A Multimodal Causal Framework for Large-Scale Ecosystem Valuation: Application to Wetland Benefits for Flood Mitigation

Anonymous Author(s)
Affiliation
Address
email

Abstract

Climate change is poised to alter wetland ecosystems through changes in temperature and precipitation patterns, compounding the already pronounced influence of human-driven wetland development. In this context, policymakers and environmental managers would benefit from accurate wetland valuations to guide their decision-making, as their choices regarding this critical natural resource directly impact flood mitigation efforts, biodiversity conservation, and economic activity. This paper introduces a novel multimodal causal framework for producing location-specific ecosystem valuations at a national scale to be used in cost-benefit policy analysis. It leverages recent advances in estimating heterogeneous treatment effects to flexibly determine how the expected impact of ecosystem-level changes—such as wetland loss via development—varies conditional on high-dimensional and multimodal measures that characterize the complex interactions between human and natural systems such as aerial satellite imagery, weather sequence data, land cover classifications, and water surface networks. From this effort, we aim to create a national database of location-specific wetland valuations in an approach that can be readily extended in estimating the effect of other interventions on ecosystems. We also plan to generate open-source feature embeddings for each U.S. wetland, embeddings that can be used to address other climate-related causal questions as well.

Ecosystems like wetlands are vital to human welfare, yet often lack market prices, complicating efforts to integrate environmental considerations into policy and economic decisions around climate change [Bar13]. Despite growing interest in quantifying ecosystem values, large-scale estimates are scarce, even though improving benefits transfer methods is recognized as necessary for increasing the role of ecosystem valuations in guiding environmental decision-making [MO09]. In this paper, we introduce a novel multimodal causal framework to generate location-specific ecosystem valuation at a national scale. Our key innovation lies in combining advances in computer vision and causal inference to leverage vast quantities of data characterizing natural and human systems—including satellite and aerial imagery, weather models and observations, geophysical and hydrologic data, administrative records, and more—to reliably model how ecosystems contribute to human well-being as a function of local systems.

We demonstrate this framework to assess the value of wetlands for flood protection. We first establish benchmarks using conventional approaches such as causal forests [WA18] to estimate Conditional Average Treatment Effects (CATEs), which characterize how local factors influence wetland flood

Submitted to Tackling Climate Change with Machine Learning: workshop at NeurIPS 2024.

protection value. However, recognizing the limitations of traditional tabular data in capturing spatial heterogeneity, we extend our framework to incorporate rich, multimodal data, including natural phenomena maps, human activity indicators, and satellite imagery. This approach aims to improve valuation precision and quantify complex environmental interactions. Our ultimate goals are to produce a national database of location-specific wetland valuations supporting conservation decisions and to generate open-source, national-scale feature embeddings for broader causal analysis in ecosystem valuation.

In what follows, we first characterize the problem tacked in this research in §1, discuss methodological questions in §2, before presenting a pilot analysis in §3 before concluding with a brief discussion.

1 Problem Formulation

As a starting point, we note that 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 ecosystems across space and time. Critically, the approach needs to be generalizable across a wide variety of settings: we want to be able to apply the same general framework to produce location-specific estimates for *any* intervention on ecosystems for which sufficiently rich data are available—enabling investigation into other policy questions related to the changing climate and the environment.

We can formalize the problem as follows: we aim to estimate the impact of an intervention, e.g., ecosystem loss, (D_i), at site i on indicators of human well-being (Y_i), taking into account all observable characteristics of the site $(\mathbf{Z}_i)$. One formulation of this general model is

$$Y_i = \beta(\mathbf{Z}_i)D_i + f(\mathbf{Z}_i) + \epsilon_i \tag{1}$$

where $\beta(\mathbf{Z}_i)$ denotes the site-specific impact of ecosystem loss on well-being. Here, $\beta(\mathbf{Z}_i)$ is a flexible function of observed characteristics, $\mathbf{Z}_i$, such that we allow for rich heterogeneity in the parameter estimate while also keeping intact the interpretation of β as the per unit impact of ecosystem loss on human well-being. We specify the impact of D_i on Y_i linearly for simplicity—in practice, this relationship will be specified flexibly using the causal machine learning techniques described below.

While the approach is general, in our application here, the outcome variable, Y_i, describes property damages from flooding, which we measure using flood insurance claims from the National Flood Insurance Program (NFIP). The treatment variable, D_i, is the conversion of wetlands to developed area, measured in log hectares (hereafter "wetland development"). We measure wetland development using data from the Coastal Change Analysis Program (C-CAP) land cover product. The treatment effect, $\beta(\mathbf{Z}_i)$, measures the elasticity between wetland development and property damages from flooding as a function of observable characteristics, $\mathbf{Z}_i$.

One innovation of this project is that $\mathbf{Z}_i$ will not only be high-dimensional, but also multimodal [Che+20], containing sources such as, $\mathbf{I_i}$ denoting aerial imagery; $\mathbf{N}_i$ describing natural systems (e.g., maps of land cover, flood zones, elevation, the water surface water network, soil types); $\mathbf{H_i}$ describing human systems (e.g., the location and values of homes, adoption of flood mitigation measures); and, $\mathbf{W_i}$ containing weather observations and climate statistics (e.g., precipitation, hurricane exposure). The set of all covariates is denoted by $\mathbf{Z_i} = \{\mathbf{I_i}, \mathbf{N_i}, \mathbf{H_i}, \mathbf{W_i}\}$.

Figure 1 shows several of these data layers at two sites in Miami-Dade, Florida. Taking full advantage of the different modalities depicted here is the core methodological challenge taken up in this research. §SI.1.1 provides further information about each data modality.

In our general framework, we will employ double machine learning (DML) and *R*-Learner methodologies [NW21] to estimate the causal effect of ecosystem loss on indicators of human well-being. DML is a procedure for estimating causal effects in observational data in the presence of high-dimensional or highly complex confounders [Che+18]. DML splits the estimation of causal effects into three prediction tasks that rely on a decomposition proposed by [Rob88] to estimate parametric components in partially linear models (one for outcome, treatment, and treatment effect). See §SI.1.2 for details; we next turn to the machine learning methods required for this approach in the multimodal ecosystem context.

2 Methods

We plan to design, train, and deploy competitive machine-learning models for the DML prediction tasks. We face several methodological challenges. First, the models must input multimodal data, $\mathbf{Z}_i$. Hence, we need to build a model that takes in inputs of varying spatial and temporal resolutions and learns to draw connections across the modalities. In addition, we want to use machine learning models that capture outputs differing in their character: specifically, while two of the DML models predict treatment and outcome as directly observable quantities, the third is learned through minimizing a causal loss.

Unified Data Representation Using Self-Supervised Learning. The first challenge to model development is that the input modalities differ significantly in how they are captured and stored, generating significant challenges due to the multi-phase, resolution, and source nature of these data [Men14]. To develop a model that processes these diverse inputs to output a single value, we will design unified representations integrating the spatial and temporal resolutions of our input data. Using self-supervised learning techniques, we will transform our input data into a high-dimensional tabular form within a unified representation space. Specifically, we will map input measurements to discrete latent codewords, forming a codebook learned purely from the data by minimizing reconstruction loss on a masked portion of the spatiotemporal arrays [FLH+22]. Consequently, all components of $\mathbf{Z}_i$ will be mapped to a latent codeword of fixed dimension d using the corresponding codebook for that modality (where the codebook represents a collection of d-dimensional vectors). A benefit of using a discrete, as opposed to continuous, latent space is that the discrete approach facilitates interpretability in the resulting representation (as latent features are characterized by presence or absence, as opposed to real-valued magnitudes) [TYG24].

Network Architecture. A related modeling challenge relates to appropriately capturing complex interactions within the input data that could lead to a robust and high-performing model using unified data representations. Unlike traditional convolutional neural networks (CNNs) that are limited by their fixed receptive fields, transformers can attend to any part of the input sequence, regardless of distance. This global attention mechanism is crucial for our task, where, e.g., the impact of wetland loss in one area may affect flood damages in distant, hydrologically connected regions. Transformers thus enable reasoning across various data modalities across space and time. See §SI.1.3 for training details.

3 Pilot Results & Conclusion

We present pilot results in Figure 2 valuations derived from tabular covariates. This figure shows the subwatershed-level estimates of CATEs (elasticities between wetland development and per hectare flood insurance claims) in panel *a* and the project impacts (effect of a 1% increase in wetland development on flood insurance claims, measured in log claims per hectare) in panel *b*. The figure also presents histograms of the estimated CATEs and projected impacts in panel *d*, along with a validation exercise showing the group average treatment effects (GATEs) for each quartile of the CATE distribution. The results are promising in that there is a notable heterogeneity signal present in the data, heterogeneity that will be further explored using the multimodal approach developed next. In this next research phase, we will expand these results to the multimodal case, resulting in a national database of location-specific wetland valuations, along with embeddings derived for each U.S. wetland.

In conclusion, in the context of climate change, decision-making critically depends on high-quality estimates of the costs and benefits of different actions taken in the context of competing policy priorities and limited resources. The approach here, we hope, will provide useful information to decision-making this critical wetland resources under threat from changing climactic conditions. □

References

[Rob88] Peter M. Robinson. “Root-N-consistent semiparametric regression”. In: *Econometrica: Journal of the Econometric Society* (1988), pp. 931–954.

[MO09] Robert Mendelsohn and Sheila Olmstead. “The economic valuation of environmental amenities and disamenities: methods and applications”. In: *Annual Review of Environment and Resources* 34 (2009), pp. 325–347.

[Bar13] Edward B. Barbier. “Valuing Ecosystem Services for Coastal Wetland Protection and Restoration: Progress and Challenges”. In: *Resources* 2.3 (2013), pp. 213–230. DOI: 10.3390/resources2030213. URL: https://www.mdpi.com/2079-9276/2/3/213.

[Men14] Xiao-Li Meng. “A trio of inference problems that could win you a Nobel prize in statistics (if you help fund it)”. In: *Past, present, and future of statistical science* (2014), pp. 537–562.

[Che+18] Victor Chernozhukov et al. “Double/debiased machine learning for treatment and structural parameters”. In: *The Econometrics Journal* 21.1 (2018), pp. C1–C68.

[WA18] Stefan Wager and Susan Athey. “Estimation and inference of heterogeneous treatment effects using random forests”. In: *Journal of the American Statistical Association* (2018).

[Che+20] Peng Chen et al. “A survey of multimodal machine learning”. In: *Chinese Journal of Engineering* 42.5 (2020), pp. 557–569.

[NW21] Xinkun Nie and Stefan Wager. “Quasi-oracle estimation of heterogeneous treatment effects”. In: *Biometrika* 108.2 (2021), pp. 299–319.

[FLH+22] Christoph Feichtenhofer, Yanghao Li, Kaiming He, et al. “Masked autoencoders as spatiotemporal learners”. In: *Advances in Neural Information Processing Systems* 35 (2022), pp. 35946–35958.

[TYG24] Sabera Talukder, Yisong Yue, and Georgia Gkioxari. “TOTEM: TOkenized Time Series EMbeddings for General Time Series Analysis”. In: *arXiv preprint arXiv:2402.16412* (2024).

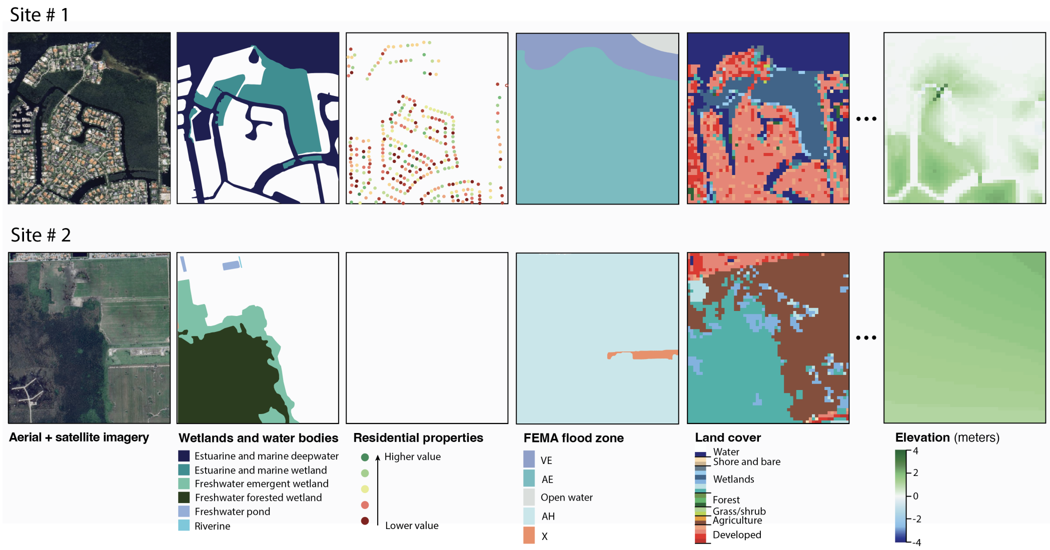

Figure 1: **Example of two wetland sites in Miami Dade, FL.** Site 1 (TOP) is a residential area along the coast. Site 2 (BOTTOM) is an inland agricultural area. Each column shows a different data layer (LEFT to RIGHT): aerial imagery, wetland polygons and types, the location and value of properties, flood zones, land cover, and elevation. These are just a few examples of the inputs our method will consider.

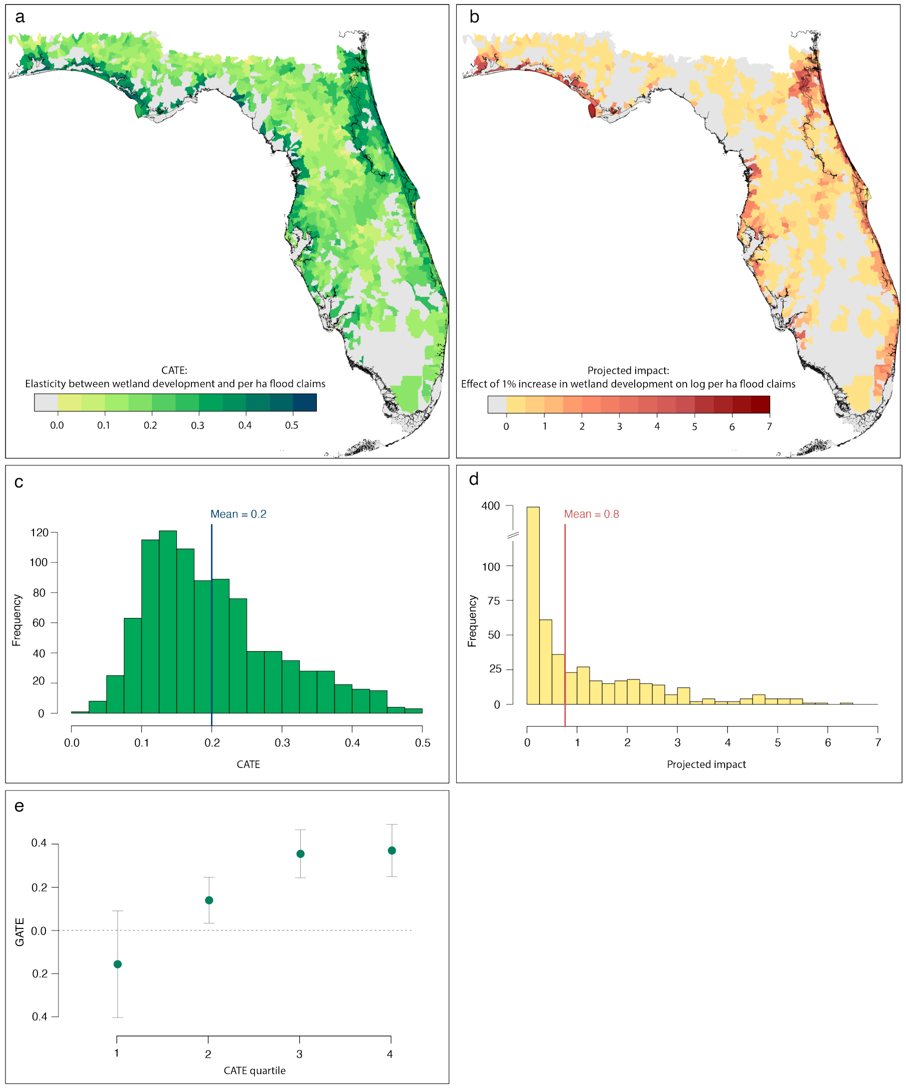

Figure 2: **Pilot analysis of wetland flood mitigation services in Florida subwatersheds.** TOP ROW shows subwatershed-level estimates of CATEs (elasticities between wetland development and per hectare flood insurance claims) in panel a and the project impacts (effect of a 1% increase in wetland development on flood insurance claims, measured in log claims per hectare) in panel b. Grey areas are omitted from the analysis because there are zero properties in these subwatersheds. MIDDLE ROW shows histograms of the CATEs in panel c and projected impacts in panel d. BOTTOM ROW is a validation exercise that shows the group average treatment effects (GATEs) for each quartile of the CATE distribution.

SI. Supplementary Information

SI.1.1 Data

Table SI.1 lists the observable factors at each site, along with their native spatial and temporal resolutions, which vary widely. We can observe our outcome variable, property damages from flooding (Y), at the daily level. The NFIP is the dominant insurer for flooding in the US, with over 4.7 million policies and \$1.28 trillion in coverage. We use NFIP claims as our outcome variable because they provide a monetized measure of property damages from flooding that can be directly employed in a cost-benefit analysis. Additionally, the richness of the data in terms of temporal span, spatial extent and granularity, and consistency in the measurement makes our analysis empirically tractable. NFIP participation is greater than 50% among homes located in floodplains because the program subsidized and coverage is required for all homeowners with federally-backed mortgage.

However, one limitation of using these data to measure flood damages is that they do not capture damages that occur outside the NFIP and will therefore likely underestimates of the flood protection services of wetlands. We can also observe exposure to extreme weather—a necessary condition for experiencing flood damages— at the daily level. We can only observe wetland area changes every five years, and data on physical features (e.g., vector data on the surface water network) tend to only be available for a single snapshot in time. Human features (e.g., the location and value of homes) are often available annually. Spatial resolutions range from individual points to coarse administrative regions. In specifying our model, we will experiment with different spatial units of analysis while allowing for interdependencies between locations that are hydrologically connected. Our study period spans the years 1990 to 2023, and all variables are available for the entire contiguous United States.

SI.1.2 Double Machine Learning Details

Robinson (1988) notes that estimating the relationship between Y_i and D_i conditional on $\mathbf{Z}_i$ is equivalent to a three-step process.

First, we model treatment, D_i, as a function of pre-treatment observables, $\mathbf{Z}_i$:

$$D_i = g(\mathbf{Z}_i) + \eta_i \tag{2}$$

Second, we model the outcome, Y_i, also as a function of $\mathbf{Z}_i$:

$$Y_i = h(\mathbf{Z}_i) + \nu_i \tag{3}$$

Finally, we regress the residuals of Equation 3 on the residuals of Equation 2. That is, we can estimate the Average Treatment Effect (ATE), β, using the equation,

$$\{Y_i - h(\mathbf{Z}_i)\} = \alpha + \beta\{D_i - g(\mathbf{Z}_i)\} + \mu_i \tag{4}$$

In the Double Machine Learning (DML) approach, the functions $g(\cdot)$ and $h(\cdot)$ can be parameterized using any machine learning model. A primary advantage of DML is its systematic approach to controlling for confounding variables, even when these confounders are high-dimensional and complex.

To estimate the potential heterogeneity in treatment effect, we propose using the *R*-Learner [NW21], which is a special case of the more general DML framework. In this estimator, the CATE is estimated by using the following estimating equation:

$$\hat{\beta} = \arg\min_{\beta} \mathbb{E}_n \left[\left(\tilde{Y}_i - \beta(\mathbf{Z}_i) \cdot \tilde{D}_i \right)^2 \right] \tag{5}$$

	Data source (1)	Variables (2)	Spatial resolution (3)	Temporal resolution (4)
Outcome				
Flood damages	National Flood Insurance Program (NFIP)	Flood insurance claims paid	Census block	Daily
Treatment				
Wetland development	USGS Coastal Change Analysis Program (C-CAP)	Land converted from wetland to developed area	30 meter	5-year
Covariates				
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 (C-CAP)	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

Table SI.1: **Variables and data formats.** 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 1990 to 2023.

where $\tilde{Y}_i = Y_i - \mathbb{E}[Y_i \mid \mathbf{Z}_i]$ and $\tilde{D}_i = T_i - \mathbb{E}[D_i \mid \mathbf{Z}_i]$ denotes the residual outcome and residual treatment in Equations 2 and 3, respectively. What distinguishes the *R*-Learner in the realm of causal inference is its non-parametric nature at the last stage (Equation 4). Unlike traditional regression approaches that might assume a specific functional form for $\beta(\mathbf{Z}_i)$, the *R*-Learner allows for a more flexible estimation. This flexibility is crucial when the true relationship between the treatment and the outcome is complex and not well-modeled by parametric forms. By not imposing a predefined shape or relationship, the R-Learner can adapt to the underlying patterns in the data, potentially leading to more accurate and insightful estimates of the treatment effects across different subpopulations. The non-parametric functions in this method are modeled using techniques from computer vision, discussed in §2.

SI.1.3 Optimization & training recipes for causal model learning

To model and train the function $\beta(\cdot)$, we must solve a structured prediction problem. Unlike $h(\cdot)$ and $g(\cdot)$, where we know the true outputs in the training set, we lack a direct signal to supervise the output of $\beta(\cdot)$. Instead, we aim for $\beta(\cdot)$ to optimize the objective in Equation 6, making the task of learning a generalizable $\beta(\cdot)$ non-trivial:

$$\mathcal{L}(\beta(\mathbf{Z_i})) = \frac{1}{N}\sum_{i=1}^{N}\left(\tilde{Y}_i - \beta(\mathbf{Z}_i)\tilde{D}_i\right)^2 \tag{6}$$

We explore structured prediction techniques that leverage unsupervised and self-supervised learning. Contrastive learning and feature regression models have been shown to produce generalizable feature representations in vision tasks.

Additionally, we will need to explore the best way to parameterize the impact of wetland loss on property damages from flooding. Equation 6 models the effect of wetland loss on flood damages linearly. This specification is appealing due to its simplicity, but the true relationship may take another form. We will explore other parametric (e.g., quadratic) and non-parametric (e.g., binned) models. We will also experiment with how to specify interdependencies between hydrologically connected areas.