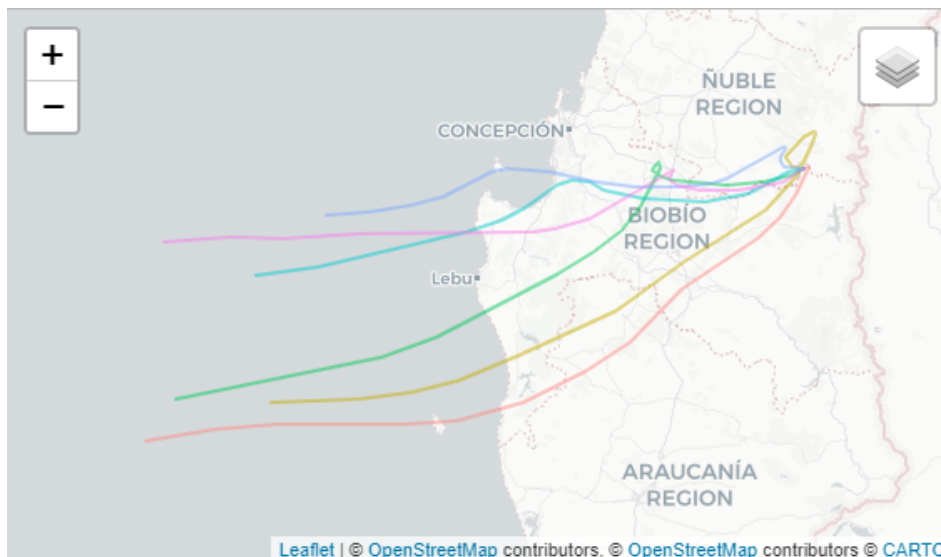


Figure 3: Represents the average of 48 hrs wind direction, within a sample of points, the lines represent the average back trajectories that arrive at an average point. Suggesting that the weight matrix should be represented as depicted in Figure 2. Further information is explained in Appendix 6.1.

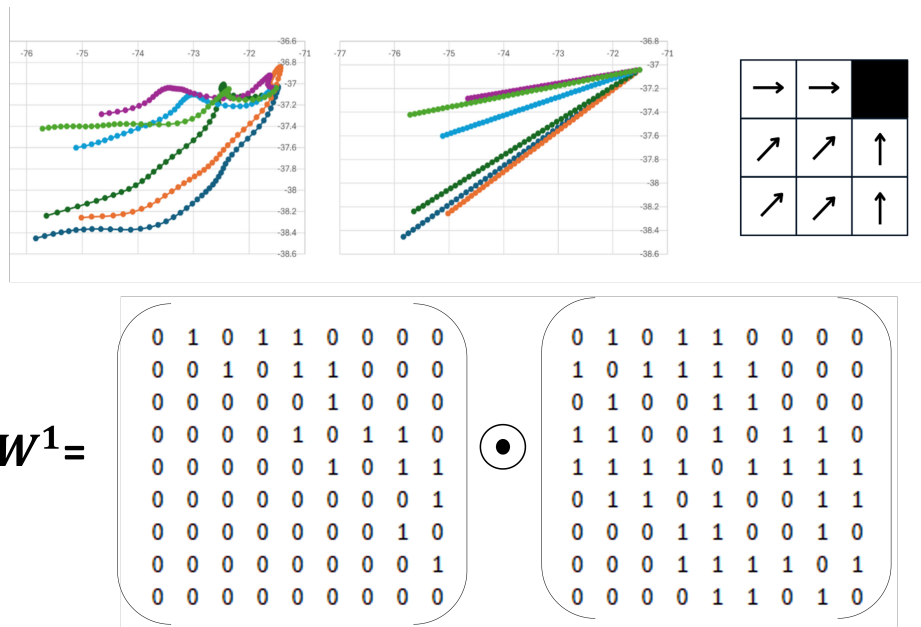


Figure 4: From left to right, the first step involves displaying the raw average wind direction, followed by its linearized form, and then an illustration of the typical wind pattern over the year and across the two-decade span. Finally, the matrix W^1 is produced by multiplying the wind direction matrix (on the left) with the matrix representing all direct neighbors (on the right). This ensures that actual data is combined with the wind direction matrix, which was created using Mata, a sublanguage of Stata. As a note, the spatial weight matrices, which partially account for wind direction and pixel contiguity, assign a value of 1 to direct neighbors and 0 otherwise. The diagonal elements are always zeros, as they represent the same unit, and other entries are zeros where there are no inter-unit connections.

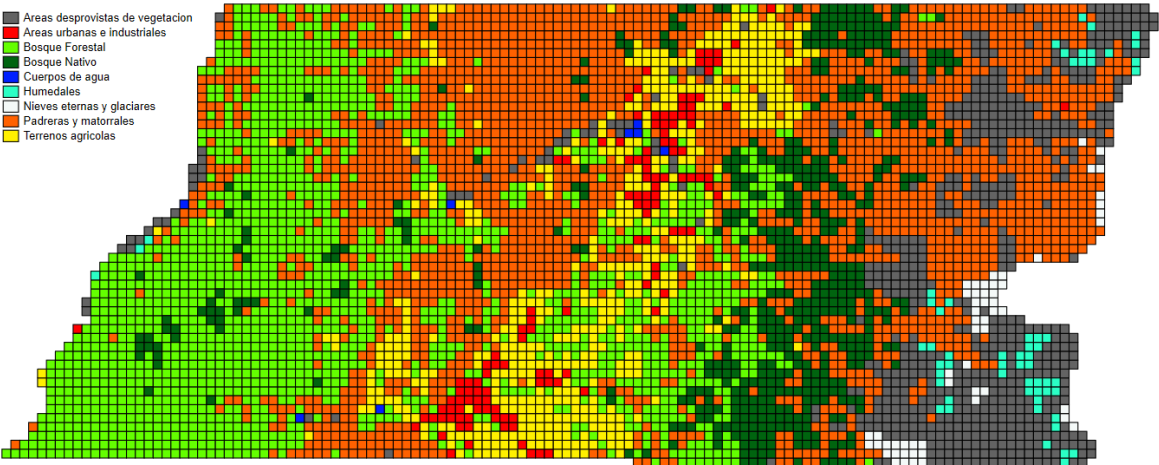


Figure 5: Land use, from the prior update of the cadastre, in an example, at 2000 meter resolution.

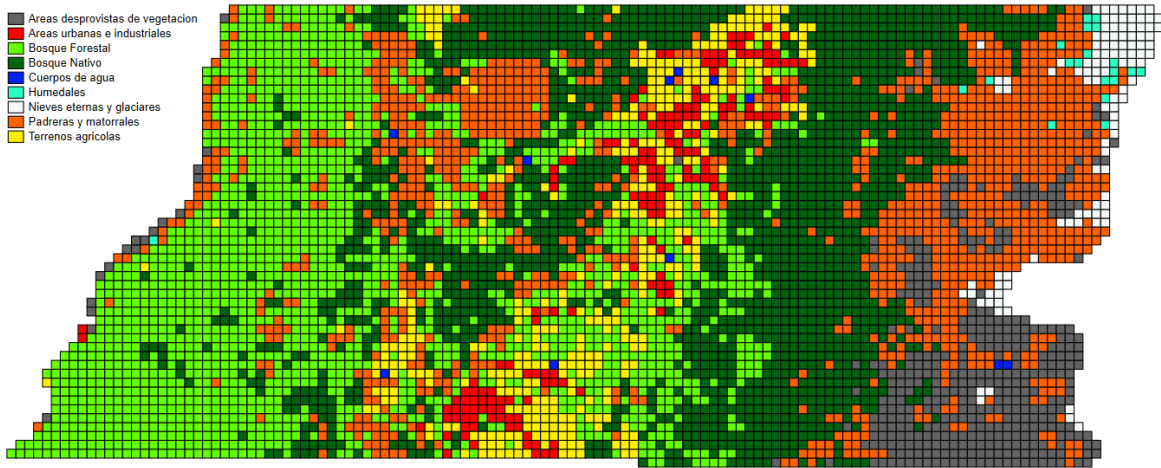


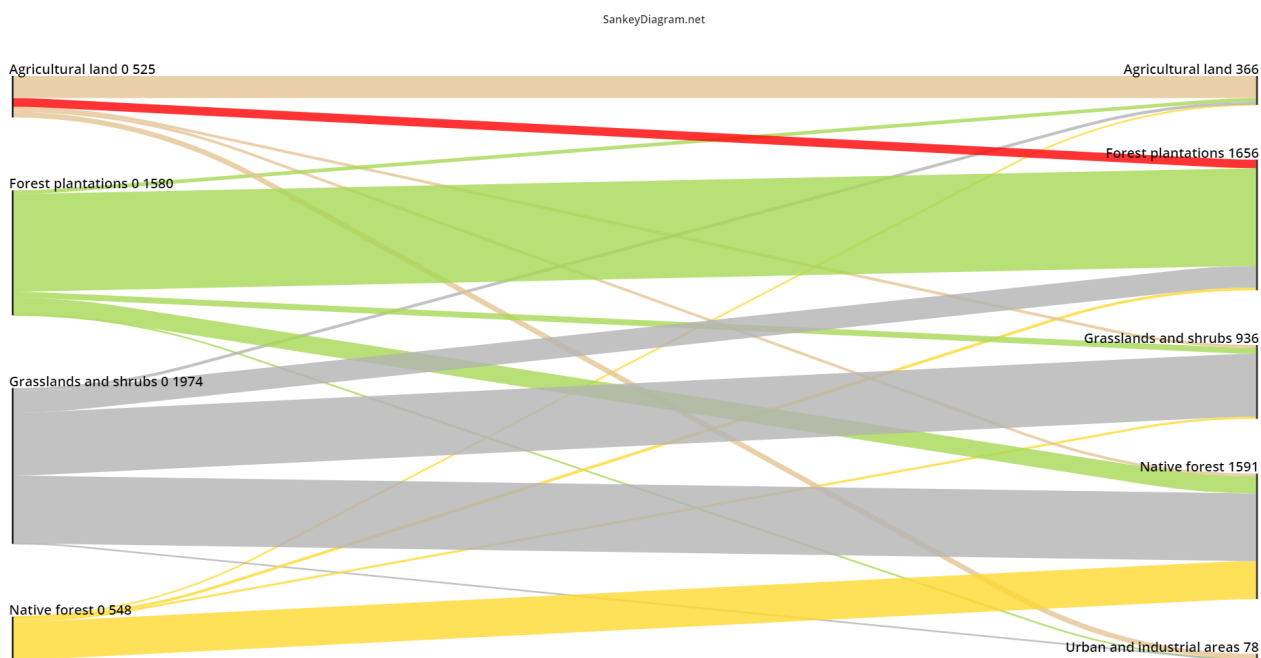
Figure 6: Land use, from the latest update of the cadastre, in an example, at 2000 meter resolution.

Land use				
Region	The latest	The prior update	Baseline	
Región de O’ Higgins	2013	2005	2001	
Región del Maule	2016	2009	1999	
Región de Ñuble*	2015	-	-	
Región de Bío Bío	2015	2008	1997	
Región de la Araucanía	2017	2007	-	

Table 6: These regions fall within the specified range of latitudes. Updates typically occur in intervals of several years, and the processed information is usually released one year after the update is made. It is important to note that the Ñuble region, created in 2017, is a subdivision of the Bío Bío region. But the municipalities remain the same.

Land Use, by type of forest			
Type of Land Use	The prior update	The latest	Change
Urban and Industrial areas	228	370	+
Agricultural land	1,390	870	-
Grasslands and shrublands	3,993	2,543	-
Forests Plantations	3,163	3,482	+
Native Forests	1,418	2,987	+
Wetlands	52	7	-
Areas devoid of vegetation	1,258	1,289	+
Glaciers and snow	127	91	-
Water bodies	41	31	-

Table 7: Land Use Change in the Maule region, from the Conaf cadastre. Taking the difference between "the prior update" decade (2001-2012) and "the latest update" decade (2012-2023) at the pixel/observation level. The same methodology is applied per municipality to analyze changes in each type of Land Use.



Graphic 7: Sankey diagram of change in the Land-Use in the most relevant categories (Agricultural Land, Forest Plantations, Grasslands and shrubs and Native Forests). The red section indicates the agricultural land that is being converted into forest plantations.

4 Empirical Approach

In this section I try to model the phenomenon to understand it, going from particular to general as recommended by [Herrera \(2015\)](#). Therefore, I make that exercise first with a Spatial-Panel Model that accounts for temporal auto-correlation and spatial auto-correlation, which would give us a clue about the spillover effect ([Elhorst, 2014](#)) then will be added the error autocorrelation term as a robustness check. Then in the second part, I will study the Maule region in a cross-sectional way in specific interesting periods, during which there is clearly a clear-cutting. The goal is to estimate and infer the indirect effects, following the recommended interpretation made by [LeSage and Pace \(2009\)](#), and finally, I would use my toy dataset to build indicators for the land use change, by municipality. With the aim to test the heterogeneous effects in the income using the datasets CASEN (National Socioeconomic Characterization Survey) to capture at least the sign of the economic value of the calculated ecosystem service.

4.1 Spatial-Panel Approach

Beginning with the simplest model, an Ordinary Least Squares (OLS) approach would be insufficient to capture the spatial autocorrelation between units. Nevertheless the spillover effect would be zero if spatial dependence were not present.:

$$Y_i = \alpha + \beta X_i + \varepsilon_i \quad (1)$$

$$Y_i = (\text{NDVI})$$

$$X_i = (\text{Evapotranspiration, Humidity and Precipitation})$$

I can calculate for each period $t \in [1, 92]$ the same cross-sectional regression and then estimate Moran's I to measure spatial auto-correlation in that regression ([Moran, 1950](#)). Then, I will calculate the average to assess the significance of the indicator (I) across the entire panel of information. ([Beenstock and Felsenstein, 2019](#)).

$$I = \frac{N}{\sum_i \sum_j W_{ij}} * \frac{\sum_i \sum_j W_{ij} \hat{u}_i \hat{u}_j}{\sum_i \hat{u}_i^2}$$

Therefore, I calculated the average for the 92 periods:

$$\bar{I} = \frac{1}{T} * \sum_{t=1}^T I_t = 4,982.3771$$

Being with a value greater than 1, it suggests the presence of spatial auto-correlation. However, I need to test this against the null hypothesis that (I) is actually equal to zero. The z-value I need to contrast with the normal distribution corresponds to a critical value of 1.96.

$$z_{obs} = \frac{\bar{I}}{\bar{V}} N(0, 1) > 1.96$$

Where V is:

$$V^2 = \frac{N^2 * \sum_i \sum_j w_{ij}^2 + 3(\sum_i \sum_j w_{ij})^2 - N \sum_i (\sum_j w_{ij})^2}{T(N^2 - 1)(\sum_i \sum_j w_{ij})^2} = 6.15967 * 10^{-10}$$

$$\frac{4,982.3771}{2.48187 * 10^{-5}} > 1.96$$

$$200,751,047.5544 > 1.96$$

After the test we can reject the null hypothesis, so there is evidence of a spatial auto-correlation, so I will use the weight matrix presented in figure 4 in the Data section. The notation is W^k where $k = 1$ because of the size of the resolution, it's not going to be necessary more than one spatial lag.

The second option for the model would be a panel regression, which eliminates constant variables but would overestimate the time autocorrelation in this case. For instance, a preliminary attempt in the Maule region yielded a coefficient of 0.7596451 for the fourth time lag. A more realistic result will be presented in the next section (Figure 8):

$$Y_{i,t} = \alpha + \beta X_{i,t} + c_i + \varepsilon_{i,t} \quad (2)$$

The Spatial-Panel model is based on the spatial econometric approach to test this relationship in the Amazon forest by (Araujo et al., 2023) which is a Spatial Panel Model that accounts for spatial and temporal autocorrelation between the pixels. The following equation is presented, incorporating weather variable controls:

$$Y_{i,t} = \alpha + \rho Y_{i,t-1} + \sum_{k=1}^K \beta_k W_{i,t}^{[k]} Y_{i,t} + \beta X_{i,t} + \varepsilon_{i,t} \quad (3)$$

$$Y_{i,t} = (\text{NDVI})$$

$$X_{i,t} = (\text{Evapotranspiration, Humidity and Precipitation})$$

The model considers the persistence over seven temporal lags, which in our setup each period spans a quarter of a year. The summation accounts for multiple levels of the weight matrix, but I will only use $k = 1$, the term $X_{i,t}$ accounts for the controls, to mitigate the issue of omitted variable bias. The results are explained in the 5.1 section.

4.2 Cross-sectional Approach

To understand the dynamics within the large dataset, I propose testing the best model (3), with the addition of constant land use information. This will be applied across two periods and in the Maule region. As the one previously presented. The model in question would be:

$$y_i = \alpha + \beta_1 W_{i,t}^{[1]} y_{i,t} + \beta X_{i,t} + \gamma Z_{i,t} + \varepsilon_{it} \quad (4)$$

$$Y_i = (\text{NDVI})$$

$$X_i = (\text{Evapotranspiration, Humidity and Precipitation})$$

$$Z_i = (\text{Agricultural land, Grasslands and shrublands, Forests Plantations, Native Forests...})$$

To avoid multicollinearity, the category "Urban and Industrial areas" is excluded from the estimation. Consequently, the results will be interpreted as differences relative to cities.

4.3 Tests for economic value

Using the empirical strategy proposed by [Bonilla-Mejía and López \(2024\)](#) but in a cross-sectional way:

$$\ln_yauth_{i,m} = \alpha + \gamma \overline{NDVI_{Fi,m}} + \mu Agro_{i,m} + \theta(\Delta \overline{NDVI_{CPi,m}}) + \kappa(Agro_{i,m} * (\Delta \overline{NDVI_{CPi,m}})) + \beta X_{i,m} + \epsilon_{i,m} \quad (5)$$

$$(\Delta \overline{NDVI_{CPi,m}}) = \begin{cases} 0 & \text{if } \overline{NDVI_{CPi,m,t}} - \overline{NDVI_{CPi,m,t-1}} \geq 0, \\ 1 & \text{if } \overline{NDVI_{CPi,m,t}} - \overline{NDVI_{CPi,m,t-1}} < 0. \end{cases}$$

The difference with the original approach lies in the use of a binary variable $\Delta \overline{NDVI_{CPi,m}}$ which denotes if there was a decrease in the average amount of commercial forestry plantations at the municipality level. Here $\ln_yauth_{i,m}$ represents the natural logarithm of the household income i , living in municipality m . $\overline{NDVI_{Fi,m}}$ is a continuous variable, being the average NDVI value of total forest (Native and commercial plantations) in municipality m , $\mu Agro_{i,m}$ is a binary variable taking value 1 if the individual works in the agricultural sector and zero otherwise. Meanwhile $X_{i,m}$ are the control variables: gender, age, the quadratic of age, number of years of education, the quadratic of the number of years of education, a binary variable if the house is in a urban or rural area, the number of persons in the household, if the person is under the poverty line and if is working.

Additional, κ is the coefficient of interest, that represent the interaction between the binary variables of $Agro_{i,m}$ and $\overline{NDVI_{CPi,m}}$. The results are tested in the CASEN survey, for the years: 2013, 2015, 2017, 2022. and a pooled version of all of them, in section 5.3

5 Results analysis

In this section, I present the results obtained from the three methodologies outlined in the previous section. The first approach addresses the initial research question of this thesis, which investigates the existence and impact of a spillover effect from forests in promoting vegetation growth through local weather regulation. The second approach incorporates land use decisions to evaluate the varying impacts of different types of forests in the Maule region, using clear-cutting practices from 2016-2017 as a specific example. Finally, the evidence from the previous results supports the exploration of the second research question: assessing the economic value of forests and determining the sign of the ecosystem services they provide in local weather regulation. The results are organized as follows: first, I test the existence of this natural phenomenon in central Chile; second, I estimate the impacts resulting from clear-cutting and present a simulation of potential conservation strategies; and finally, I calculate the benefits of forests at the municipal level while examining the hypothesis of a mechanism involving the agricultural sector.

5.1 Results for the Spatial-Panel Approach

Therefore, the results from the Spatial-Panel approach by region are presented in Table 8. The panel approach removes any constants that may appear in the error term, such as soil type or distance to the nearest body of water. The main beta of the indirect impact of neighbors vegetation index values can be interpreted as the average spillover effect of a marginal increase by the first degree neighbors in the wind back trajectory. As the NDVI ranges from 0 to 1, only if the neighbors reach the hypothetical higher value in the range, I mean 1, as can be seen in Figure 9 (only forests get near that value), then the expected average impact would be of 0.36175 for example for the Maule region. The significance an positive values indicates the evidence of forests ability to promote vegetation growth through the water cycle process.

When comparing these results with the ones made in the Amazon Rainforest with the Leaf Area Index (LAI) ¹ as the dependant value and without any variable controls made by [Araujo et al. \(2020\)](#) they found that time-lags are the coefficient with the higher magnitude for the interpretation, a difference between evergreen amazon rain forest and the type of vegetation in Chile is seasonality, the higher time-lag in the Chilean case is the quarter a year ago, reflecting the difference in the type of vegetation, as can be seen in Figure 8.

Including the control variables might initially seem controversial due to potential endogeneity issues. However, the justification for their use is as follows: Precipitation appears random at the pixel level each quarter, while Humidity follows a consistent gradient from the coast to the interior of the country. Evapotranspiration also varies by region each quarter, and although more vegetation leads to higher evapotranspiration—potentially fueling this cycle—the data also indicate that croplands in the central valley exhibit significant evapotranspiration during the hotter summer quarters, despite

¹For further information on the difference between LAI and NDVI, please refer to Appendix 7.4

	O´higgins NDVI	Maule NDVI	Ñuble NDVI	Bio bío NDVI
PP normalized	0.01240*** (0.00068)	-0.00191*** (0.00051)	-0.00247** (0.00084)	0.00064 (0.00089)
ET normalized	-0.00468*** (0.00048)	0.00043 (0.00032)	-0.00128* (0.00058)	0.00372*** (0.00049)
HU normalized	0.03940*** (0.00119)	0.02206*** (0.00106)	0.03640*** (0.00173)	0.03446*** (0.00132)
W^1 NDVI	0.26177*** (0.00051)	0.36175*** (0.00050)	0.10808*** (0.00021)	0.25716*** (0.00042)
Constant	0.04676*** (0.00005)	0.05321*** (0.00004)	0.05995*** (0.00006)	0.06748*** (0.00005)
Time-lags	✓	✓	✓	✓
Pseudo R-squared	0.7586	0.6870	0.7165	0.6036
Observations	487390	991950	438685	877625
Periods	92	92	92	92

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 8: The results of estimating model 3 by region over the two-decade period from 2001 Q1 to 2023 Q4, incorporating control variables for Precipitation, Evapotranspiration, and Humidity, are normalized on a 0 to 1 scale to match the dependent variable, which also falls within this range. This normalization simplifies the interpretation of the results. Additionally, the time lags are visualized in Figure 8.

low NDVI values, showing that captures the amount of water released by irrigation system in croplands. Furthermore, trials using temperature data, along with example graphics, can be found in Appendices 7.5 and 7.6. Ultimately, these variables are included as they are enhance the conditions for vegetation growth.

Time lags of Table 8

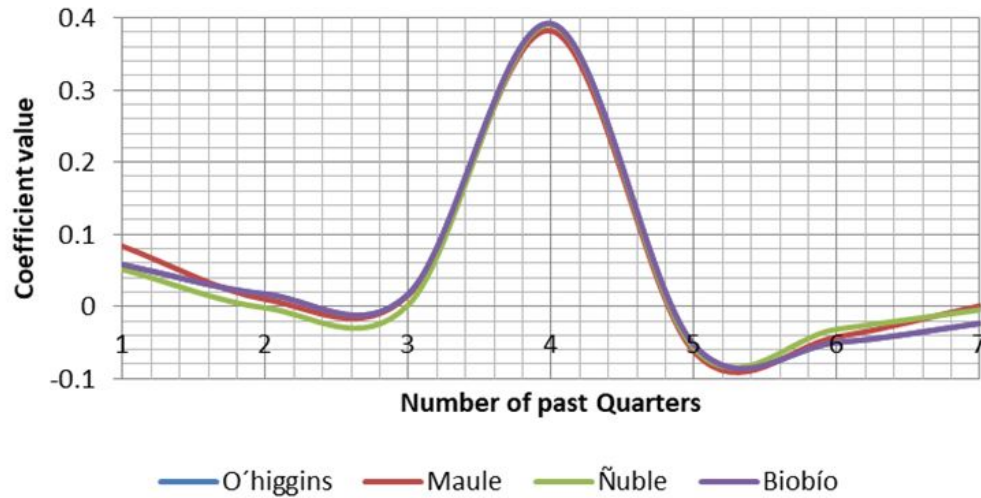


Figure 8: These are the time lags from Table 8, covering the four regions in the study, reveal a high and significant peak at the fourth time lag, demonstrating the persistence of pixel values from one year prior. This peak is consistently significant at the 0.01% level. In comparison to results from the Amazon forest—where the best predictor is the first time lag due to the evergreen nature. For the mixed and seasonal forests in this study, the fourth time lag proves to be a better predictor than the immediate past.

The positive and significant coefficients of NDVI for first-degree neighbors provide evidence of the natural phenomenon that forest units regulate local weather. The indirect impact, often referred to as "spillover," can be interpreted as the impact of neighbour vegetation. However, given the diversity of land uses involved, I remain cautious and at least accept the existence of correlation. This raises an important question: is the clustering of forest units a natural occurrence, a result of human decisions in their placement, or a combination of both? Specifically, in the case of commercial forest plantations, Table 9 offers insight. It presents a spatial-panel regression that includes a spatial autocorrelation term for the errors. This term is significant and carries substantial weight. The spatial autocorrelation coefficient of NDVI diminishes in almost every region except for Biobío, suggesting that an unobserved spatial factor is influencing the spatial distribution more than the vegetation itself.

A plausible explanation for Biobío's distinct result could be the presence of more native forests near the coast. In light of this unclear outcome, I will examine whether land use data provides further clarity. In the next section (4.2), I will delve deeper into this question, considering that the error term may reflect land use choices made by landowners. I will explore the assumption that land use remains relatively constant over at least a decade and propose an alternative approach.

	O´higgins NDVI	Maule NDVI	Ñuble NDVI	Biobío NDVI
PP normalized	0.04455*** (0.00169)	-0.00900*** (0.00118)	-0.01221*** (0.00206)	-0.00032 (0.00113)
ET normalized	-0.00078 (0.00110)	0.00574*** (0.00073)	0.01254*** (0.00140)	0.00792*** (0.00063)
HU normalized	0.04102*** (0.00294)	0.05181*** (0.00242)	0.08852*** (0.00418)	0.04921*** (0.00170)
W^1 NDVI	-0.00126 (0.00098)	0.03061*** (0.00110)	0.01194*** (0.00054)	0.18377*** (0.00111)
W^1 Error NDVI	0.33795*** (0.00089)	0.43402*** (0.00105)	0.12724*** (0.00056)	0.12118*** (0.00140)
Constant	0.04488*** (0.00005)	0.05169*** (0.00004)	0.05840*** (0.00006)	0.06707*** (0.00005)
Time-lags	✓	✓	✓	✓
Pseudo R-squared	0.9142	0.9246	0.8244	0.7256
Observations	487390	991950	438685	877625
Periods	92	92	92	92

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 9: The result of estimating model 3 by region, in the two decades span from 2001 Q1 to 2023 Q4, but with the difference with Table 8, is the incorporation of the autocorrelation term of the error in the row of W^1 Error NDVI the coefficients of NDVI remain significant but with a lower magnitude except by the O´higgins region which is not significant anymore.

5.2 Results for the Spatial Cross-sectional Approach

The reason for studying the dataset by period is to capture fixed or constant effects that may be overlooked in the Panel approach. Consider, for instance, the impact of land use, which reflects human decisions influencing what can grow in a given area. Assuming that land use remains constant and valid over a decade, is possible to analyze the relationship between land use and NDVI values. Specifically, this approach allows to highlight the differences between native and commercial forest plantations. Evenmore, it enables to identify minimal evidence of a naive difference before and after clear-cutting.

Figures 1 and 2 present the NDVI values for the Maule region for the fourth quarter of 2016 and the fourth quarter of 2017. In the lower-left corner, near the city of Constitución, a significant hollow can be observed, representing a clear-cutting of mature forest plantations, as depicted in Figure 2. Table 10 displays the results of model (5) applied four times: the first two columns correspond to 2016, and the last two columns correspond to 2017, highlighting the differences when estimating the fixed effects of Land Use, before and after the clear-cutting.

	2016q4	2016q4	2017q4	2017q4
	NDVI	NDVI	NDVI	NDVI
NDVI				
PP normalized	5.64927*** (0.14836)	3.30560*** (0.11283)	1.40831*** (0.11669)	0.85341*** (0.09085)
ET normalized	0.22250*** (0.04578)	0.06967 (0.03751)	0.49965*** (0.04890)	0.34874*** (0.03977)
HU normalized	0.91270*** (0.01669)	0.43638*** (0.01371)	0.84426*** (0.01596)	0.38424*** (0.01277)
Commercial Forests		0.03732*** (0.00644)		0.01713* (0.00707)
Native Forests		0.06504*** (0.00657) (0.00727)		0.08348*** (0.00726) (0.00801)
Constant	-0.70426*** (0.03715)	-0.10278*** (0.03121)	-0.57832*** (0.03827)	-0.13399*** (0.03182)
W				
NDVI	0.08985*** (0.00626)	0.11661*** (0.00465)	0.09281*** (0.00726)	0.13845*** (0.00520)
Pseudo	0.4952	0.7036	0.3993	0.6384
R-cuadrado	11670	11670	11670	11670

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 10: The results of the estimation of Model 4 for the periods of 2016 Q4 and 2017 Q4 in the Maule region, before and after clear-cutting, are presented here. The third and fifth columns display the outcomes while controlling for the fixed effects of land use types. Only the results for commercial forests and native forests are shown; however, these results were derived using all types of land use, excluding urban and industrial areas. Therefore, the results should be interpreted as the differences relative to these excluded land use types.

As expected, in the fourth quarter of 2016, the first-degree neighbor W^1 coefficient $\beta_1 = 0.11661$ is lower than the value obtained from the panel approach, but still significant. Consistent with the conclusions from the results in section 4.1, this finding suggests the presence of the natural phenomenon within this weight matrix.

When In the 2016 model, the Land Use data is incorporated as control binary variables, taking the value of one for specific types of land use and zero otherwise. The coefficients for Native Forests and Plantations are both significant, but the magnitude for Native Forests is notably higher. This suggests that native forests, by being more complex, with multiple layers of vegetation, usually higher NDVI values and being clustered, have a stronger spatial effect. Nevertheless, after the clear-cutting in the 2017 models, I observed that the forest's ability to self-regulate by influencing local weather conditions seemed to diminish for the commercial plantations. The previously observed relationships changed significantly, with the direction and magnitude of control variables deviating from the averages calculated in the results of section 4.1. However, upon adding fixed effects of land use, it became evident that the substantial spatial effect of

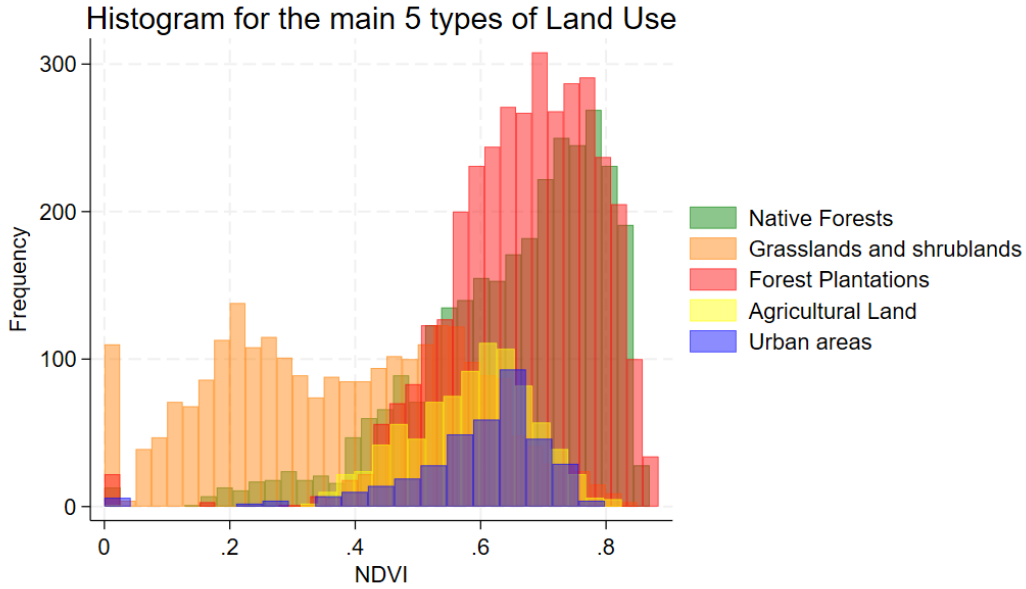


Figure 9: Frquency of the type of Land Use in 2017 Q4

first-degree neighbors could be attributed to the remaining native forests located on the right side of the map near the Andean mountains (See Figures 5 and 6). This is supported by the continued significance of the coefficient for Native Forests (0.08348^{***}), whereas the coefficient for Plantations only retains significance at the 5% level.

The impact of including the type of Land Use is noteworthy, as evidenced by the substantial increase in the Pseudo R2 value. This highlights the significance of incorporating this previously omitted variable into the analysis. As discussed in the introduction, it underscores the economic challenge in the human decision about land use, where utility maximization is constrained by natural factors. With these results in mind, is that for taking the decision, is necessary an amplified valuation of the ecosystem service of forests, the small simulation will explore that.

5.2.1 Simulation

Now, considering the simulation, let's explore a scenario in which a government aims to identify optimal locations for forest conservation that would positively impact productive units, such as agricultural land or plantations. To inform this decision, we can use NDVI values from the fourth quarter of 2016, both before clear-cutting and after the shock, as presented in Table 11. The data indicate that commercial forest plantations have a slightly higher average NDVI value compared to native forests but show a lower standard deviation. The simulation examines the hypothetical situation of what would happen if a pixel had remained intact instead of being cut down.

	time	Mean	Std. dev.	Minimum	Maximum
Native Forests	2016 Q4	.6672151	.1208077	0	.8837
Forest Plantations	2016 Q4	.6144844	.1383705	0	.8782
Native Forests	2017 Q4	.6484445	.1486165	0	.8695
Forest Plantations	2017 Q4	.6493789	.1576794	0	.8719

Table 11: It's presented for both type of forest, in the two periods of study the: the mean, standard deviation and the range of NDVI for Native Forests and Plantations.

The simulation examines the hypothetical scenario in which a forest unit cut in 2017 Q4 would have remained intact prior to clear-cutting. The selected point, marked by the yellow dot in Figure 8, reveals a small but significant global impact. The NDVI value before the cut was 0.8204, while it decreased to 0.3297 in a pixel of commercial plantations. Figure 8 illustrates the difference between the two predictions based on this small change at the described point. Interestingly, this change generates an expansion wave. The simulation suggests that preserving the observation with ID number 23577 as a forest could positively influence distant agricultural units in the direction of the wind. This approach could serve as a valuable tool for identifying which forest units significantly impact neighboring agricultural areas or could be applied in a randomized controlled trial conducted by the forestry industry.

Global Spillover Effect

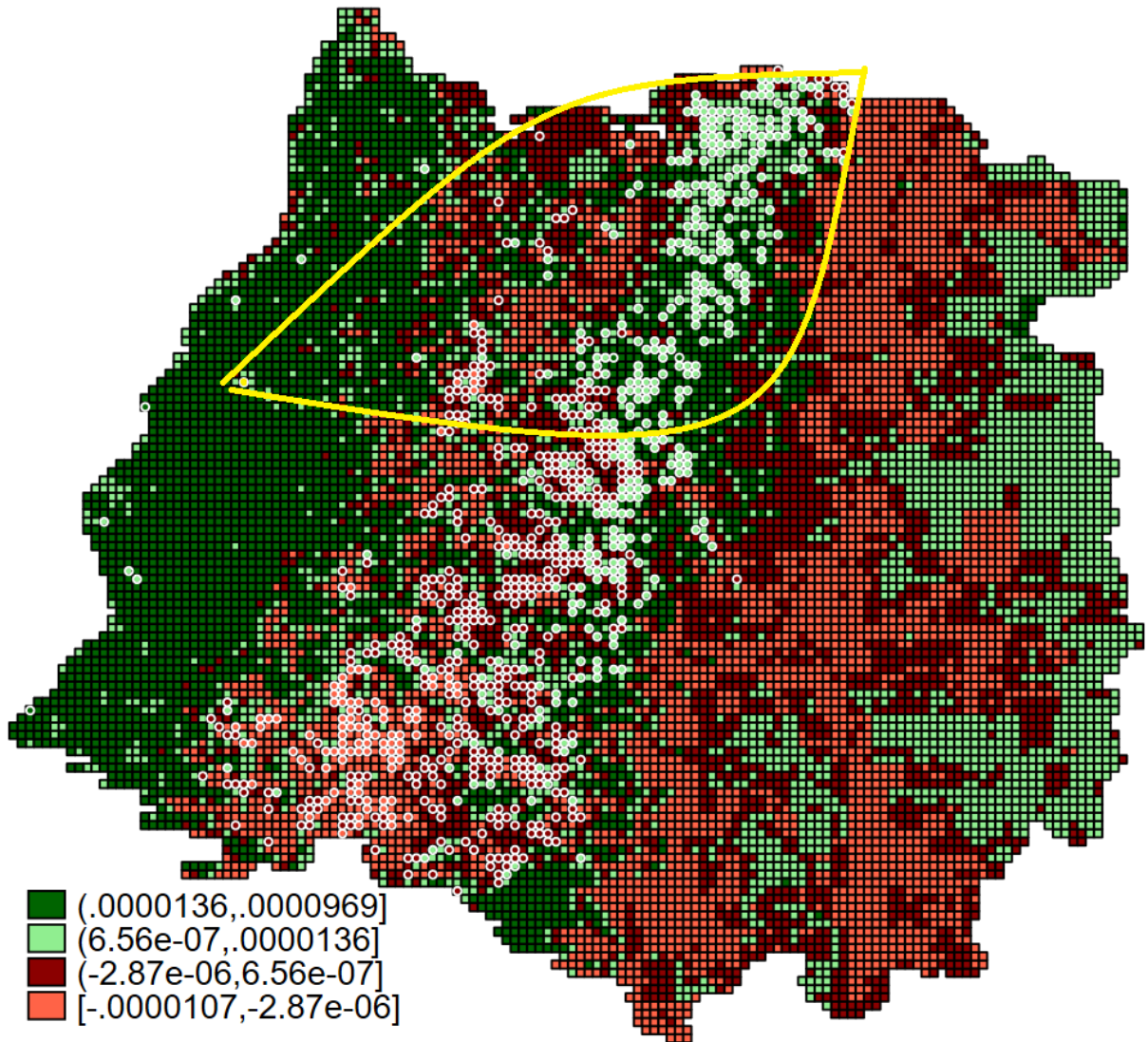


Figure 10: Global Spillover effect, the values are the difference in the prediction between before and after the change in the value of NDVI observation with ID's number 23577 in the Maule region in the year 2017 fourth quarter.

5.3 Results for the tryouts, for economic value

Finally, to study the economic value of native forests and commercial forest plantations, the identification approach involved analyzing municipality-level data on changes of levels in vegetation amount. Tables 12 and 13 present the results from estimating equation (5). In both tables the dependant variable is "ln_yauth" represents the natural logarithm of the autonomous income of the household.

Using the CASEN survey and land use change information from Table 7 at the municipality level, this analysis examines household income. Although the survey collects individual observations, household income is aggregated based on family members,

providing broader coverage since a household could be affected by the income of any family member. The control variables, as previously described, include the individual's gender, age, the quadratic term of age, years of education, the quadratic term of education, and the area of residence—coded as zero for rural areas and one for urban areas. The variable "number" represents the number of members in the household, while "lp" is a binary variable indicating poverty status: zero if the person is above the poverty line and one if otherwise. Additionally, "employed" is a binary variable, equal to one if the individual was working at the time of the interview. The complete results can be found in Appendix 7.8.

The objective of this section was to measure how the "NDVI Forests" that is the average NDVI of any type of forest vegetation in the municipality as an indicator of forest quantity impacts the household income. Higher NDVI values at the municipal level are correlated with higher income, a trend consistent across the last four CASEN surveys, as shown in Table 12. The "Pool" column aggregates these four surveys, with the estimation controlling for year-to-year variations in case of weather anomalies. These results assume that the individuals in the sample are similar to those in a random sample and that the variables are comparable. However, an issue with these results is the potential inflation that may contaminate the findings.

Regarding the mechanism of interest, the three variables in Table 12 aim to determine whether commercial forestry positively impacts income through weather regulation. The hypothesis posits that in municipalities experiencing a decrease in average NDVI—indicative of clear-cutting—individuals working directly or indirectly in the agricultural industry would see a decrease in their income. This impact is expected to be captured by the interaction between these two binary variables. This interaction would reflect the differences between municipalities that underwent clear-cutting and those that did not, comparing farmers to non-farmers. The results are ambiguous; while this hypothesis holds true for 2022—one of the driest years of the recent drought—other surveys show the opposite sign, suggesting that the negative externalities may outweigh the benefits of this ecosystem service.

	(2022)	(2017)	(2015)	(2013)	(Pool)
	ln_yauth	ln_yauth	ln_yauth	ln_yauth	ln_yauth
NDVI Forests	0.10270** (0.04784)	0.27540*** (0.05100)	0.26923*** (0.04085)	0.08373** (0.04155)	0.14109*** (0.02197)
Decrease NDVI of Commercial Forests	-0.04832*** (0.01041)	0.03017*** (0.00800)	-0.04059*** (0.00724)	-0.00794 (0.00737)	-0.02143*** (0.00381)
B_Farmer * De- crease NDVI of Commercial Forests	-0.06431** (0.03124)	0.06773*** (0.02131)	-0.01603 (0.01991)	0.03978* (0.02192)	0.02511** (0.01032)
Binary_Farmer	0.00319 (0.02845)	-0.04278** (0.01743)	-0.01060 (0.01690)	-0.03164*** (0.01136)	-0.04057*** (0.00802)
Constant	12.22845*** (0.04532)	11.75730*** (0.04836)	11.79281*** (0.03877)	11.76584*** (0.03929)	11.79599*** (0.02128)
Control variables	✓	✓	✓	✓	✓
Control by year	X	X	X	X	✓
Control by region	✓	✓	✓	✓	✓
Pseudo R-squared	0.3855	0.3878	0.4288	0.4953	0.4398
Number of observa- tions	42584	46753	56834	47452	193623
Municipalities	98	98	98	98	98

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 12: Results of model number five with the CASEN survey, but controlling with the region level, in the CASEN surveys of 2022, 2017, 2015 and 2013, and also an extension with a pool of the four surveys.

To incorporate the municipality level, the key difference between Table 12 and Table 13 is that the former controls for regions, while the latter controls directly for municipalities. Both tables compare against the category with more observations Table 12 compares to the Maule region, while Table 13 compares to the municipality of Rancagua in the O'Higgins region. Due to multicollinearity, the average NDVI by municipality variable was lost, but the results remain robust, with consistent signs. The pooled results suggest that, on average, a decrease in the NDVI of commercial forests benefits farmers. However, 2022 remains a significant outlier worth further investigation. This positive relationship between a decrease in average NDVI and income may be influenced by factors such as land use changes to farmlands, road expansions, or other underlying processes.

	(2022)	(2017)	(2015)	(2013)	(Pool)
	ln_yauth	ln_yauth	ln_yauth	ln_yauth	ln_yauth
Decrease NDVI of Commercial Forests	0.12066** (0.05571)	-0.33627*** (0.06388)	0.00124 (0.04382)	-0.42795*** (0.07686)	-0.00538 (0.00422)
B_Farmer * De- crease NDVI of Commercial Forests	-0.06010* (0.03280)	0.06827*** (0.02219)	-0.02853 (0.02119)	0.06806*** (0.02307)	0.03002*** (0.01039)
Binary_Farmer	0.00172 (0.02979)	-0.03018* (0.01785)	0.00343 (0.01773)	-0.03215*** (0.01165)	-0.03719*** (0.00808)
Constant	12.30483*** (0.06196)	12.04828*** (0.03183)	12.19142*** (0.02903)	11.99639*** (0.02795)	12.05892*** (0.01545)
Control variables	✓	✓	✓	✓	✓
Control by year	X	X	X	X	✓
Control by munici- pality	✓	✓	✓	✓	✓
Pseudo R-squared	0.3978	0.3980	0.4387	0.5035	0.4457
Number of observa- tions	42584	46753	56834	47452	193623
Municipalities	98	98	98	98	98

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 13: Results of model number five with the CASEN survey, but controlling with the municipality level, in the CASEN surveys of 2022, 2017, 2015 and 2013, and also an extension with a pool of the four surveys.

6 Conclusion

The benefits of forests in terms of local climate regulation stem from their ability to transport moisture into the continent’s interior, their role in the water cycle, and their capacity to lower temperatures. Although this ecosystem service is challenging to quantify, methods have been developed to measure both the existence and magnitude of this effect. However, accurately determining its economic value remains difficult.

This thesis builds on the evidence and models developed for the Amazon rainforest and applies them to Chile’s south-central region. This area is characterized by extensive commercial forests and small remnants of native forests, with varying land uses. Since most forests are concentrated along the coast, the removal or change in land use of commercial plantations—the second-best provider of this ecosystem service after native forests—would result in a loss of these benefits. Finally, the economic value of this service is assessed through the agricultural industry.

A spatial and temporal model was used to measure the average effect per region over the past two decades, within the latitudes of 30° and 38° south in Chile, at a resolution of 2000 meters. This model aimed to capture the effect of vegetation in

itself, using wind direction measured through the spatial matrix W . The results were statistically significant at the 0.1% level, with a low magnitude ranging between 0.1 and 0.4. This persistence within that range also highlighted seasonal variability, as the forest is not completely evergreen. However, tests for error autocorrelation suggest the presence of an additional, larger unmeasured effect—except in the Biobío region. Where, the phenomenon remains relevant, possibly due to the greater concentration of native forest remnants along the coast compared to the other three regions.

Regarding what this unmeasured autocorrelation could be, a plausible option could be the land use decision. For this reason, testing in two periods before and after deforestation, where it is verified that the spatial effect is lost due to this shock and that more stable land use as the Native forests can have a better effect in determining the level of vegetation and the land use of the neighboring units.

About the economic value, while a positive relationship is found between municipalities with higher average forest vegetation levels and higher household incomes, the mechanism through the agricultural sector remains unclear. It is also uncertain whether the negative externalities of commercial forests outweigh the positive ones, though the results suggest that this may be the case.

Lastly, it remains to be determined what distinguishes the Biobío region from the other three, allowing it to maintain the magnitude of spatial autocorrelation. A plausible solution would be testing at different levels of resolution to address the Modifiable Areal Unit Problem (MAUP) strengthens the robustness of the results but demands greater computational capacity. Additionally, questions regarding historical average wind direction remain unresolved, with the potential for non-parametric models to better capture the connections between pixels.

In closing, the results confirm the existence of this phenomenon at varying magnitudes, highlighting that native forests are the optimal land use choice to sustain this positive externality. The role of commercial forests as a second-best option remains uncertain. However, the methodologies used in the simulation could be expanded by incorporating databases with direct information from agricultural fields adjacent to planted forests. Additionally, conducting randomized controlled experiments could offer valuable insights for measuring this effect. This remains a promising area for future research.

References

- Abatzoglou, J. T., Dobrowski, S. Z., Parks, S. A., and Hegewisch, K. C. (2018). Ter-
raclimate, a high-resolution global dataset of monthly climate and climatic water
balance from 1958–2015. *Scientific Data*, 5(170191).
- Araujo, R., Assunção, J., Hirota, M., and Scheinkman, J. (2023). Estimating the
spatial amplification of damage caused by degradation in the amazon. *Proceedings
of the National Academy of Sciences*, 120(46).

- Araujo, R., Costa, F., and Sant´Anna, M. (2020). Efficient forestation in the brazilian amazon.
- Aybar, C., Wu, Q., Bautista, L., Yali, R., and Barja, A. (2020). rgee: A r package for interacting with google earth engine. *The Journal of Open Source Software*, 5(51):2272.
- Balocchi, F., Galleguillos, M., Rivera, D., Stehr, A., Arumi, J. L., Pizarro, R., Garcia-Chevesich, P., Iroum´e, A., Armesto, J., Herv´e-Fern´andez, P., Oyarz´un, C., Barri´a, P., Little, C., Mancilla, G., Y´epez, S., Rodriguez, R., White, D. A., Silberstein, R., Neary, D., and Ram´ırez de Arellano, P. (2022). Forest hydrology in chile: Past, present, and future. *Journal of Hydrology*, 616.
- Banerjee, O., Cicowicz, M., Macedo, M., Malek, Z., Verburg, P. H., Goodwin, S., Vargas, R., Rattis, L., Brando, P. M., Coe, M. T., Neill, C., and Marti, O. D. (2021). An amazon tipping point: The economic and environmental fallout. *IDB Working Paper Series*, 1(1):1–103.
- Beenstock, M. and Felsenstein, D. (2019). The econometric analysis of non-stationary spatial panel data. *Springer*, pages 1–275.
- Boisier, J. P., Ciais, P., Ducharne, A., and Guimberteau, M. (2015). Projected strengthening of amazonian dry season by constrained climate model simulations. *Nature Climate*, 5(7):656–660.
- Bonilla-Mejía, L. and L´opez, E. (2024). Downsides of green industries: socioeconomic impact of commercial forestry in chile.
- Castle, S. E., Miller, D. C., Merten, N., Ordonez, P. J., and Baylis, K. (2022). Evidence for the impacts of agroforestry on ecosystem services and human well-being in high-income countries: a systematic map. *Environmental Evidence*, 11(10).
- Cheng, A. T., Sims, K. R., and Yi, Y. (2023). Economic development and conservation impacts of china’s nature reserves. *Journal of Environmental Economics and Management*, 121:102848.
- Clara Zemp, D., Schleussner, C.-F., Barbosa, H. M. J., Hirota, M., Montade, V., Sampaio, G., Staal, A., Wang-Erlandsson, L., and Ramming, A. (2017). Self-amplified amazon forest loss due to vegetation-atmosphere feedbacks. *Nature Communications*, 8(1):14681.
- Copernicus (2023). Copernicus, europe’s eyes on earth, land monitoring system. Accessed: 2024-06-30.
- Depetris-Chauvin, N., Fern´andez Olmos, M., Hallak, J. C., and Mosquera, J. S. (2023). Quality, vertical integration and adaptability. *Working paper, Red Nacional de Investigadores en Economía (RedNIE)*, 221.
- Didan, K. (2021). Modis/terra vegetation indices 16-day l3 global 500m sin grid v061 [data set]. *NASA EOSDIS Land Processes Distributed Active Archive Center.*, Accessed 2024-03-10:from <https://doi.org/10.5067/MODIS/MOD13A1.061>.

- Donaldson, D. and Storeygard, A. (2016). The view from above: Applications of satellite data in economics. *Journal of Economic Perspectives*, 30(4):171–198.
- Elhorst, J. P. (2014). Spatial panel models. *Handbook of regional science*, 3.
- Farrokhi, F., Kang, E., Pellegrina, H., and Sotelo, S. (2023). Deforestation: A global and dynamic perspective. (6).
- Fetzer, T. and Marden, S. (2017). Take what you can: property rights, contestability and conflicts. *The Economic Journal*, 127(601):757–783.
- Foster, A. D. and Rosenzweig, M. R. (2003). Economic growth and the rise of forests. *The Quarterly Journal of Economics*, 118(2):601–637.
- Heilmayr, R., Echeverría, C., Fuentes, R., and Lambin, E. (2016). A plantation-dominated forest transition in chile. *Applied Geography*, 75:71–82.
- Herrera, M. (2015). Econometría espacial usando stata. breve guía aplicada para datos de corte transversal. *Conicet-Ielde*, 1:1–55.
- Hirota, M., Holmgren, M., Van Nes, E., and Scheffer, M. (2011). Global resilience of tropical forest and savanna to critical transitions. *Science*, 334(6053):232–235.
- Kalinin, A. V., Sims, K. R., Meyer, S. R., and Thompson, J. R. (2023). Does land conservation raise property taxes? evidence from new england cities and towns. *Journal of Environmental Economics and Management*, 119:102782.
- Laurans, Y., Rankovic, A., Billé, R., Pirard, R., and Mermet, L. (2013). Use of ecosystem services economic valuation for decision making: Questioning a literature blindspot. *Journal of Environmental Management*, 119:208–2019.
- LeSage, J. and Pace, R. K. (2009). Introduction to spatial econometrics. *Champan and Hall/CRC*.
- Little, C., Lara, A., McPhee, J., and Urrutia, R. (2009). Revealing the impact of forest exotic plantations on water yield in large scale watershed in south-central chile. *Journal of Hydrology*, 374:162–170.
- Makarieva, A. and Gorshkov, V. (2007). Biotic pump of atmospheric moisture as driver of the hydrological cycle on land. *Petersburg Nuclear Physics Institute*, 3:1–21.
- McNally, A. (2018). Fldas noah land surface model l4 global monthly 0.1 x 0.1 degree (merra-2 and chirps), greenbelt, md, usa. *Goddard Earth Sciences Data and Information Services Center (GES DISC)*, Accessed 2024-03-10:from: doi:10.5067/5NHC22T9375G.
- Moran, P. A. (1950). Notes on continuous stochastic phenomena. *Biometrika*, 37(1/2):17–23.
- Murtazashvili, I., Murtazashvili, J., and Salahodjaev, R. (2019). Trust and deforestation: A cross-country comparison. *Forest Policy and Economics*, 101:111–119.

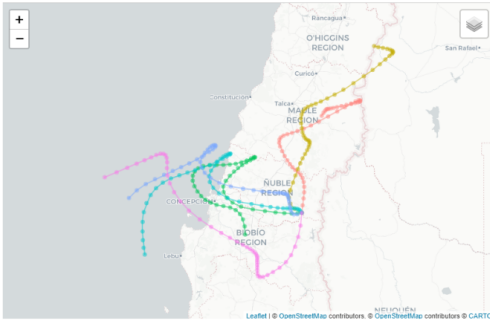
- NASA (2000). Normalized difference vegetation index (ndvi). Accessed: 2024-06-30.
- Piazza, A. and Roy, S. (2019). Irreversibility and the economics of forest conservation. *Economic Theory*, 69:667–711.
- Poblete Hernández, P., Gysling Caselli, J., Álvarez González, V., Bañados Munita, J. C., Kahler González, C., Pardo Velásquez, E., Soto Aguirre, D., and Baeza Rocha, D. (2023). Anuario forestal 2023. *Instituto Forestal, Chile.*, (192):280.
- Raga, F. (2009). The chilean forestry sector and associated risks. *Trebol*, pages 10–19.
- Ruiz, V., Munizaga, J., and Salazar Burrows, A. (2017). Forest plantations and its extension towards urban areas in metropolitan valparaiso and associations with the increase of wildfires. *Investigación Geografía Universidad de Chile*, 54:23–40.
- Sagoff, M. (2011). The quantification and valuation of ecosystem services. *Ecological Economics*, 70(3):497–502.
- Shukla, J., Nobre, C., and Seller, P. (1990). Amazon deforestation and climate change. *Science*, 247(4948):1322–1325.
- Sing, B., Jeganathan, C., Rathore, V., Atkinson, P., Behera, M., Singh, C., Dash, J., and Roy, P. (2023). The relationship between central indian terrestrial vegetation and monsoon rainfall distributions in different hydroclimatic extreme years using time-series satellite data. *Theoretical and Applied Climatology*, 155(1):45–69.
- Smith, C., Baker, J., and Spracklen, D. (2023). Tropical deforestation causes large reductions in observed precipitation. *Nature*, 615(7951):270–275.
- Spracklen, D., Arnold, S., and Taylor, C. (2012). Observations of increased tropical rainfall preceded by air passage over forests. *Nature*, 489(7415):282–285.
- Staal, A., Koren, G., Tejada, G., and V Gatti, L. (2023). Moisture origins of the amazon carbon source region. *Environmental Research Letters*, 18(4):044027.
- Staver, A. C., Archibald, S., and Levin, S. A. (2011). The global extent and determinants of savanna and forest as alternative biome states. *Science*, 334(6053):230.232.
- Stein, A. F., Draxler, P., Rolph, G. D., Stunder, B., Cohen, M. D., and Ngan, F. (2015). Noaa’s hysplit atmospheric transport and dispersion modeling system. *American Meteorological Society*.
- Teixeira, A., Silveira, B., Leroy, J., Medeiros, G., and Börner, J. (2021). Deforestation reduces rainfall and agricultural revenues in the brazilian amazon. *Natural Communications*, 12(2591).
- Verbesselt, J., Umlauf, N., Hirota, M., Holmgren, M., Van Nes, E., Herold, M., Zeileis, A., and Scheffer, M. (2016). Remotely sensed resilience of tropical forests. *Nature Climate Change*, 6(11):1028–1031.

7 Appendix

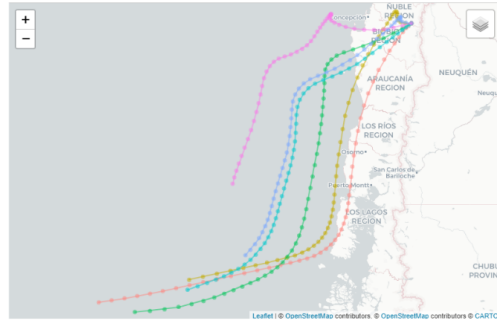
7.1 Back trajectories average

To calculate the average back trajectory for central-south Chile, I deliberately selected three specific points: (-35.00,-71.00); (-37.00,-71.50); (-39.00,-72.00) Then, using the program 'SplitR' and the HYSPLIT model, I calculated the back trajectories for the fifth and sixth day of every first month, for each quarter of the year, aiming to capture the seasonal variations (in the southern hemisphere). Additionally, I calculate the average of the back trajectories with four specific years, chosen deliberately for analysis: 2001; 2008; 2016; 2023. (I'm averaging 48 datasets, with 294 observations each one)

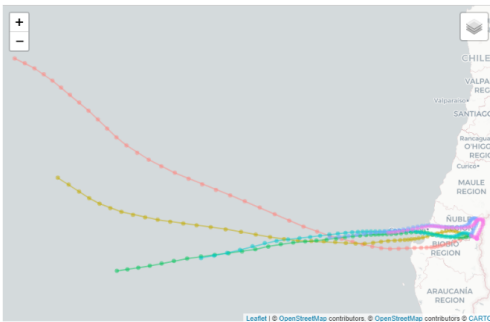
For example, this is how is built the average for the point (-37.00, -71.50) and looks like this in 2023:



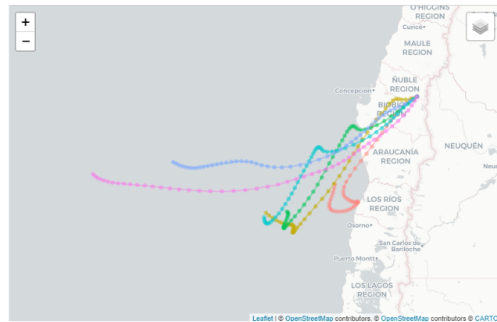
First quarter of 2023 on January 5th and 6th



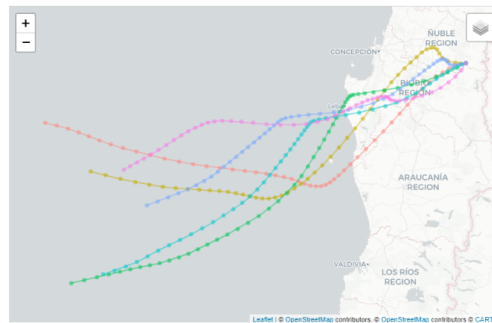
Second quarter of 2023 on April 5th and 6th



Third quarter of 2023 on July 5th and 6th



Fourth quarter of 2023 on October 5th and 6th



Average of 2023

Figure 11: Example of the construction of the average of the back trajectory.

7.2 Exploratory investigation

In Figure (1), I present exploratory evidence of this phenomenon. The blue lines depict annual precipitation data from various stations of the Dirección Meteorológica de Chile (from the DGAC: Dirección General De Aeronáutica Civil) spanning central-south Chile, with lighter shades representing northern regions and darker shades representing southern regions. Notably, as one moves from north to south, there is a general increase in annual precipitation, with the exception of the Valdivia station, located near the coast and surrounded by national parks with native forests.

Of particular interest is the peculiar correlation observed since 2000 between the red and orange lines, representing the annual production of wood and furniture in Chile, respectively. The red line depicts nominal prices, while the orange line shows prices adjusted for inflation (data sourced from the Central Bank of Chile). Notably, two peaks in wood production coincide with dry years characterized by below-average precipitation. To achieve an increase in wood production, the clearcutting of mature trees is necessary. Consequently, the loss of commercial forests appears to be inversely correlated with decreases in annual precipitation. Notable coincidences occurred in 2007 and 2016.

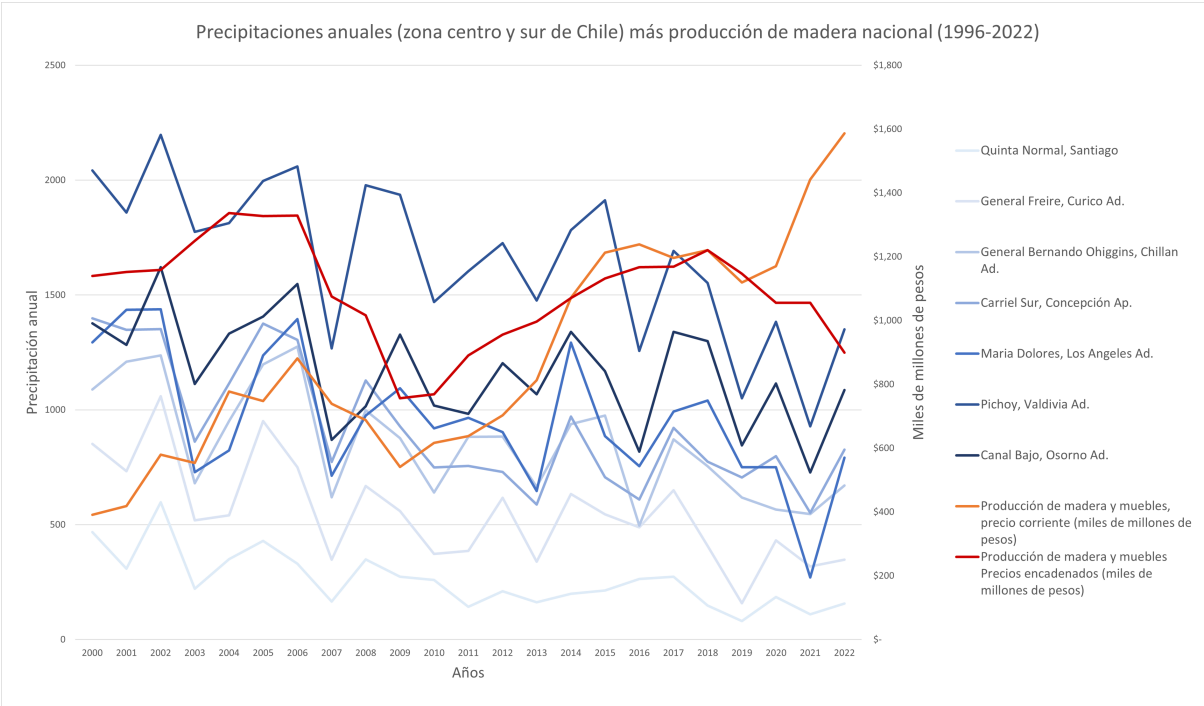


Figure 12: Original creation, with data from DGAC and the Central Bank of Chile

7.3 Land Use

Refer to Figure 18 to view the complete map of Land Use, based on the latest cadastre information from CONAF, within the study range.

7.4 Difference between LAI (Leaf Area Index) and NDVI (Normalized Difference Vegetation Index):

The reason for choosing the NDVI (The Normalized Difference Vegetation Index) as the vegetation indicator, was made after weighing the following information:

Both satellite remote sensing indices are used to measure and monitor vegetation on a large scale, and the values produced by both indices are typically correlated. However, the methods by which they measure vegetation are quite different. LAI, being the older method, and its most used definition ([Copernicus, 2023](#)) is that it measures half of the total area of leaves, since it only takes the image from above and not from other angles. It is also a good indicator of leaf density:

$$LAI = \text{Leaf Area } m^2 / \text{Floor Area } m^2$$

However, one problem that the index faces is that it only takes one layer of leaves, so the tree canopy covers other plant layers, It is also sensitive to seasonality. In summary, it is just an index of the density of the vegetation cover.

On the other hand, NDVI takes advantage of light waves, which plants reflect when photosynthesis (from 700 to 1100 nm being red and almost infrared light waves). So the greater the concentration of plants, the greater the reflection of these waves, that are captured by the sensors. It is defined as the normalized difference because it is the subtraction between the reflection of visible light and near-infrared light. The more visible light that is reflected, the lower the density of plants, and the opposite, the lower the reflection of visible light, means the greater the reflection of almost infrared light, which is why there would be a higher density of vegetation.

$$NDVI = (NIR - VIS) / (NIR + VIS)$$

Being: NIR (Near Infrared) and VIS (Visible) The Index ranges between -1 to 1. Values close to 1 indicate the densest possible vegetation, while a value of zero signifies no vegetation. Values close to -1 indicate other types of surfaces, such as water, snow, and clouds ([NASA, 2000](#)).

However, NDVI presents challenges when estimating vegetation in urban areas, as the level of reflection is compromised. However, it is preferable to use NDVI as an indicator of vegetation health. NDVI is more complex and goes beyond merely quantifying the number of leaves; In its favor, is already a normalized index and is the one that more comprehensively captures vegetation compositions.

Source links: [LAI](#) [NDVI](#)

7.5 Example of the data and regression with the temperature levels

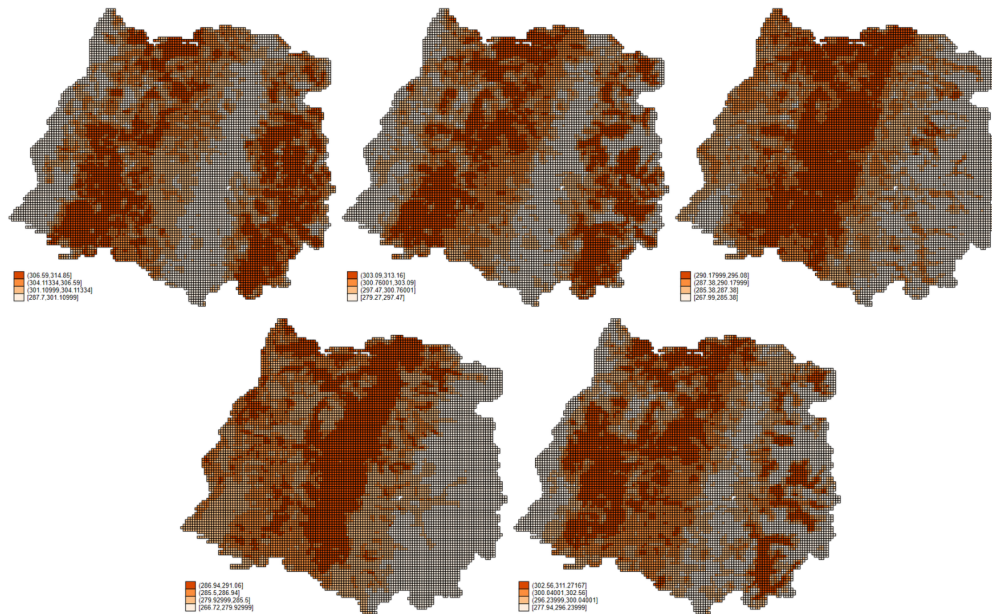


Figure 14: Exmple of the data in the Maule region with the Land Surface Temperature Data, from left to right, in the periods of 2016q4, 2017q1, 2017q2, 2017q3 and 2017q4

Another control variable considered for the spatial-panel model was Land Surface Temperature (LST), which shows a negative correlation with the amount of vegetation, resulting in a negative average impact value of LST in the pixel. However, further investigation is required due to potential endogeneity issues. The question arises whether lower temperature values create better conditions for vegetation growth, or whether large amounts of vegetation, such as forests, are responsible for reducing temperature. My main concern is the latter, but this relationship requires more study.

	O´higgins NDVI	Maule NDVI	Ñuble NDVI	Biobío NDVI
PP normalized	0.00900*** (0.00068)	-0.00241*** (0.00051)	-0.00437*** (0.00086)	0.00254** (0.00088)
ET normalized	0.02798*** (0.00073)	0.00174*** (0.00047)	0.01060*** (0.00094)	0.00832*** (0.00086)
HU normalized	0.05512*** (0.00122)	0.02329*** (0.00111)	0.04894*** (0.00186)	0.03247*** (0.00138)
LST normalized	-0.07552*** (0.00130)	-0.00268** (0.00084)	-0.03124*** (0.00197)	-0.01028*** (0.00208)
W^1 NDVI	0.26324*** (0.00051)	0.37105*** (0.00051)	0.10482*** (0.00022)	0.10486*** (0.00016)
Constant	0.04642*** (0.00005)	0.05302*** (0.00004)	0.06051*** (0.00007)	0.06663*** (0.00005)
Time-lags	✓	✓	✓	✓
Pseudo R-squared	0.7796	0.3645	0.7213	0.6972
Observations	486795	991015	432055	869125
Periods	92	92	92	92

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 14: The result of estimating model 3 by region, in the two decades span from 2001 Q1 to 2023 Q4, but with the difference with Table 8, is the incorporation of the contro variable Land Surface Temperature (LST).

7.6 Graphics of example of the Data Set for the other variables

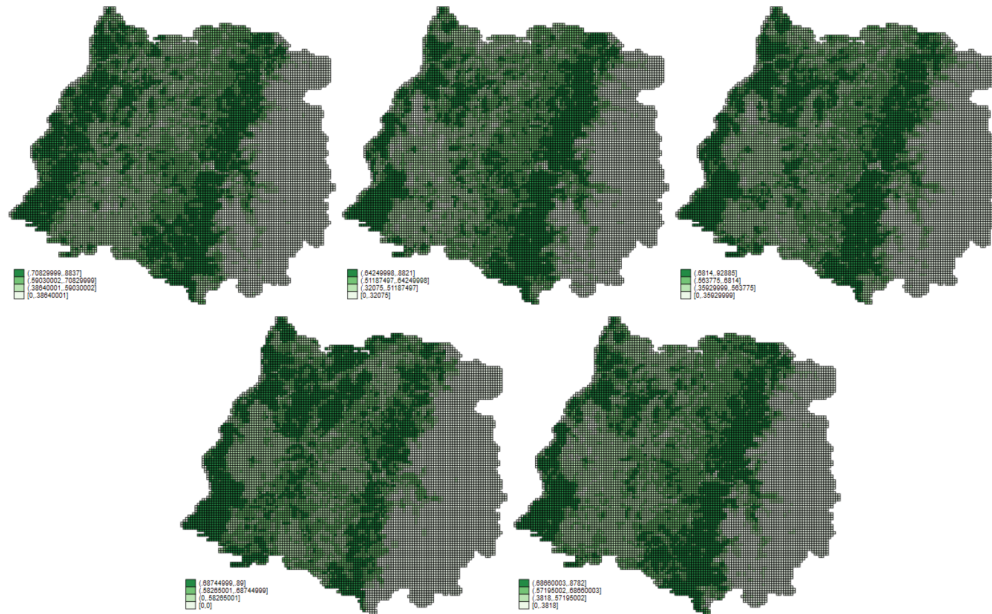


Figure 15: Exmple of the data in the Maule region with the NDVI Data, from left to right, in the periods of 2016q4, 2017q1, 2017q2, 2017q3 and 2017q4

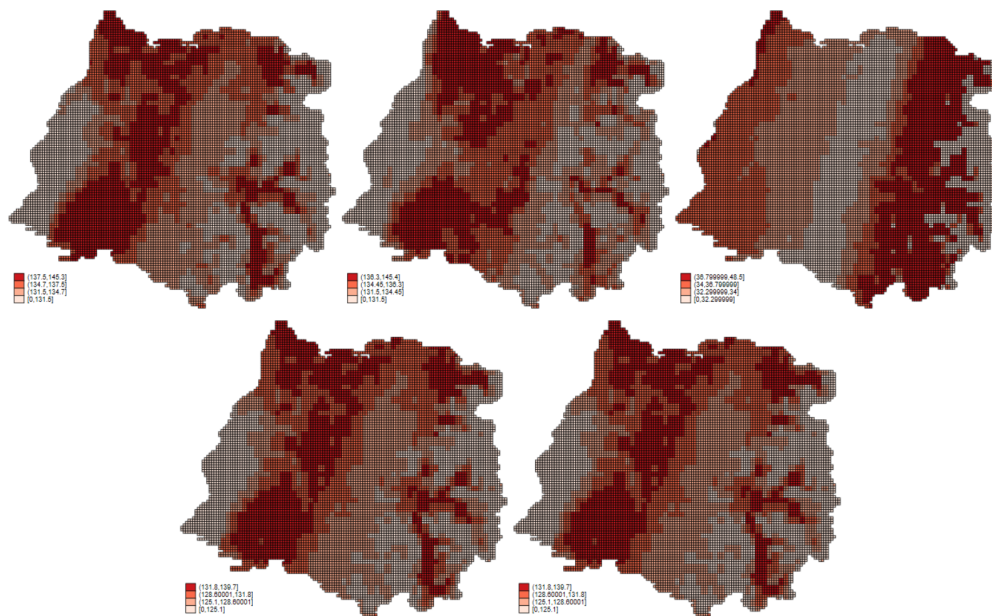


Figure 16: Exmple of the data in the Maule region with the Evapotranspiration Data, from left to right, in the periods of 2016q4, 2017q1, 2017q2, 2017q3 and 2017q4

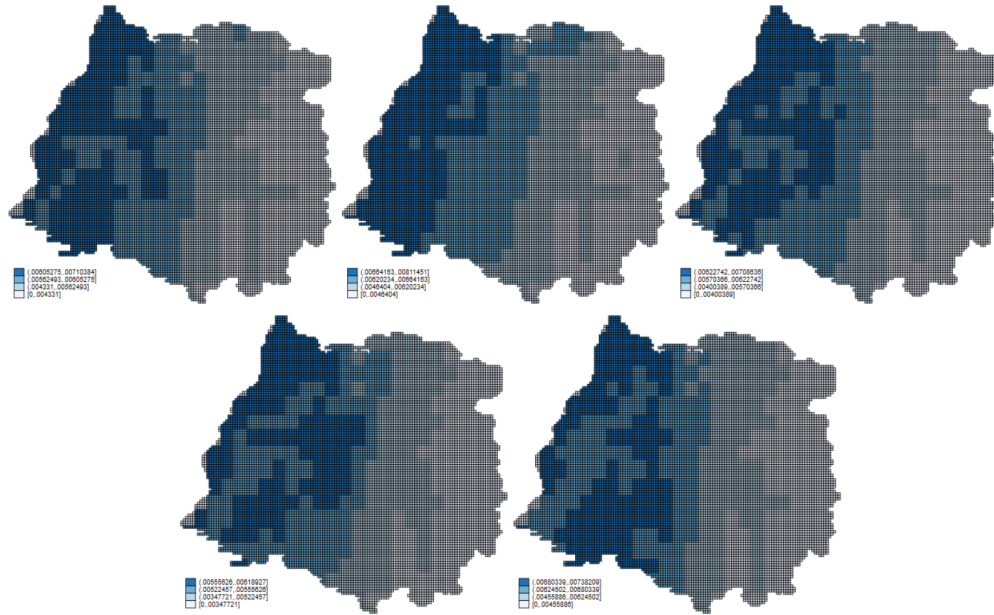


Figure 17: Exmple of the data in the Maule region with the Humidity Data, from left to right, in the periods of 2016q4, 2017q1, 2017q2, 2017q3 and 2017q4

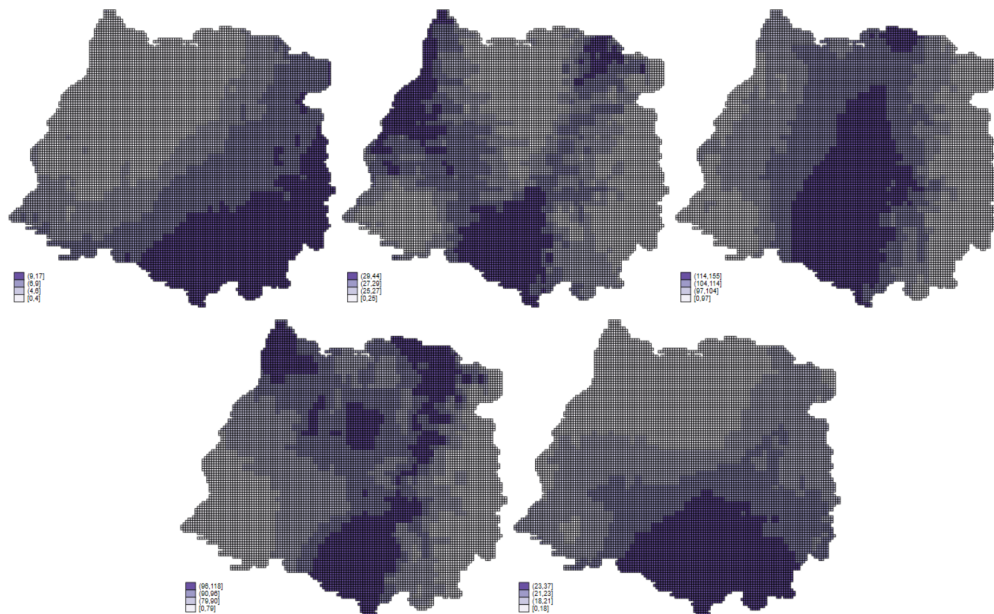


Figure 18: Exmple of the data in the Maule region with the Precipitation Data, from left to right, in the periods of 2016q4, 2017q1, 2017q2, 2017q3 and 2017q4

7.7 Regression by type of forest

The following tables present the raw results from regressions conducted separately for each type of forest: Native Forests and Commercial Plantations, using the

latest land use data from the SIT. The spatial-panel regression focuses on the last decade, from 2011 to 2023. The results indicate that the average impact of Native Forests is slightly higher than that of Commercial Plantations. However, these results may be imperfect due to the presence of isolated pixels ("islands") in the data—pixels without neighboring observations—resulting from only selecting the sample of pixels containing these types of forests.

	Commercial Plantations	Native Forests
	NDVI	NDVI
W^1 NDVI	0.07911*** (0.00023)	0.09504*** (0.00023)
Constant	0.04692*** (0.00004)	0.07347*** (0.00007)
Time-lags	✓	✓
Pseudo R-squared	0.3922	0.9246
Observations	569037	991950
Periods	56	56

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 15: The result of estimating model 3 by type of forests in the four regions, in the last decades span from 2011 Q1 to 2023 Q4. But also the spatial-panel regression was made with seven time-lag.

7.8 Results Tables of the economic value section

Results when controlling by the region:

	(2022)	(2017)	(2015)	(2013)	(Pool)
	ln_youth	ln_youth	ln_youth	ln_youth	ln_youth
Gender	0.02590*** (0.00750)	-0.00135 (0.00725)	0.00563 (0.00648)	0.00680 (0.00635)	0.00756** (0.00345)
Age	-0.00151 (0.00109)	0.00066 (0.00101)	0.00144 (0.00091)	0.00265*** (0.00092)	0.00089* (0.00049)
Age2	0.00002* (0.00001)	0.00001 (0.00001)	-0.00001 (0.00001)	-0.00001 (0.00001)	0.00000 (0.00001)
Years of education	-0.01526*** (0.00352)	0.00262 (0.00358)	-0.00300 (0.00312)	-0.00397 (0.00299)	-0.00520*** (0.00164)
Years of education2	0.00353*** (0.00016)	0.00292*** (0.00018)	0.00316*** (0.00016)	0.00339*** (0.00015)	0.00327*** (0.00008)
Number	0.21478*** (0.00249)	0.19089*** (0.00229)	0.18050*** (0.00211)	0.18840*** (0.00194)	0.19136*** (0.00110)
Employed	0.37183*** (0.00869)	0.29070*** (0.00832)	0.23490*** (0.00742)	0.21423*** (0.00727)	0.27442*** (0.00396)
LP	-1.31876*** (0.01867)	-1.13454*** (0.01292)	-1.17567*** (0.01007)	-1.13731*** (0.00883)	-1.16999*** (0.00571)
Rural	-0.11591*** (0.00887)	-0.10304*** (0.00869)	-0.06111*** (0.00759)	-0.08317*** (0.00769)	-0.08926*** (0.00410)
$\overline{\text{NDVI}}$ Forests	0.10270** (0.04784)	0.27540*** (0.05100)	0.26923*** (0.04085)	0.08373** (0.04155)	0.14109*** (0.02197)
Decrease $\overline{\text{NDVI}}$ of C.F.	-0.04832*** (0.01041)	0.03017*** (0.00800)	-0.04059*** (0.00724)	-0.00794 (0.00737)	-0.02143*** (0.00381)
B_Farmer * De- crease $\overline{\text{NDVI}}$ of C.F.	-0.06431** (0.03124)	0.06773*** (0.02131)	-0.01603 (0.01991)	0.03978* (0.02192)	0.02511** (0.01032)
Binary_Farmer	0.00319 (0.02845)	-0.04278** (0.01743)	-0.01060 (0.01690)	-0.03164*** (0.01136)	-0.04057*** (0.00802)
Year 2015					0.03468*** (0.00451)
Year 2017					0.08614*** (0.00468)
Year 2022					0.36836*** (0.00537)
Constant	12.22845*** (0.04532)	11.75730*** (0.04836)	11.79281*** (0.03877)	11.76584*** (0.03929)	11.79599*** (0.02128)
Control by region	✓	✓	✓	✓	✓
Pseudo R-squared	0.3855	0.3878	0.4288	0.4953	0.4398
Number of observations	42584	46753	56834	47452	193623

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 16: Results of model number five with the CASEN survey, but controlling with the region level, in the CASEN surveys of 2022, 2017, 2015 and 2013, and also an extension with a pool of the four surveys, where the year of 2013 is the category left behind for the regression. Also the region category also left for interpretation is the Biobío region.

Results when controlling by the Municipality:

	(2022)	(2017)	(2015)	(2013)	(Pool)
	ln_youth	ln_youth	ln_youth	ln_youth	ln_youth
Gender	0.02595*** (0.00744)	-0.00018 (0.00721)	0.00648 (0.00644)	0.00580 (0.00632)	0.00788** (0.00344)
Age	-0.00120 (0.00108)	0.00076 (0.00100)	0.00131 (0.00090)	0.00249*** (0.00091)	0.00092* (0.00048)
Age2	0.00002 (0.00001)	0.00001 (0.00001)	-0.00001 (0.00001)	-0.00001 (0.00001)	-0.00000 (0.00001)
Years of education	-0.01596*** (0.00347)	0.00169 (0.00355)	-0.00374 (0.00310)	-0.00507* (0.00297)	-0.00576*** (0.00163)
Years of education2	0.00341*** (0.00016)	0.00285*** (0.00018)	0.00311*** (0.00016)	0.00334*** (0.00015)	0.00319*** (0.00008)
Number	0.21444*** (0.00252)	0.18983*** (0.00229)	0.17916*** (0.00209)	0.18759*** (0.00195)	0.19041*** (0.00110)
Employed	0.36817*** (0.00865)	0.28647*** (0.00831)	0.23409*** (0.00738)	0.21380*** (0.00725)	0.27284*** (0.00395)
LP	-1.30970*** (0.01856)	-1.12007*** (0.01291)	-1.16105*** (0.01000)	-1.12047*** (0.00879)	-1.15697*** (0.00570)
Rural	-0.04849*** (0.00961)	-0.06523*** (0.00909)	-0.03509*** (0.00793)	-0.05376*** (0.00803)	-0.04985*** (0.00429)
Decrease $\overline{\text{NDVI}}$ of C.F.	0.12066** (0.05571)	-0.33627*** (0.06388)	0.00124 (0.04382)	-0.42795*** (0.07686)	-0.00538 (0.00422)
B_Farmer * De- crease $\overline{\text{NDVI}}$ of C.F.	-0.06010* (0.03280)	0.06827*** (0.02219)	-0.02853 (0.02119)	0.06806*** (0.02307)	0.03002*** (0.01039)
Binary_Farmer	0.00172 (0.02979)	-0.03018* (0.01785)	0.00343 (0.01773)	-0.03215*** (0.01165)	-0.03719*** (0.00808)
Year 2015					0.02565*** (0.00462)
Year 2017					0.07801*** (0.00462)
Year 2022					0.34447*** (0.00534)
Constant	12.30483*** (0.06196)	12.04828*** (0.03183)	12.19142*** (0.02903)	11.99639*** (0.02795)	12.05892*** (0.01545)
Municipality	✓	✓	✓	✓	✓
Pseudo R-squared	0.3978	0.3980	0.4387	0.5035	0.4457
Number of observations	42584	46753	56834	47452	193623

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 17: Results of model number five with the CASEN survey, but controlling with the municipality level, in the CASEN surveys of 2022, 2017, 2015 and 2013, and also an extension with a pool of the four surveys, where the year of 2013 is the category left behind for the regression. Also the region category also left for interpretation is the Biobío region.

7.9 Treatment of missing values

To understand the decision made during the correction process for missing values is necessary to see Figure 17, which has the missing values for the NDVI variable. The red squares represent the missing values for the fourth quarter of 2016, accounting for 1,652. Most of these missing values are located along the coastline or high in the Andean mountains, with a few exceptions in major cities within the central valley and near Concepción city, which describes the Biobío river. To avoid gaps in the data, the decision was made to change the missing values to zeros. The areas with missing values are the major cities of Rancagua, Temuco, Chillán, and Concepción, which generally have near-zero vegetation. Also, the mountainous regions are likely to consist of snow or rocks, while the coastal areas might include seashores due to the grid size. A similar problem is seen across the periods so the same decision was taken.

Moreover, as indicated in Appendix 6.4, NDVI values below zero typically represent water, snow, or clouds. These account for only 11 out of 33,336 observations in the 2016, fourth quarter, approximately 0.033% of the data. This percentage remains consistent across other periods. To facilitate interpretation, the decision was made to adjust the NDVI range from zero to one.

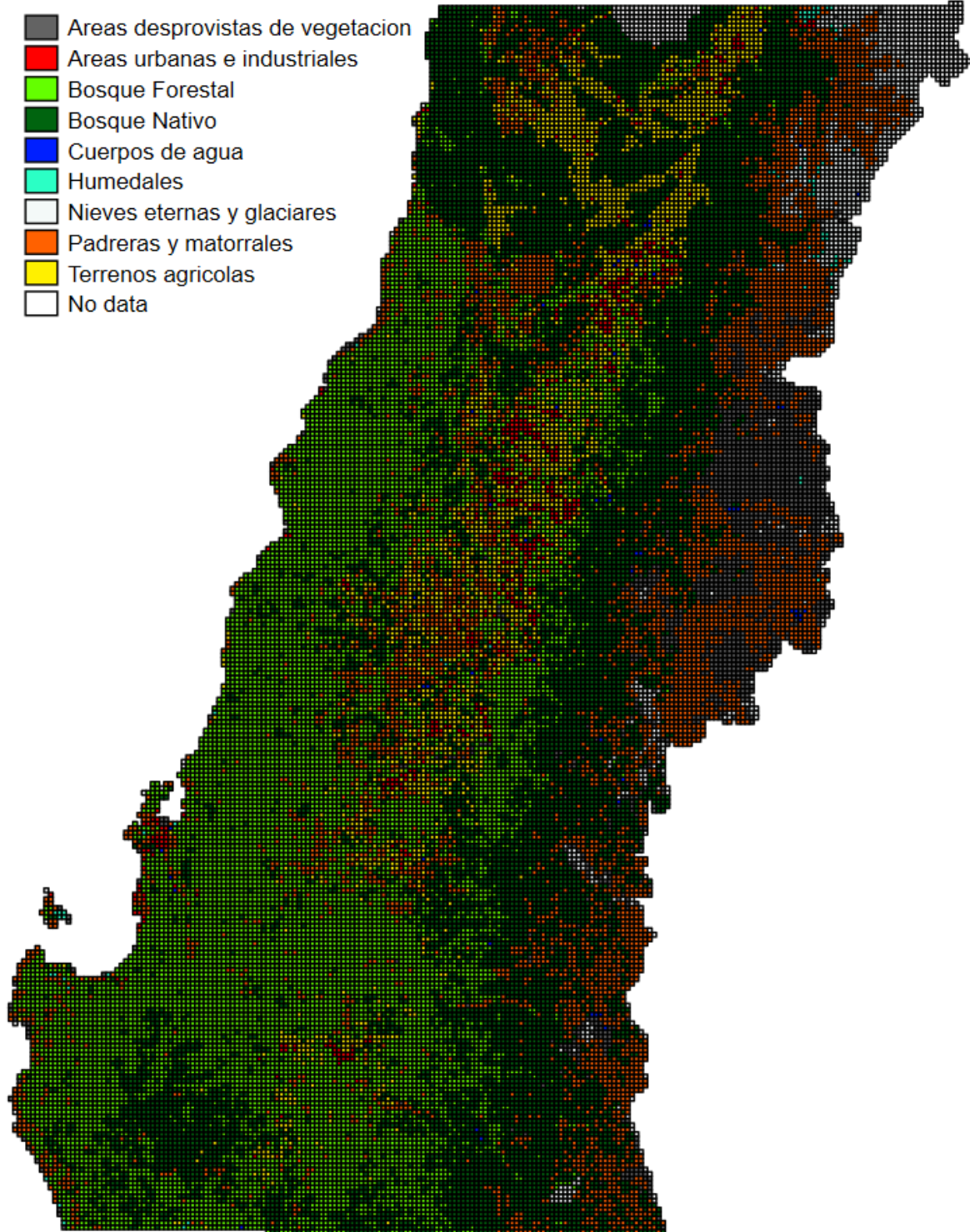


Figure 13: Land use, from the latest cadastre, in the total range of study, without any correction for the missing values.

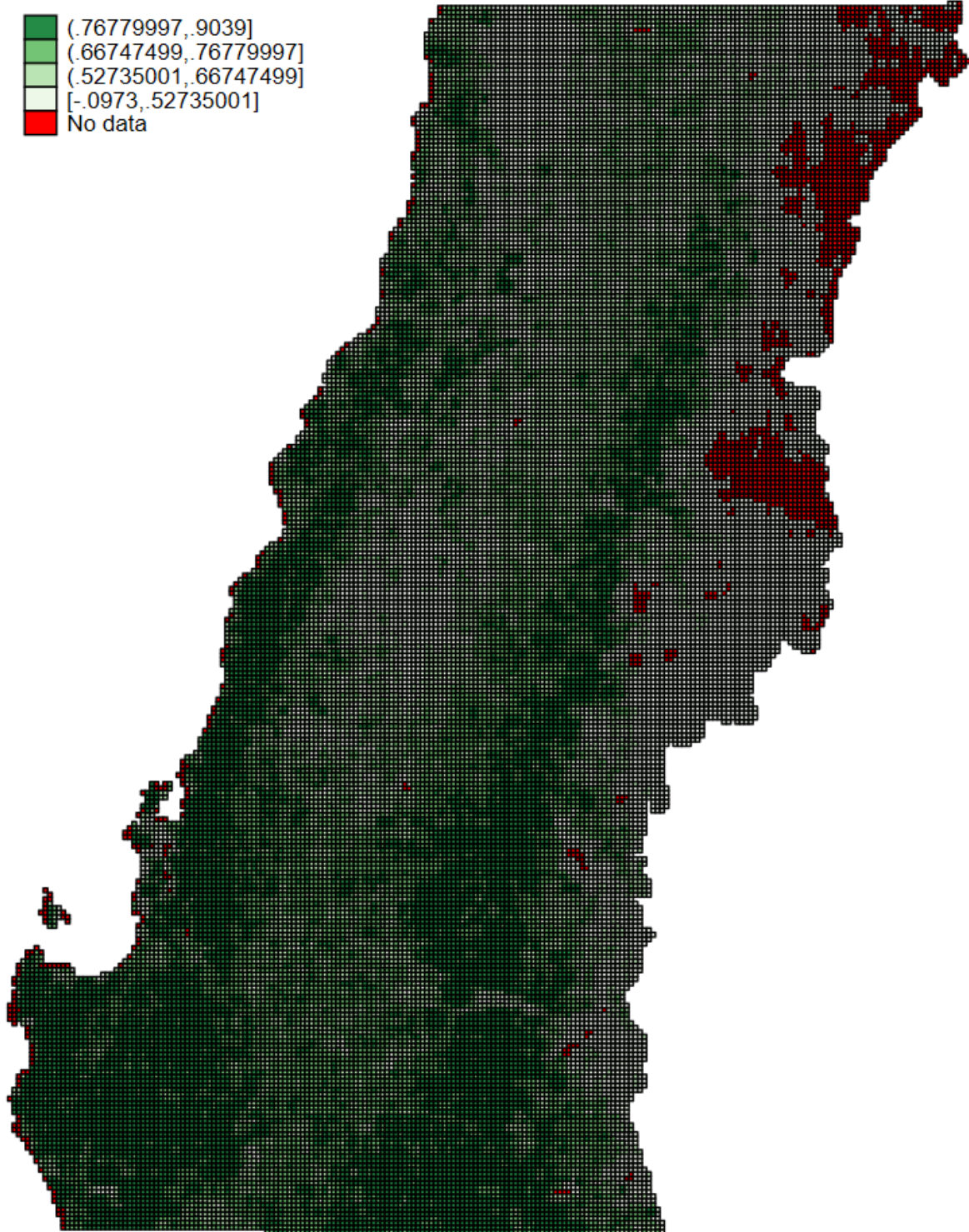


Figure 19: Missing values for the NDVI variable at the year 2016, fourth quarter. In the total range of study.