

Heterogeneous Wetland Benefits for Flood Protection: A Multimodal Causal Framework

By SAYEDMORTEZA MALAEKEH*

We developed a multimodal causal framework by integrating Vision Transformers with the R-Learner framework to more accurately capture heterogeneity in estimating wetland benefits for flood protection. Using detailed, multi-scale, and multi-spatiotemporal-resolution data, we applied this framework to wetlands across Florida between 1996 to 2016. We find that each hectare of wetland loss increases flood damages by \$195.29 on average, with maximum benefits reaching \$512.29 in high-benefit areas. This framework demonstrates significant improvements in capturing heterogeneity compared to linear models or more advanced techniques relying solely on tabular data, increasing the Ranked Average Treatment Effect (RATE) ratio from 0.09 (sd = 0.09) in the tabular-only model to 0.23 (sd = 0.09) in the multimodal model. Key drivers of heterogeneity include land cover, exposed property values, and geography, highlighting the importance of location-specific factors in determining wetland benefits. Our results demonstrate the critical role of wetlands in flood mitigation and provide policymakers with a robust tool for prioritizing conservation efforts.

Keywords: R-Learner, Heterogeneous Treatment Effect, Vision Transformer, Wetland, Non-market valuation, Multimodal Causal Framework

Humanity depends critically on Earth’s ecosystems and the services they provide, such as food, fuel, clean air, and natural hazard protection. Over the past century, these ecosystems have undergone rapid and extensive transformations, fueled by economic growth and the demands of an expanding global population. While these changes have supported human development, they have also imposed significant ecological costs. Quantifying these costs has been challenging since nature’s goods and services often lack prices or markets (Assessment, 2005). Where should ecosystems be converted to other land uses, and where should they be

* This research is a collaborative project with Hannah Druckenmiller and Connor Jerzack, initiated during my visit to Caltech, and we plan to publish it together. The preliminary results presented here were drafted solely at my request to serve as a writing sample for my application process, highlighting my research interests. I conducted the analyses reported here independently, and this draft has not been reviewed by either Hannah or Connor. However, both Hannah and Connor contributed significantly during our weekly discussions and through their intellectual contributions, which is why I use “we” throughout the full text. Hannah provided the processed tabular data, while Connor contributed a set of representations, from which I identified the optimal configurations. Both have supervised this work with exceptional patience and guidance, for which I am sincerely grateful. Portions of the introduction were adapted from the grant proposal for this project. Table 1 and Figure 2 are also created by Hannah. I kindly request that this document not be circulated or shared.

preserved to maximize societal benefits? Addressing this question requires accurate geographic estimates of the benefits provided by nature. In this study, we develop a scalable, quantitative, and rigorous method for modeling nature’s benefits as a function of local human and physical features. Our approach provides insights into key questions of geographic science: How do ecosystem services vary across space, particularly along the rural-urban gradient or across diverse landscape structures? What local factors drive differences in the benefits ecosystems provide? And how do natural versus human land uses deliver the greatest societal benefits?

We focus on the concept of ecosystem services, which captures the goods and services that nature provides to society. This framework allows policymakers to evaluate the benefits of nature alongside the economic costs of conservation. Despite the growing policy relevance of ecosystem services, real-world applications remain scarce due to challenges in generating accurate, scalable estimates (Everard, 2014; Mendelsohn and Olmstead, 2009). Traditional “benefits transfer” methods, which extrapolate ecosystem service estimates from one context to another, often fail to capture local heterogeneity in natural, social, and built systems (Everard, 2021).

Our work is most closely related to the literature that empirically quantifies damages resulting from environmental changes, such as studies quantifying the impact of temperature on economic productivity (Burke et al., 2015), pollution on health (Schlenker and Roberts, 2009), and climate on crop yields (Deryugina et al., 2019). Flooding is among the most frequent and costly natural disasters in the United States and worldwide (Wing et al., 2020). Wetlands serve as a nature-based solution to mitigate flood risk by acting as natural sponges that absorb and slow floodwaters. Previous studies have established a link between wetland loss and increased property damages from flooding (Costanza et al., 2021; Sun and Carson, 2020; Taylor and Druckenmiller, 2022; Aronoff and Rafey, 2023). These studies highlight that the flood protection services provided by wetlands are spatially heterogeneous; however, the methods used to model this heterogeneity are often simplistic. For instance, Taylor and Druckenmiller (2022) estimate heterogeneous effects by interacting wetland loss at specific locations with a limited set of local factors, such as population density, one at a time. In reality, the effectiveness of wetland flood protection services depends on various characteristics, including wetland type (e.g., estuarine emergent, freshwater forested), natural features (e.g., vegetation, soil, depth), spatial positioning relative to water networks (e.g., coastal, adjacent to rivers, isolated), proximity to human systems (e.g., nearby properties), and the interactions among these factors.

To address these limitations in capturing heterogeneity when estimating wetland benefits, we leverage recent advances in causal deep learning to develop a novel multimodal causal framework for estimating heterogeneous ecosystem services. Specifically, we integrate Vision Transformers with the *R*-Learner framework, a flexible approach in causal inference that separates the modeling of treat-

ment effects from outcome prediction to estimate heterogeneous treatment effects. This enables us to determine how the benefits of wetlands vary as a function of detailed local features, effectively incorporating both structured and unstructured data for more nuanced and accurate analysis.

Our analysis uses detailed, multi-scale, and multi-spatiotemporal-resolution data to estimate the location-specific flood mitigation benefits of wetlands across Florida. We employ flood claims data spanning decades, along with a rich set of covariates including elevation, land cover, proximity to shorelines, and sociodemographic information. These data allow us to estimate the conditional average treatment effect (CATE) of wetland development on flood damages, offering the first scalable framework for estimating ecosystem services at this level of granularity.

This study makes several key contributions by advancing the understanding of wetland benefits and providing actionable insights for environmental decision-making. First, we conduct a detailed analysis of location-specific wetland benefits, estimating the causal effects of wetland development on flood damages at a highly granular level to uncover localized variations in their protective value. To achieve this, we integrate Vision Transformer architecture with the *R*-Learner framework, a novel approach that combines advanced computer vision techniques with causal inference to estimate heterogeneous treatment effects. This integration allows us to capture spatial heterogeneity in wetland benefits and demonstrates the importance of multimodal data in identifying key drivers such as exposed property value, elevation, land cover, and proximity to shorelines.

Our findings provide spatially explicit estimates that can guide policymakers in prioritizing conservation efforts, such as targeting wetlands near high-risk flood zones or urban areas with significant flood mitigation needs. Additionally, the scalability of our framework makes it applicable beyond Florida to other regions with rich multimodal datasets and adaptable to broader ecosystem valuation challenges, such as assessing the role of forests in agricultural productivity or urban greening projects in regulating extreme heat.

We find that each hectare of wetland loss increases flood damages by \$195.29 on average, with maximum benefits reaching \$512.29 in areas with high exposure to flood risks. Our analysis highlights the limitations of linear models in capturing the complex heterogeneity of wetland benefits. Even advanced techniques like the *R*-Learner framework, when applied without multimodal data, fail to fully capture this richness. By integrating multimodal data into the framework, we significantly enhance the ability to model heterogeneity, increasing the Ranked Average Treatment Effect (RATE) ratio from 0.09 (SD = 0.02) in the tabular-only model to 0.23 (SD = 0.04) in the multimodal framework. This improvement not only provides more accurate estimates but also equips policymakers with a powerful tool for informed decision-making. By quantifying heterogeneity more effectively, our approach underscores the critical role of wetlands in mitigating climate change impacts and supports sustainable management practices.

The paper is structured as follows: Section I outlines our data, Section II introduces the causal identification strategy, validation procedures, and our multimodal causal framework. Section III presents the results, Section IV concludes the paper with a short discussion and policy implications.

I. Data

We measure property damages from flooding using flood insurance claims from the National Flood Insurance Program (NFIP). The NFIP is the dominant insurer for flooding in the United States, with over 1.8 million policies and \$466 million in coverage in Florida. We use NFIP claims because they provide a monetary measure of property damages from flooding that can be directly applied in cost-benefit analyses. Furthermore, the richness of the data—spanning temporal, spatial, and granular dimensions, along with its consistency—makes our analysis empirically tractable. NFIP participation exceeds 50% among homes located in floodplains, driven by the program’s subsidies and the requirement for coverage on federally-backed mortgages. However, a limitation of using these data is that they do not account for damages outside the NFIP, potentially underestimating the flood protection services provided by wetlands.

The conversion of wetlands to developed areas is measured in log hectares and referred to as “wetland development” hereafter. We quantify wetland development using data from the Coastal Change Analysis Program (C-CAP) land cover product. We used the difference in wetland area between 1996 and 2016 as our treatment variable. Additionally, we computed five-year averages of flood damages because flood insurance loss payments in the NFIP dataset are highly variable across individual years due to the infrequent nature of flood events. Figure 1 illustrates the differences in wetland development and flood claims (both in logarithmic scale) from 1996 to 2016 in Panels (a) and (b). Panel (c) shows a scatter plot illustrating the positive relationship between wetland development and flood claims differences. However, this relationship is influenced by factors such as wetter climates and rapid population growth, which often lead to more wetlands, urban expansion, and higher flood claims, complicating the causal interpretation.

For control variables, we collected extensive multimodal data for each site, encompassing aerial imagery, natural system features (e.g., land cover, flood zones, elevation, surface water networks, and soil classifications), human system attributes (e.g., property locations, values, and flood mitigation measures), as well as weather and climate metrics (e.g., precipitation patterns and hurricane exposure).

Table 1 summarizes the data variables used in the analysis, highlighting their sources, spatial resolutions, and temporal resolutions. The outcome variable, flood damages, is observed daily at the census block level, while wetland development—the primary treatment variable—is available every five years at a 30-meter resolution. Data on extreme weather events, like precipitation and hurricane exposure, are recorded daily. Human systems data, such as property locations,

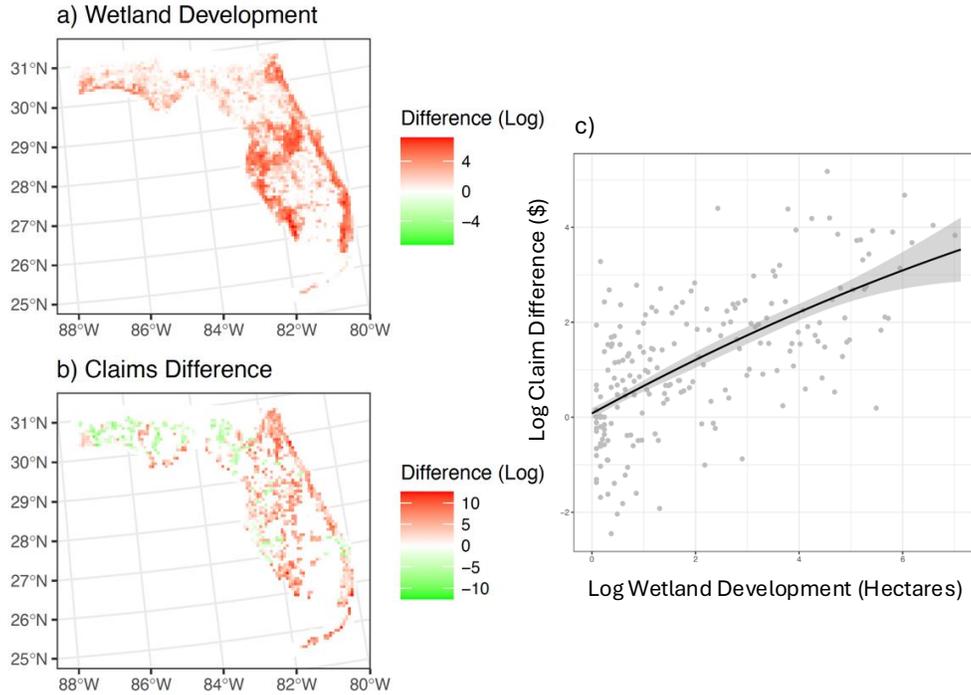


Figure 1. : Increased wetland development corresponds to higher flood claim values. *Note: a) the spatial distribution of wetland development from 1996 to 2016, b) the spatial distribution of flood claim differences from the 5-year window centered around 1996 and 2016, and c) the association between wetland development and flood claim changes. Wetland development represents the conversion or loss of wetland areas using data from the the Coastal Change Analysis Program (C-CAP). Flood claim differences are derived from the National Flood Insurance Program (NFIP) data. Both variables are aggregated to the grid cells across Florida*

values, and flood mitigation measures, are updated annually. Physical attributes, including the surface water network and soil characteristics, are typically static snapshots. Spatial resolutions vary from precise point data to broader administrative or hydrological regions, allowing flexibility in model specification and the incorporation of interdependencies across hydrologically connected locations. This detailed data framework supports robust modeling of ecosystem and human system interactions.

Figure 2 demonstrates the multimodality of data sources used in our analysis, using two wetland sites in Florida as examples. Site 1, shown at the top, represents a coastal residential area, while Site 2, at the bottom, depicts an inland agricultural region. Each column showcases a different data layer, including aerial imagery, wetland polygons and types, property locations and values, FEMA flood zones, land cover classifications, and elevation measurements. The integration of

Table 1—: Variables and data formats.

	Data source	Variables	Spatial resolution	Temporal resolution
	(1)	(2)	(3)	(4)
<u>Outcome</u>				
Flood damages	National Flood Insurance Program (NFIP)	Flood insurance claims paid	Census block	Daily
<u>Treatment</u>				
Wetland development	USGS Coastal Change Analysis Program (CCAP)	Land converted from wetland to developed area	30 meter	5-year
<u>Covariates</u>				
Residential properties	CoreLogic	Property coordinates, assessed values, and characteristics (e.g. stories)	Points	Annual
FEMA flood zones	National Flood Hazard Layer (NFHL)	12 different risk classes	Polygons	Static
Sociodemographics	American Community Survey	Population density, median income, race/ethnicity, etc.	Census tract	Annual
Adoption of flood mitigation measures	National Flood Insurance Program (NFIP)	Community Rating System (CRS) score	Census block	Annual
Land cover	USGS Coastal Change Analysis Program (CCAP)	24 land cover classes, including 6 different wetland types	30 meter	5-year
Surface water network	National Hydrography Dataset	Flowlines, water resource type, flow rates, etc.	Vector	Static
Soil characteristics	Gridded National Soil Survey Geographic Database (gNATSGO)	Depth to water table, soil taxonomy, hydric rating, flooding frequency, ponding frequency	30 meter	Static
Ecoregion	Environmental Protection Agency	Level IV classification	Polygons	Static
Climate	PRISM Climate Group	Precipitation, temperature, vapor pressure deficit, solar radiation, cloud transmittence	800 meter	Static
Precipitation	PRISM Climate Group	11 precipitation indicators (e.g. Rx1day, Rx5day, SDII, R10mm, R20mm)	800 meter	Daily
Hurricane exposure	NOAA HURDAT + wind field model	Maximum wind speed, Power Dissipation Index (PDI)	800m	Daily
Aerial imagery	National Aerial Imagery Program (NAIP)	RGB + Infrared	Submeter	Static

Note: We collect extensive high-resolution data on natural and human systems at each location to inform our estimation of location-specific ecosystem values. Data come from a variety of sources (column 2) and have different spatial and temporal resolutions (columns 4 and 5). Our study period spans the years 1996 to 2026

these diverse data modalities is a key contribution of this study, allowing for the precise estimation of the heterogeneous causal effects of wetland development on flood damages.

All data are aggregated at the scene level, with each scene covering approximately 58.98 km^2 . This resolution results in a total of 3,566 scenes, ensuring complete coverage of the state of Florida (Figure 3). Aggregating data at this scale preserves spatial patterns and variability while maintaining a manageable granularity for tabular analysis. This represents the finest resolution possible without significant loss of information from the spatial resolutions of the variables. The integration of images with differing spatial resolutions will be discussed further in the Empirical framework section.

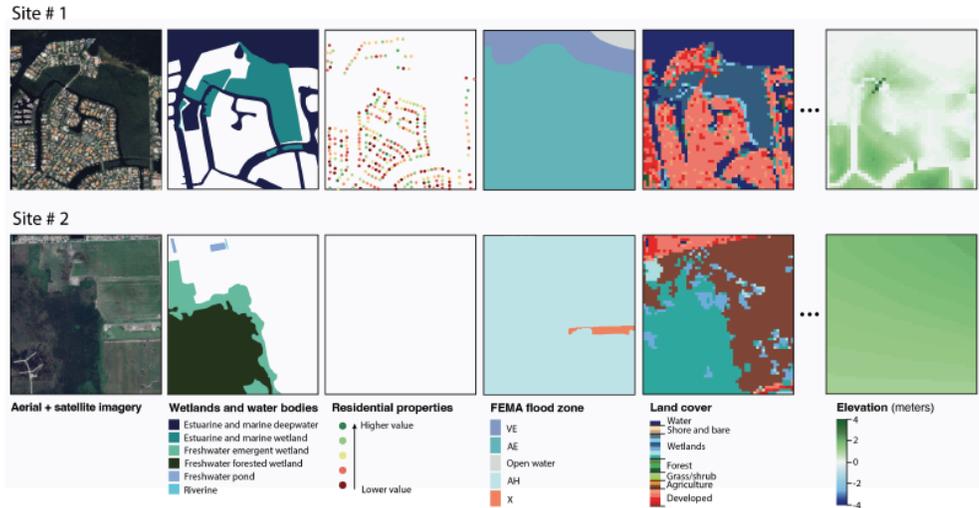


Figure 2. : Example of two wetland sites in Maimi Dade, FL. Note: Site 1 (TOP) is a residential area along the coast. Site 2 (BOTTOM) is an inland agricultural area. Each column shows a different data layer (LEFT to RIGHT): aerial imagery, wetland polygons and types, the location and value of properties, flood zones, land cover, and elevation.

II. Empirical Framework

Our empirical approach has two primary objectives: (i) to estimate the *causal effect* of ecosystem changes on indicators of human well-being and (ii) to estimate *heterogeneous effects* that allow for rich variation in ecosystem services across space. A key challenge is that wetland extent is correlated with other factors influencing flood damages. For instance, regions with higher economic development often have better flood protection infrastructure but may also experience greater economic losses during extreme events due to higher property values. As a result, we observe a positive correlation between flood protection infrastructure

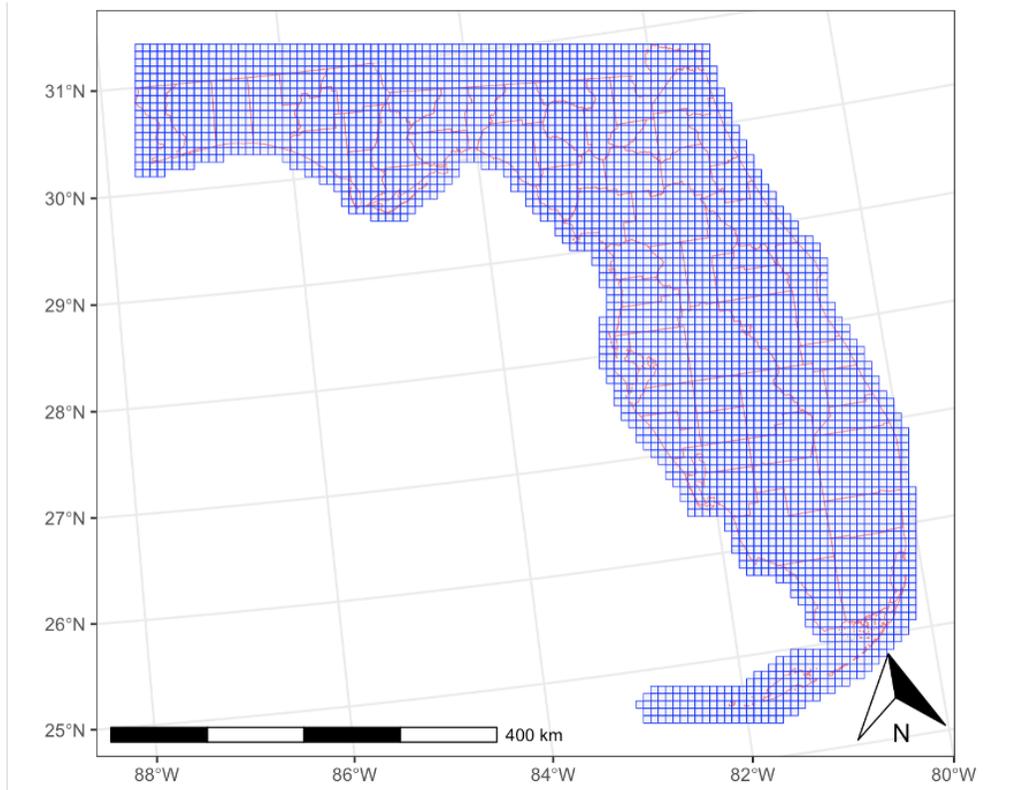


Figure 3. : Overlay of scenes and county boundaries in Florida. *Note: The grid overlay represents the spatial unit of analysis, where each scene corresponds to a cell with an approximate area of 58 km², and the red lines depict county boundaries. These spatial units form the foundation for integrating wetland development, flood damage data, and covariates, enabling a consistent analysis of heterogeneity across geographic scales in tabular format. The uniform grid design supports the aggregation of multi-resolution data, ensuring comparability while maintaining spatial granularity.*

and economic losses. However, this relationship should not be interpreted as flood protection causing greater losses, as confounding factors, such as regional wealth, likely drive the correlation. Similarly, the positive correlation between economic development and flood damage raises concerns about omitted variable bias.

Our primary identification strategy leverages the *R-Learner* framework (Nie and Wager, 2021) with long-differenced treatment and outcome variables to estimate the long-term effects of wetland development on flood damage, taking into account potential adaptation. This approach addresses time-invariant unobserved confounders but may still be sensitive to time-varying unobservables that affect both the treatment and outcome. Despite relying on relatively strong assumptions of unconfoundedness compared to quasi-experimental methods, this strategy offers greater generalizability. It allows for flexible modeling of heterogeneity using advanced methods and scalability, including the integration of computer vision

techniques to fully utilize multimodal data. To further validate our causal estimates, we employ a natural experimental approach with difference-in-differences, which is discussed in detail in Section III.C.

We want to examine the impact of wetland development ΔT_i at the site i on flood damage ΔY_i , taking into account all the observable characteristics of the site X_i . X_i itself is set of variables $X_i = \{I_i, N_i, H_i, W_i\}$ in which I_i is aerial imagery, N_i is natural systems, H_i is human systems, and W_i is weather and climate data. The general model is

$$(1) \quad \Delta Y_i = g(X_i, \Delta T_i) + f(X_i) + \epsilon_i$$

where $g(X_i, \Delta T_i)$ captures the complex relationship between wetland development and flood damage, including interactions with site characteristics that cannot be identified using conventional causal techniques. Taylor and Druckenmiller (2022) assumed a linear causal relationship between T and Y , which provides valuable insights but may oversimplify the complexity of this setting.

We are particularly interested in the average treatment effect (ATE), defined as the expectation over the joint distribution of X_i and ΔT_i :

$$\text{ATE} = \mathbb{E} \left[\frac{\partial g(X_i, \Delta T_i)}{\partial \Delta T_i} \right].$$

Similarly, the conditional average treatment effect (CATE) can be estimated by conditioning on specific site characteristics:

$$\tau(x) = \text{CATE}(x) = \mathbb{E} \left[\frac{\partial g(X_i, \Delta T_i)}{\partial \Delta T_i} \mid X_i = x \right].$$

Specifically, we propose using the R -Learner framework developed by Nie and Wager (2021) to estimate heterogeneous treatment effects by reformulating the estimation procedure into three prediction tasks built on machine learning and deep learning approaches.

The next sections proceed as follows: In Section II.A, we briefly describe how causal identification and the estimation of heterogeneous treatment effects using the R -Learner framework are reformulated as prediction tasks. Section II.B introduces our multimodal causal framework, highlighting the methodological innovations achieved by combining computer vision with the R -Learner framework. Section II.C sheds light on the heterogeneity analysis, and Section II.D explains the validation of our results and framework using a natural experimental approach with difference-in-differences.

A. R -Learner and Heterogenous Treatment Effect Estimation

We formulate the problem using the potential outcomes framework (Neyman, 1923; Rubin, 1974), where for each site i , we observe features X_i , a long-differenced

outcome ΔY_i , and a long-differenced treatment ΔT_i . We assume the existence of potential outcomes $\Delta Y_i(\Delta T)$ for each treatment level ΔT . The goal is to estimate the conditional average treatment effect (CATE):

$$(2) \quad \tau^*(X_i) = \mathbb{E} \left[\frac{\partial \Delta Y_i(\Delta T)}{\partial \Delta T} \mid X_i \right].$$

Under the assumption of unconfoundedness, $\{\Delta Y_i(\Delta T)\} \perp \Delta T_i \mid X_i$, the conditional mean outcome and treatment propensity are defined as:

$$(3) \quad m^*(X_i) = \mathbb{E}[\Delta Y_i \mid X_i], \quad e^*(X_i) = \mathbb{E}[\Delta T_i \mid X_i].$$

The Robinson decomposition (Robinson, 1988) rewrites the estimation problem as:

$$(4) \quad \Delta Y_i - m^*(X_i) = (\Delta T_i - e^*(X_i))\tau^*(X_i) + \epsilon_i.$$

If the conditional average treatment effect (CATE) is constant, i.e., $\tau(X) = \tau$ for all X , the following estimator is semiparametrically efficient for τ under unconfoundedness (Chernozhukov et al., 2018, 2024; Robinson, 1988):

$$(5) \quad \hat{\tau} = \frac{\frac{1}{n} \sum_{i=1}^n (\Delta Y_i - m^*(X_i)) (\Delta T_i - e^*(X_i))}{\frac{1}{n} \sum_{i=1}^n (\Delta T_i - e^*(X_i))^2}.$$

Although this estimator assumes a constant treatment effect, the formulation can be extended to estimate $\tau^*(X_i)$ by minimizing the empirical loss, motivated by the *R-Learner* framework (Nie and Wager, 2021):

$$(6) \quad \hat{\tau} = \arg \min_{\tau} \frac{1}{n} \sum_{i=1}^n [(\Delta Y_i - m^*(X_i)) - (\Delta T_i - e^*(X_i))\tau(X_i)]^2 + \Lambda_n(\tau),$$

where $\Lambda_n(\tau)$ is a regularization term. This framework supports flexible machine learning and deep learning techniques for calculating the mean treatment and outcome propensity scores, as well as for optimizing the minimization objective function to capture the complex relationship between wetland development (ΔT) and flood damages (ΔY). CATE estimates are obtained using 5-fold estimation to ensure robust and unbiased evaluation. In each fold, the model is trained on 80% of the data, while CATEs are predicted on the held-out 20%, preventing data leakage and overfitting. We utilize a variety of models, including *linear* models with regularization, *forest-based* methods, *gradient boosting* models, and *neural networks*. Among these, we find that using random forests for the first two stages and causal forests (Wager and Athey, 2018) to solve the local moment equation provides the most robust results. A detailed comparison of these approaches is included in Appendix A1.

B. Multimodal Causal Framework

Current approaches to modeling heterogeneous effects predominantly rely on algorithms that process tabular data, such as causal forests or gradient boosting methods. However, these methods often fail to capture the rich, complex interactions present in multimodal datasets (Jerzak and Daoud, 2023; Jerzak et al., 2023). Such datasets may include satellite imagery, weather observations, surface water network vector data, residential property values, raster data on elevation and soil characteristics, and administrative data on sociodemographics. Aggregating this information into tabular formats can obscure critical local-level interactions. For instance, the relative location of a wetland to surface water networks and residential properties may significantly influence its flood protection value. A wetland situated between floodwater sources and residential properties is more likely to mitigate flood damages effectively than one located behind the properties.

In this work, we focus on developing a multimodal causal framework to improve causal effect estimation by harnessing the strengths of diverse data sources and representations. By integrating multimodal data, the framework can better capture the interactions between human and natural systems, providing more nuanced and accurate insights into causal relationships.

As argued by Belloni et al. (2014), orthogonalization is crucial in causal inference with high-dimensional data, particularly when certain features strongly predict treatment propensities but are weakly predictive of outcomes. A multimodal framework addresses this limitation by inherently supporting orthogonalization, effectively capturing global spillovers that are especially significant in image-based settings. These spillovers enhance the prediction of both treatment and outcomes given a set of covariates. By integrating rich representations from multimodal data, this framework enables robust causal effect estimation, effectively accounting for confounding factors and complex interactions that traditional tabular-based methods often fail to capture.

The proposed multimodal causal framework integrates representations from imagery and tabular data to estimate causal effects as shown in Figure 4. To handle the high dimensionality of pre-treatment image array M_i , we employ a representation extraction function $\phi: \mathcal{M} \rightarrow \mathbb{R}^d$ that maps images to lower-dimensional feature vectors. We use transformer architectures for image embeddings due to their ability to capture both local and global dependencies, unlike convolutional neural networks (CNN), which rely on locality and translation invariance (Dosovitskiy et al., 2021). Transformers’ attention mechanisms allow reasoning across spatially and temporally dispersed observations, making them ideal for our project. For example, they can model how rainfall in one location impacts flood damages in hydrologically connected distant areas, enabling more accurate causal effect estimation. First, images are processed using a pre-trained Vision Transformer (ViT) architecture (Dosovitskiy et al. 2021), specifically the ViT-B/16 model. The input covariates as images are divided into fixed-size patches (e.g., 16×16 pixels),

and each patch is flattened into a lower-dimensional vector through a linear transformation. Positional embeddings are added to retain spatial context, and the sequence of patches is fed into a transformer encoder consisting of 12 transformer blocks with multi-head self-attention layers and feed-forward networks. This results in high-dimensional embeddings ($768-D$) that effectively capture spatial and structural features from the images. To reduce dimensionality while retaining key information, Principal Component Analysis (PCA) is applied to the embeddings, selecting components that explain a specified variance (e.g., 80%). The resulting processed representations serve as compact, variance-preserving summaries of the visual data. The processed embeddings (V_i) from the ViT are concatenated with

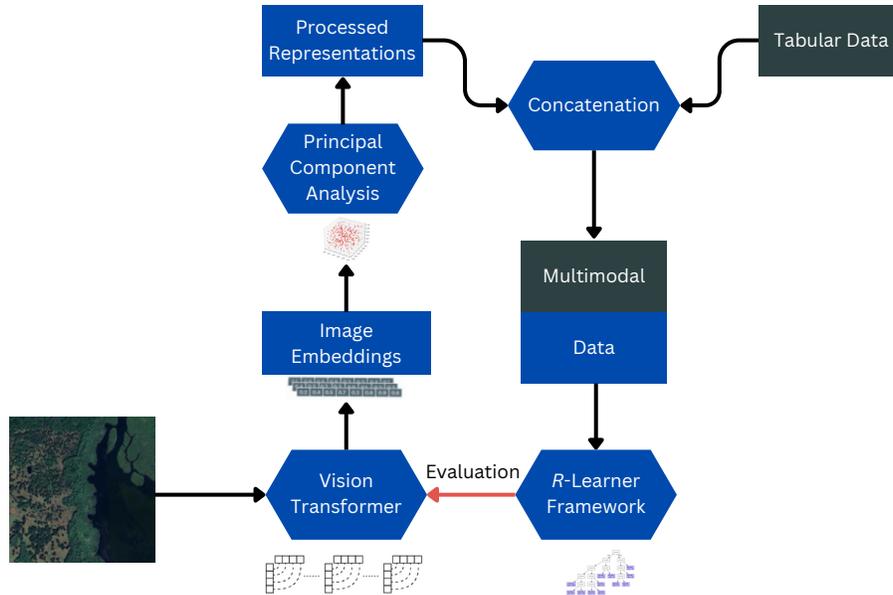


Figure 4. : The schematic of multimodal causal framework developed in this study. *Note: It integrates Vision Transformer (ViT) embeddings and tabular data for causal estimation. Image embeddings are extracted using ViT-B/16, reduced via PCA, and combined with tabular covariates to create a multimodal dataset. The model passes this dataset to the R-Learner framework, which provides evaluation metrics. An iterative process explores all combinations of representations and PCA values to identify the optimal configuration, ensuring effective capture of treatment heterogeneity.*

tabular covariates (\mathbf{X}_i) to form a multimodal dataset, which is then input into the R -Learner framework (Nie and Wager, 2021). The R -Learner reformulates the causal estimation problem into a loss minimization task. This process iteratively

searches over combinations of PCA-reduced representations and evaluates their ability to explain treatment selection and outcomes and heterogeneity measures from CATEs, ensuring optimal representations are used for causal estimation.

To fix ideas, the optimization proceeds by comparing the heterogeneity signal of the representations extracted from $\phi(M_i, k)$ against the baseline heterogeneity signal from the tabular model D_i . The goal is to identify the combination of representation $\phi \in \Phi$ and PCA level $k \in K$ that maximizes the improvement in heterogeneity μ over the baseline.

$$(7) \quad \max_{\phi \in \Phi, k \in K} \{ \mathbb{E}[\mu(h_\theta(\phi(M_i, k)))] - \mathbb{E}[\mu(h_\theta(D_i))] \},$$

where Φ is the set of candidate representation extraction functions. M_i represents the spatial imagery for unit i . $k \in K$ specifies the level of PCA retained (e.g., the number of components). h_θ maps the extracted representations to treatment and outcome spaces. μ measures the heterogeneity of causal effects given the representations. D_i represents the baseline tabular model for comparison.

Using the multimodal causal framework, we aim to understand and compare the dynamics of the following quantities, both separately and jointly:

Tabular CATE:

$$(8) \quad \tau(x) = \mathbb{E} \left[\frac{\partial \Delta Y_i(\Delta T)}{\partial \Delta T} \mid \mathbf{X}_i = x \right],$$

where \mathbf{X}_i represents tabular covariates.

Image CATE:

$$(9) \quad \tau(m) = \mathbb{E} \left[\frac{\partial \Delta Y_i(\Delta T)}{\partial \Delta T} \mid V_i = v \right],$$

where V_i is the image representation extracted from the ViT model.

Multimodal CATE (Image + Tabular):

$$(10) \quad \tau(v, x) = \mathbb{E} \left[\frac{\partial \Delta Y_i(\Delta T)}{\partial \Delta T} \mid V_i = v, \mathbf{X}_i = x \right],$$

where V_i represents the image embeddings concatenated with tabular features \mathbf{X}_i .

Additionally, we seek to quantify how much the optimal image representations enhance our ability to capture the presence of heterogeneity which is discussed in Section II.C.

C. Assessing Heterogeneity

At first, we evaluate how each of our models performs in practice, focusing on their ability to capture effect heterogeneity across different settings (tabular, image, and multimodal). Since true CATE estimates are unobserved, directly

assessing model performance is challenging. However, we can address this issue by leveraging the Rank-Weighted Average Treatment Effect (RATE) (Yadlowsky et al., 2024). This measure provides a way to quantify and compare the performance of different CATE estimation methods in the absence of ground truth individual treatment effects. Furthermore, we generalize the RATE metric to accommodate continuous treatment settings.

The RATE ratio quantifies how effectively each model prioritizes units with high treatment effect heterogeneity. By integrating over the TOC curve, we measure the gain in treatment effect lift under different prioritization rules, allowing us to determine which representation type performs best in capturing heterogeneity. The RATE ratio is:

$$(11) \quad \text{RATE} = \int_0^1 a(q) \left[\mathbb{E} \left(\frac{\partial \Delta Y_i}{\partial \Delta T_i} \mid S(Z_i) \geq 1 - q \right) - \mathbb{E} \left(\frac{\partial \Delta Y_i}{\partial \Delta T} \right) \right] dq,$$

where $S(Z_i)$ is the scoring function applied to the representation Z_i (tabular, image-only, or multimodal) to prioritize units. The scoring function $S(\cdot)$ is derived from a data sample independent of the one used for calculating the treatment effect estimates. $a(q) = 1$ is the weighting function using the Area Under the Treatment Effect Curve (AUROC).

Secondly, we estimate the best linear fit using forest predictions (on held-out data) as well as the mean forest prediction as regressors using the following equation:

$$(12) \quad \Delta Y_i - m^*(X_i) = \alpha(\Delta T_i - e^*(X_i)) + \beta(\hat{\tau}(X_i) - \bar{\tau})(\Delta T_i - e^*(X_i)) + \epsilon_i,$$

in which the coefficients α and β evaluate CATE prediction performance. $\alpha = 1$ indicates correct average predictions, while $\beta = 1$ reflects accurate heterogeneity capture. The slope β measures the covariance between predicted and true CATE. A significant $\beta > 0$ confirms heterogeneity, while values below 0 are not meaningful.

D. Quasi-experimental approach for validation

We will validate the R -Learner causal estimates by comparing the parameter estimates from R -Learner with those from a quasi-experiment that uses downstream wetlands as a natural counterfactual for upstream wetlands. The primary challenge for inference in our setting is that wetland extent is correlated with other factors that drive flood damages. For example, communities with wetter climates and more frequent flooding tend to have more wetlands. Indeed, there is a positive correlation between wetland extent and flood damage, but this relationship should not be interpreted as wetlands causing flooding since confounding factors (including precipitation) could be driving the correlation. Similarly, flood damages and real estate development are positively correlated, further raising concerns about omitted variables.

The intuition behind the quasi-experimental setup is that, for non-coastal areas, flood risk should be affected by changes in upstream wetland area but not changes in downstream wetland areas. We can therefore use changes in downstream wetland areas as a natural counterfactual for changes in upstream wetland areas. Following Taylor and Druckenmiller (2022), we will estimate the impact of wetland loss on flood damages using an upstream-downstream difference-in-differences model:

$$(13) \quad \Delta Y_i = \alpha + \theta \Delta T_i^{\text{Up}} + \lambda \Delta T_i^{\text{All}} + \Delta X_i \Omega + \zeta_i$$

where ΔT_i^{Up} denotes changes in wetland area upstream of location i and ΔT_i^{All} denotes changes in wetland area both upstream and downstream of location i . All other variables are defined as in Equation 1. The spatial extent of upstream versus downstream wetland area will be computed using the National Hydrography Dataset flow direction over the geographic extent of the watershed. The coefficient of interest, θ , represents the differential effect of upstream wetlands on flood damages. This upstream-downstream framework effectively uses downstream changes in wetland area to control for time-varying factors unrelated to flooding that drive both changes in overall wetland extent and changes in flood damages. Consider real estate development as an example. Wetlands are lost to urban expansion in places with high population growth, which in turn experience greater flood damages due to the larger housing stock. As a result, wetland loss could be misidentified as *causing* flood damages, when in reality the growth in housing stock is the driving cause of the increased flood damages. The upstream-downstream natural experiment addresses this concern under the assumption that real estate development is not systematically biased toward either upstream or downstream areas relative to a given location.¹ We will compare our estimates of ATE from the *R-Learner* approach and θ from the upstream-downstream approach. If the two values are similar, this will increase our confidence in the causal interpretation of the *R-Learner* estimates since the upstream-downstream approach does not rely on the same conditional independence assumption.

III. Results

A. Main Results

We evaluate a range of models to determine the best approach for estimating treatment effects, including *linear models* with regularization, *forest-based* methods, *gradient boosting* models, and *neural networks*, all optimized with parameter tuning. From this point onward, all results are derived using random forests in the first two stages and causal forests (Wager and Athey, 2018) to address the

¹This assumption is essential for valid inference in the upstream-downstream framework. Taylor and Druckenmiller (2022) tested this and other endogeneity concerns and find no evidence of systematic differences in upstream versus downstream areas.

local moment equation as it performed better than other combination of machine learning prediction algorithms. A comprehensive comparison of these methods and their performances are presented in Appendix A1.

Table 2 presents the Average Treatment Effect (ATE) estimates from three models: the tabular model, the image-only model, and the multimodal model. These models estimate the causal impact of wetland development on flood damages from 1996 to 2016.

As shown in the first row, wetland development significantly increases flood damages across all three models. A 1% increase in wetland development is associated with a 0.2317% to 0.4042% increase in flood damages. While the ATE values for the tabular and multimodal models are similar (0.2317 and 0.2438, respectively), the image-only model (0.4042) shows a substantially higher ATE. This discrepancy might arise because image embeddings alone may not capture the rich, contextual information (such as sociodemographic factors) provided by the tabular data.

In the second row, the dollar impact per hectare of wetland development is reported. Each hectare of wetland development is associated with a change in flood damages of \$185.73 (tabular), \$324.03 (image-only), and \$195.29 (multimodal) at the scene level ($\sim 58 \text{ km}^2$). Again, the multimodal model’s estimates are slightly higher than the tabular model, but the difference is modest. These numbers are comparable to the findings of Taylor and Druckenmiller (2022), where their long-differenced model reported an impact of \$229.2 per hectare at the ZIP-code level.

The results suggest that our *R*-Learner framework coupled with transformer architecture provides estimates similar to traditional linear methods and quasi-experimental approaches in ATE, but we argue that it allows for a more nuanced exploration of heterogeneity which will be discussed in Section III.B. We demonstrate how our multimodal causal framework captures heterogeneity in flood damage impacts that tabular models alone fail to reveal.

B. Heterogeneity

Figure 5 illustrates the Conditional Average Treatment Effect (CATE) estimates obtained from tabular, image-only, and multimodal models. All models exhibit significant heterogeneity in CATE values, reflecting the ability of the *R*-Learner framework to capture complex treatment effect variation. The multimodal and tabular models demonstrate relatively high and similar correlations, suggesting comparable patterns in the spatial distribution of treatment effects, despite differences in the absolute CATE estimates between the two models. In contrast, the image-only model displays substantially different CATE values and lower correlation with the other two models, aligning with the differences observed in the estimated Average Treatment Effects (ATE).

While the observed correlations provide a straightforward comparison of the CATE estimates, it is important to note that correlation is a linear measure and

Table 2—: The Effect of Wetland Development on Flood Damages

<i>Dependent Variable: Changes in Flood Claims (%)</i>			
	Tabular	Image	Multimodal
Wetland Development (%)	0.2317*** (0.0445)	0.4042*** (0.0380)	0.2438*** (0.0430)
<i>Dependent Variable: Changes in Flood Claims (US\$)</i>			
Wetland Development (hectare)	185.73*** (35.67)	324.03*** (30.45)	195.29*** (34.50)
Tabular Controls	✓		✓
Image Embeddings		✓	✓
N	3566	3566	3566

Note: In all three models, R-Learner framework with random forest in the first two stages and causal forest in the last stage is used to estimate average treatment effects (ATE). Covariates include property data (CoreLogic), FEMA flood zones, sociodemographics (ACS), flood mitigation measures (NFIP), land cover (C-CAP), surface water and soil characteristics, climate and precipitation metrics (PRISM), hurricane exposure (NOAA), and high-resolution aerial imagery (NAIP). Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors in parentheses.

may not fully capture non-linear relationships between the CATEs produced by different models. Nonetheless, it serves as a useful starting point for assessing the relative agreement between models in identifying treatment effect heterogeneity.

The analysis of the TOC curve (Figure 6) focuses on the estimated ATE for the top q -percentile of units compared to the overall ATE of the sample. The results demonstrate that the multimodal model consistently identifies higher ATE values in the top percentiles than the image-only and tabular models. This indicates that the multimodal model excels at capturing heterogeneity in treatment effects across the sample, effectively leveraging the combined strengths of tabular and image data. The image-only model, while performing better than the tabular model, still falls short of the multimodal model. The tabular model, in contrast, struggles to capture the complexity of treatment heterogeneity inherent in the data.

Figure 7, which presents the RATE ratio estimates for the three models, reinforces these findings. The RATE estimate for the multimodal model (0.23) is significantly different from zero, validating its ability to detect meaningful heterogeneity in treatment effects. The image-only model also shows some capacity for capturing heterogeneity but is less effective than the multimodal approach. In contrast, the tabular model’s RATE estimate is not significantly different from zero (0.09), underscoring its limitations in accounting for nuanced relationships, particularly those requiring spatial or visual contextualization.

To further assess the heterogeneity across the models in Table 3, we estimate the best linear fit using forest predictions (on held-out data) as the mean forest prediction as regressors, as described in Equation 12, to evaluate the heterogeneity captured by the models. The coefficient α is significantly different from zero

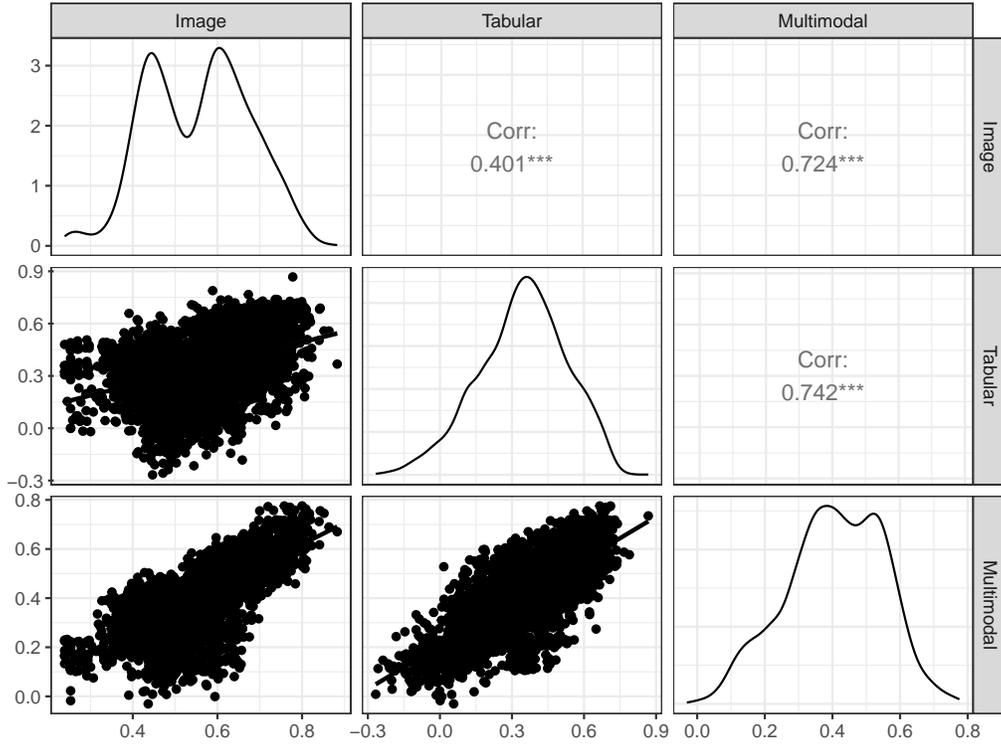


Figure 5. : Pairwise correlation of Conditional Average Treatment Effect (CATE) estimates from the tabular, image only, and multimodal models. *Note: The image-only model uses embeddings extracted exclusively from Vision Transformer (ViT), capturing spatial features from images. The multimodal model combines image embeddings with tabular data through an optimized concatenation process, leveraging the strengths of both data modalities.*

across all three models, indicating that each model accurately captures the average treatment effect. However, only the multimodal model demonstrates the ability to capture heterogeneity, as evidenced by β being greater than zero, while β values in other models are not significantly meaningful.

These results are critical for accurately estimating the benefits of wetlands for flood protection, as they highlight the necessity of multimodal causal framework for capturing the spatial and economic heterogeneity of wetland benefits. This capability is essential for guiding resource allocation, planning interventions, and shaping policies to enhance resilience against flood damages in vulnerable regions.

Figure 8 illustrates the spatial heterogeneity in CATE estimates using the multimodal causal framework, measuring the elasticity between wetland development and per-hectare flood claims across Florida. The CATE values range from 0 to 0.64, translating to wetland benefits of up to \$512.65 in flood claim reductions per hectare. Higher CATE values are predominantly observed near coastal regions, particularly along the Gulf Coast and the southeastern urbanized areas, as well

Table 3—: Using best linear fit using forest predictions (on held-out data) as well as the mean forest prediction as regressors to assess heterogeneity

Model	α (Mean Forest Prediction)	β (Differential Forest Prediction)
Tabular	1.14516*** (0.36228)	-0.22823 (0.53929)
Image	0.98250*** (0.12960)	0.57734 (0.71380)
Multimodal	0.98363*** (0.29947)	1.33410* (0.64371)

Note: A significant α indicates that the the average prediction produced by the forest is correct, while β assesses the model's capacity to adequately capture underlying treatment effect heterogeneity. The multimodal model shows strong performance and its superior ability to integrate information from tabular data and image embeddings to capture both average effects and heterogeneity more effectively than the tabular-only or image-only models. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. One-sided heteroskedasticity-robust (HC3) Standard errors in parentheses.

Table 4—: Comparison of Variables Between Lowest and Highest CATE Groups

	Lowest CATE	Highest CATE	SMD
Exposed Capital			
Property Value (Million USD)	22.96 (7.03)	35.28 (15.08)	0.96
Geography			
Elevation (m)	31.92 (14.46)	10.36 (10.94)	1.68
Depth to water table (m)	107.27 (4.71)	65.56 (4.13)	8.65
Distance to shore (km)	43.35 (2.07)	8.52 (1.40)	13.75
Land Cover (share of land area)			
Developed	0.02 (0.05)	0.03 (0.06)	0.11
Agriculture	0.07 (0.17)	0.20 (0.20)	0.68
Vegetation	0.11 (0.09)	0.06 (0.09)	0.69
Forest	0.25 (0.21)	0.13 (0.16)	0.66
Wetland	0.36 (0.30)	0.33 (0.19)	0.12
High-risk Flood Zones (share of land area)			
A zones	0.12 (0.14)	0.22 (0.33)	0.41
V zones	0.00 (0.04)	0.08 (0.15)	0.62
Number of observations	421	421	

Note: Columns (1) and (2) show the mean (SD) characteristics of scenes with low and high CATEs, respectively. We define low CATE scenes as those with CATEs in the first quartile and high CATE areas as those with CATEs in the fourth quartile. Column (3) shows the standardized mean difference (SMD). Standard deviations in parentheses.

as riverine systems such as the St. Johns River. Conversely, lower CATE values are found in the interior regions of Florida, particularly around the Lake Okechobee watershed. These areas are characterized by fewer urban developments, resulting in a weaker relationship between wetland development and flood claims. This pattern suggests that wetland benefits are highly context-dependent, with

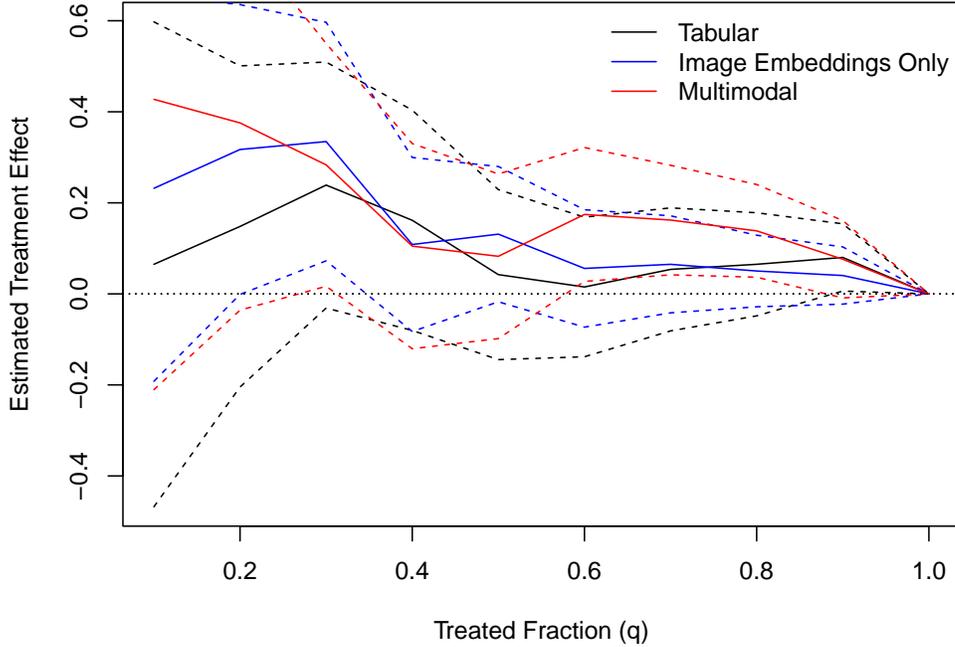


Figure 6. : Targeting operator characteristic curves for the tabular, image only, and multimodal models. *Note: The multimodal model consistently demonstrates higher estimated treatment effects compared to the tabular and image-only models across all treated fractions, highlighting its superior ability to integrate information from both data modalities. 95% confidence interval in dash lines.*

their flood mitigation value varying based on geography, urbanization, and flood exposure.

While the map provides valuable insights into spatial variability, it is challenging to fully understand the drivers of heterogeneity through visual inspection alone. To gain deeper insights, we compare the characteristics of regions in the top quartile (25th percentile) of CATE values with those in the lowest quartile across multiple dimensions, such as property value, geography, and high-risk flood zones. This comparison enables a more comprehensive understanding of the factors contributing to the observed spatial differences in treatment effects.

Table 4 compares the mean values of the covariates in the scenes with low CATEs (Column 1) and high CATEs (Column 2). The large standardized mean differences (SMDs) reported in Column 3 indicate that these two groups are systematically different along many dimensions. Locations with higher wetland flood protection services tend to have greater exposed property values (35.28 million USD compared to 22.96 million USD), a higher proportion of land in high-risk flood zones (A and V zones), lower elevation (10.36 m compared to 31.92 m), closer proximity to the shoreline (8.52 km compared to 43.35 km), and

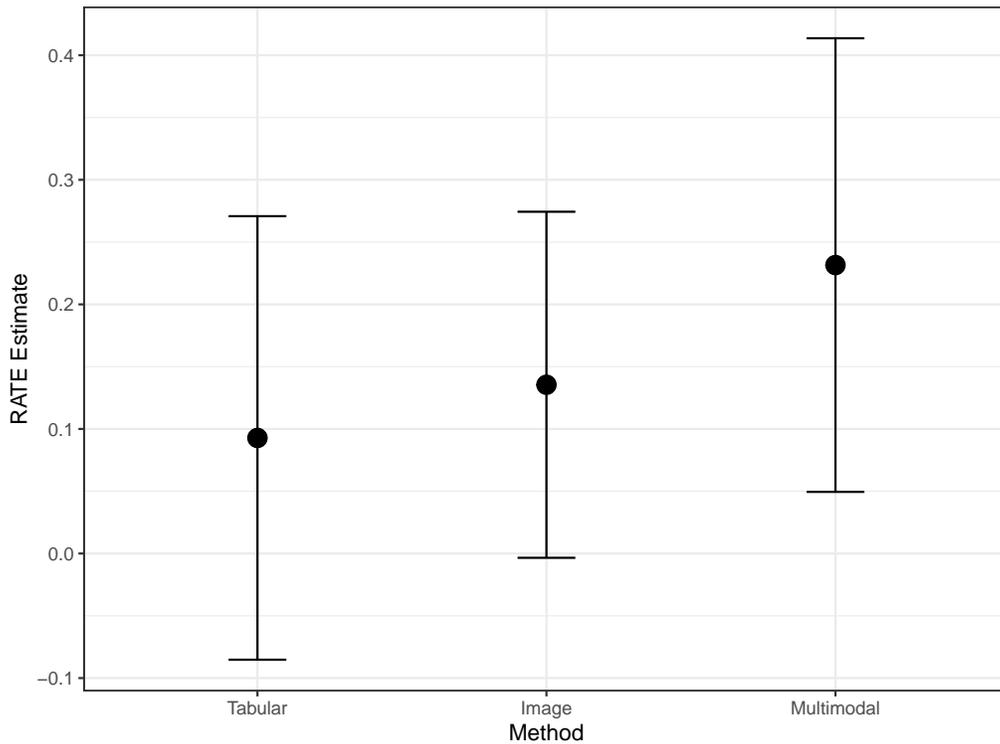


Figure 7. : Estimated Ranked Average Treatment Effect (RATE) ratio for the tabular, image only, and multimodal models. *Note: The multimodal model achieves the highest RATE estimate, demonstrating its superior ability to rank and capture treatment effect heterogeneity compared to the tabular and image-only models. The error bars represent 95% confidence intervals.*

shallower water table depths (65.56 m compared to 107.27 m). Additionally, these areas exhibit a higher share of developed and agricultural land, alongside lower proportions of forest, vegetation, and wetland cover.

These patterns align with intuitive expectations. Wealthier areas with more expensive properties and infrastructure face higher economic risks, leading to amplified flood damage claims when wetlands are developed. Lower elevations increase susceptibility to flooding due to proximity to water bodies and reduced drainage capacity, making wetlands in these areas crucial for flood mitigation. Shallow water tables make the land more prone to saturation, where wetland loss disrupts natural water storage capacity, further escalating flood risks. Proximity to shorelines exposes areas to storm surge and tidal flooding, where wetlands act as essential buffers. Wetland development in these regions removes this natural defense, intensifying flood vulnerability.

Land use patterns further explain the heightened flood risks in areas with greater wetland services. Developed and agricultural areas rely heavily on wet-

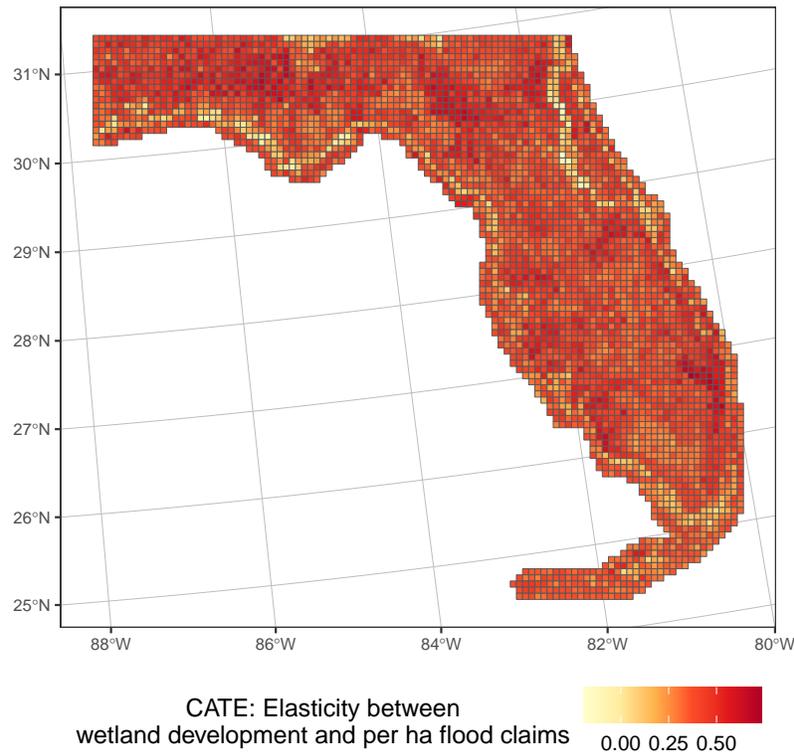


Figure 8. : Spatial heterogeneity of CATEs using multimodal model *Note: The multimodal causal framework effectively captures fine-resolution heterogeneity by incorporating site-specific characteristics, such as land cover, flood risk, and property values. This granular analysis supports more precise and targeted conservation planning.*

lands for stormwater management, and their development increases surface runoff and flood damages. Forested and vegetated wetlands serve critical roles in absorbing water, reducing surface runoff, and slowing floodwaters. Their loss diminishes these protective functions, leading to heightened flood impacts. Similarly, high and moderate-risk flood zones are naturally prone to flooding, and wetland development removes crucial barriers that could otherwise reduce damages. Collectively, these geographic, economic, and land use factors highlight the critical role of wetlands in flood protection and emphasize the amplified risks associated with their development in vulnerable regions.

Together, these findings illustrate how property values, land cover, geographic features, and flood zone classifications interact to create complex spatial heterogeneity in wetland benefits, underscoring the importance of multimodal analyses to capture these dynamics fully. Traditional tabular models alone cannot adequately account for such intricate relationships, highlighting the value of integrating diverse data sources in environmental impact studies.

C. Robustness

This section is under development. Our results demonstrate robustness across a range of model specifications. We validate the Average Treatment Effect (ATE) using the *R*-Learner framework by comparing it with the outcomes of an upstream-downstream natural experiment analyzed through a differences-in-differences approach described in Section II.D. The estimated ATE falls within the range reported in previous studies (Taylor and Druckenmiller, 2022). Additionally, we will show that the results remain consistent across various parameter settings.

IV. Discussion and Policy Implications

This study presents a significant advancement in accurately valuing wetland benefits for flood mitigation by integrating state-of-the-art methodologies from computer vision and causal inference. Using our multimodal causal framework, which combines Vision Transformer-based image embeddings with the *R*-Learner framework, we overcome the limitations of traditional tools designed for tabular data analysis. Existing approaches, while effective at modeling localized ecosystem services, often fail to capture the richness and complexity of multi-modal data. Our approach addresses this gap, enabling a detailed examination of the spatial and economic heterogeneity in wetland benefits.

Traditional benefit-transfer methods, which apply generalized estimates to new contexts, are often unreliable due to their inability to account for local geographic and socio-economic conditions. In contrast, our framework produces spatially resolved, site-specific valuations that enable policymakers and resource managers to make more informed decisions. For example, areas with high exposed property value, lower elevation, and proximity to flood-prone zones benefit the most from wetland conservation. With these insights, resource managers can compare the flood protection services of different wetlands and prioritize those with the highest potential to reduce flood damages. Moreover, by integrating heterogeneous data sources, our framework allows for a more nuanced understanding of ecosystem services that extends beyond simple tabular analyses.

While this study focuses on wetlands in Florida, our framework is versatile and generalizable to other regions and applications. It could be used to evaluate ecosystem services globally wherever rich multimodal datasets are available. For example, in agricultural regions, the framework could assess the impact of forest loss on crop yields by considering how forests regulate local microclimates, such as water cycles and solar radiation. Similarly, in urban areas, it could quantify the cooling effects of urban greening projects, such as trees and green roofs, on reducing extreme heat events, which are linked to mortality and morbidity. By identifying the most effective locations and strategies for ecosystem conservation, our framework provides a robust tool for optimizing environmental policies and investments.

Our findings show that incorporating image-based data and spatial variables

significantly enhances the ability to capture heterogeneity in ecosystem services. The RATE ratio, which quantifies the ability to target high-benefit regions, increased from 0.09 in the tabular model to 0.23 in the multimodal case. Additionally, the average per-hectare value of wetlands ranged from \$185.73 in the tabular model to \$195.29 in the multimodal model. Importantly, the multimodal framework uncovered heterogeneous benefits of wetlands, with maximum CATE values translating to monetary benefits as high as \$512.65 per hectare.

These results provide policymakers with more accurate and actionable data to inform conservation decisions. By prioritizing wetlands with the highest flood mitigation benefits, resource managers can ensure that conservation investments yield maximum societal returns. Moreover, our framework advances the methodological literature in causal inference by addressing challenges related to continuous, spatially structured covariates.

In conclusion, this study underscores the transformative potential of spatially explicit and multimodal frameworks for environmental decision-making. By providing quantitative, credible estimates of wetland benefits, our framework lays the groundwork for more reliable cost-benefit analyses at local, regional, and global scales. Policymakers can use these insights to guide conservation efforts, optimize resource allocation, and mitigate the impacts of climate change and environmental degradation. This study marks an important step toward a future where ecosystem services are valued with the rigor and precision necessary to support sustainable development and human well-being.

REFERENCES

- Aronoff, D. and Rafey, W. (2023), Conservation priorities and environmental offsets: Markets for florida wetlands, Technical report, National Bureau of Economic Research.
- Assessment, M. E. (2005), *Ecosystems and Human Well-being: Synthesis*, Island Press.
- Belloni, A., Chernozhukov, V. and Hansen, C. (2014), ‘High-dimensional methods and inference on structural and treatment effects’, *Journal of Economic Perspectives* **28**(2), 29–50.
- Burke, M., Hsiang, S. M. and Miguel, E. (2015), ‘Global non-linear effect of temperature on economic production’, *Nature* **527**(7577), 235–239.
- Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W. and Robins, J. (2018), ‘Double/debiased machine learning for treatment and structural parameters’, *arXiv preprint arXiv:1608.00060* .
- Chernozhukov, V., Hansen, C., Kallus, N., Spindler, M. and Syrgkanis, V. (2024), ‘Applied causal inference powered by ml and ai’, *rem* **12**(1), 338–402.

- Costanza, R., Anderson, S. J., Sutton, P., Mulder, K., Kubiszewski, I., Wang, X., Lo, C. and Martinez, L. (2021), ‘The global value of coastal wetlands for storm protection’, *Global Environmental Change* **70**, 102328.
- Deryugina, T., Heutel, G., Miller, N. H., Molitor, D. and Reif, J. (2019), ‘The mortality and medical costs of air pollution: Evidence from changes in wind direction’, *American Economic Review* **109**(12), 4178–4219.
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J. and Hounsby, N. (2021), ‘An image is worth 16x16 words: Transformers for image recognition at scale’, *International Conference on Learning Representations (ICLR)* .
URL: <https://arxiv.org/abs/2010.11929>
- Everard, M. (2014), *Ecosystem Services: Key Issues*, Routledge.
- Everard, M. (2021), *Ecosystem Services: Key Issues*, 2nd edn, Routledge, London. eBook Published 30 December 2021.
- Jerzak, C. T. and Daoud, A. (2023), ‘Causalimages: An r package for causal inference with earth observation, biomedical, and social science images’, *arXiv preprint arXiv:2310.00233* .
URL: <https://arxiv.org/abs/2310.00233>
- Jerzak, C. T., Johansson, F. and Daoud, A. (2023), Image-based treatment effect heterogeneity, in ‘Proceedings of the Second Conference on Causal Learning and Reasoning (CLeaR)’, Vol. 213 of *Proceedings of Machine Learning Research (PMLR)*, pp. 531–552.
- Mendelsohn, R. and Olmstead, S. (2009), ‘The economic valuation of environmental amenities and disamenities: Methods and applications’, *Annual Review of Environment and Resources* **34**, 325–347.
- Nie, X. and Wager, S. (2021), ‘Quasi-oracle estimation of heterogeneous treatment effects’, *Biometrika* **108**(2), 299–319.
- Robinson, P. M. (1988), ‘Root-n consistent semiparametric regression’, *Econometrica: Journal of the Econometric Society* pp. 931–954.
- Schlenker, W. and Roberts, M. J. (2009), ‘Nonlinear temperature effects indicate severe damages to us crop yields under climate change’, *Proceedings of the National Academy of Sciences* **106**(37), 15594–15598.
- Sun, F. and Carson, R. T. (2020), ‘Coastal wetlands reduce property damage during tropical cyclones’, *Proceedings of the National Academy of Sciences* **117**(11), 5719–5725.

- Taylor, C. and Druckenmiller, H. (2022), ‘Data and code for: Wetlands, flooding, and the clean water act’, *American Economic Review* **112**(4), 1334–1363.
- Wager, S. and Athey, S. (2018), ‘Estimation and inference of heterogeneous treatment effects using random forests’, *Journal of the American Statistical Association* **113**(523), 1228–1242.
- Wing, O. E., Pinter, N., Bates, P. D. and Kousky, C. (2020), ‘New insights into us flood vulnerability revealed from flood insurance big data’, *Nature Communications* **11**(1), 1444.
- Yadlowsky, S., Fleming, S., Shah, N., Brunskill, E. and Wager, S. (2024), ‘Evaluating treatment prioritization rules via rank-weighted average treatment effects’, *Journal of the American Statistical Association* .
URL: <https://doi.org/10.1080/01621459.2024.2393466>