

# Structured Pixels: Satellite Imagery as the Cause in Causal Effect Estimation

Chien Lu  
Trinity College Dublin  
[luc4@tcd.ie](mailto:luc4@tcd.ie)

Thomas Chadefaux  
Trinity College Dublin  
[thomas.chadefaux@tcd.ie](mailto:thomas.chadefaux@tcd.ie)

## Abstract

We present *Structured Pixels (SP)*, a causal inference model that positions satellite imagery as a cause/treatment in a causal graph, rather than merely a proxy for outcomes or confounders. Built on the generalized Robinson decomposition and a two-step, *R*-learner-inspired algorithm, *SP* uses learned latent representations to partial out confounding influences and isolate the causal effect. Its modular training pipeline supports integration with diverse machine learning models across domains. We evaluate *SP* using semi-synthesized datasets on two tasks: the impact of environmental conditions on mosquito populations and the influence of coastal characteristics on dark vessel prevalence. *SP* consistently outperforms baseline methods, and its learned representations capture meaningful environmental patterns. We further demonstrate *SP*'s applicability by re-examining the relationship between deforestation and agricultural productivity with real-world data; the results align with prior work. These findings highlight *SP*'s potential to advance GeoAI for environmental monitoring and resource management.

**Keywords:** GeoAI, Satellite Imagery, Causal Inference, Representation Learning

## 1. Introduction

GeoArtificial Intelligence (GeoAI) has gained prominence in modern information systems, driven by the growing availability of geographic data and the advancement of machine learning (Darwish, 2025). This combination has shown potential to address pressing global challenges, improve our understanding of complex phenomena, and inform decision-making

across key areas such as environmental monitoring and maritime security. Satellite imagery data plays an important role in GeoAI, enabling large-scale, long-term information gathering in situations where ground-based surveys are impractical. It has proved valuable in various fields such as environmental science (Ford et al., 2009) and socio-political research (Witmer, 2015).

The intersection of causal inference and satellite imagery has gained increasing attention (Jerzak et al., 2023). Causal inference provides a framework for estimating the effect of a cause on an outcome from observational data (Pearl, 2009), quantifying the impact of interventions or natural changes on outcomes.

However, most of the proposed causal inference methods involving satellite imagery have limited them to supporting roles. These approaches often focused on binary treatments and mainly use satellite imagery to extract outcome variables of interest, e.g., tree cover (Gordon et al., 2023), Normalized Difference Vegetation Index (NDVI) (Fick et al., 2021), or as proxies for confounders (Conlin, 2024; Jerzak et al., 2022). Although they are valuable, they are not applicable when addressing causal questions in which environmental features captured by satellite imagery naturally serve as the driving force for change. Satellite imagery captures environmental factors in an integrated manner, collectively influencing the outcome of interest. Viewing the broader environmental factors as the main driving force gives rise to intriguing causal questions, e.g., “*What would the mosquito population be if the environment in this area changed in a specific way?*”, “*Would the presence of dark vessels differ if the coastline had different characteristics?*”, “*Is the crop production affected by forest protection?*”.

These *what-if* questions reflect the perspective of treating the environment as the primary contributing

factor. Conceptualizing these environmental factors as the cause requires a rigorous analytical framework for quantifying the isolated impacts that can help researchers or policy-makers better understand environmental interactions and draw more credible conclusions. Nevertheless, directly treating satellite imagery as the cause remains largely unexplored. One reason is that treating satellite imagery as a cause introduces challenges due to its high-dimensional nature, which includes numerous spatially and spectrally organized environmental variables.

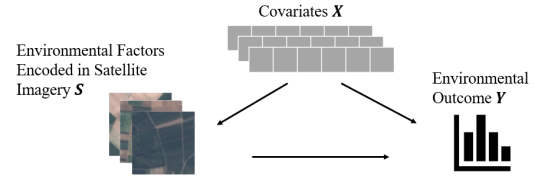
We propose *Structured Pixels* (SP), a novel method that positions satellite imagery as cause within a causal graph. SP employs a two-step estimation algorithm to isolate the causal estimand based on generalized Robinson decomposition (Kaddour et al., 2021), a recently proposed theory for causal effect estimation. Together with recent advancements in machine learning for tackling high-dimensional causal inference problems (Schölkopf, 2022), SP extracts a set of essential embedding vectors to disentangle confounding effects in the observational data. This yields an accurate estimation of causal effects through a modular, flexible model training pipeline, placing satellite imagery at the core of the causal inference framework and allowing crucial “*what-if*” questions to be addressed. Our contributions are threefold:

- **Satellite imagery as Cause:** SP pioneers formalizing satellite imagery as cause, bridging causal inference and satellite imagery analysis.
- **Novel representation approach:** SP learns representation vectors from observational data for causal effect estimation in satellite imagery with a flexible, modular training algorithm.
- **Empirical validation:** SP is evaluated using semi-synthesized and real-world datasets, addressing key issues: environmental impacts on mosquito populations, coastal features’ influence on dark vessel prevalence, and deforestation effects on agricultural production.

This paper is organized as follows: Section 2 introduces the problem setting and key terminologies. The proposed model and training algorithm are described in Section 3. Experiments and real-world data analysis are presented in Section 4 and 5, respectively. Finally, Section 6 discusses and concludes this paper.

## 2. Preliminaries and Problem Setting

Let  $i$  denote the index of an observed region, where  $\{y^{(i)}, s^{(i)}, x^{(i)}\}$  is a set of the corresponding



**Figure 1. Causal relations between environmental factors encoded in the satellite imagery ( $S$ ), covariates ( $X$ ), and outcome ( $Y$ ) variables.**

observational data, which contains:

- $y^{(i)} \in \mathbb{R}$ : A real-valued outcome variable of interest in the region  $i$ , such as the size of the mosquito population or the total number of dark vessels in the area.
- $s^{(i)} \in \mathbb{R}^{C \times H \times W}$ : The satellite imagery of the region, which encodes essential environmental features that influence the outcome.  $C$  denotes the number of channels (or bands), and  $H$  and  $W$  are the height and width, respectively.
- $x^{(i)} \in \mathbb{R}^K$ : Covariates that introduce confounding effects in the observational data. That is, they can influence both the formation of the observed satellite imagery  $s^{(i)}$  and the outcome variable  $y^{(i)}$ . For example, human activities can simultaneously affect the appearance of the environment in a crop field, thereby differing satellite imagery, as well as the mosquito population. Here, we use  $K$  to denote the total number of covariates.

We hypothesize an alternative scenario in which the same covariates,  $x^{(i)}$ , are observed in a counterfactual environment, represented by an alternative satellite image  $\bar{s}^{(i)}$  (e.g., depicting a different crop field or coastal area). Our goal is to estimate the effect of replacing the original environment with this hypothetical, counterfactual one. To quantify this effect, we use the Conditional Average Treatment Effect (CATE) (Heckman et al., 1997), a standard causal estimand. Here, the treatment, which is encoded in the satellite imagery  $\bar{s}^{(i)}$ , generates a counterfactual outcome  $\bar{y}^{(i)}$ , conditioning on  $x^{(i)}$ . For example, we might estimate the change in mosquito population resulting from replacing a forest image with a crop-field image for region  $i$ . An illustration of causal relations between variables can be found in Figure 1. In this causal problem, the treatments are two distinct environments, represented by satellite images  $\bar{s}$  and  $s$ . The estimand CATE is mathematically defined as:

$$\tau(\bar{s}, s, \mathbf{x}) \triangleq E[Y | \text{do}(\mathbf{S} = \bar{s}), \mathbf{X} = \mathbf{x}] - E[Y | \text{do}(\mathbf{S} = s), \mathbf{X} = \mathbf{x}]. \quad (1)$$

Since directly experimenting with a counterfactual environment  $\bar{s}$  is often infeasible, we infer these effects from observational data, relying on two standard causal inference assumptions (de Luna and Johansson, 2006; Imbens and Wooldridge, 2009):

**Assumption 1 (Unconfoundedness)** *A treatment  $s$  is independent of potential outcome  $y$  given covariates  $\mathbf{x}$ , that is,  $P(Y \leq y | \text{do}(\mathbf{S} = s), \mathbf{X} = \mathbf{x}) = P(Y \leq y | \mathbf{S} = s, \mathbf{X} = \mathbf{x})$ .*

**Assumption 2 (Positivity)** *For every subpopulation defined by  $\mathbf{X} = \mathbf{x}^{(i)}$ , there's a non-zero probability of any treatment, i.e.,  $0 < P(s | \mathbf{X} = \mathbf{x}^{(i)}) < 1$  for all  $s$ .*

With these assumptions, we can reformulate the causal estimation of CATE in (1) using observational data to:

$$\tau(\bar{s}, s, \mathbf{x}) = E[Y | \mathbf{S} = \bar{s}, \mathbf{X} = \mathbf{x}] - E[Y | \mathbf{S} = s, \mathbf{X} = \mathbf{x}]. \quad (2)$$

This reformulation hinges on unconfoundedness and positivity assumptions, allowing us to estimate CATE from observational distributions rather than conducting direct experiments. Note that, while one can directly train an arbitrary prediction model to obtain the statistical estimands  $E[Y | \mathbf{S} = \bar{s}, \mathbf{X} = \mathbf{x}]$  and  $E[Y | \mathbf{S} = s, \mathbf{X} = \mathbf{x}]$  from observational data, this direct estimation suffers bias (Chernozhukov et al., 2018; Künzel et al., 2019) due to the association between  $\mathbf{X}$  and  $\mathbf{S}$ , this confounding effect makes it difficult to identify the distinct contribution of  $\mathbf{S}$  to  $Y$ . We propose SP to mitigate this bias when relying on observational data.

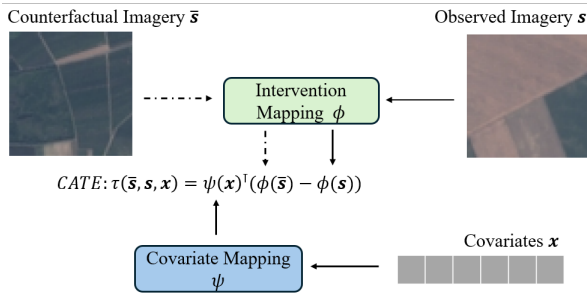


Figure 2. Illustration of CATE estimation.

### 3. Proposed Method

In this section, we present *Structured Pixels* (SP). SP is grounded in a recently proposed theory of generalized

Robinson decomposition (Kaddour et al., 2021), an extension of the classical R-learner (Nie and Wager, 2021) where the treatment is a binary variable. The key notion of the theory is to generalize the R-learned and view each  $s$  as a “structured treatment” which has a corresponding set of treatment features  $\phi(s) \in \mathbb{R}^d$ , an embedding vector residing in a  $d$ -dimensional reproducing kernel Hilbert space (RKHS) with the mapping  $\phi(\cdot) : \mathbb{R}^{C \times H \times W} \rightarrow \mathbb{R}^d$ . The contribution of the treatment to the outcome  $Y$ , conditional on covariates  $\mathbf{X}$ , is assumed to take the form of an inner product of embedding vectors:

$$Y = \psi(\mathbf{X})^\top \phi(\mathbf{S}) + \varepsilon \quad (3)$$

where  $\psi(\cdot) : \mathbb{R}^K \rightarrow \mathbb{R}^d$  is another mapping for  $\mathbf{X}$  and  $\varepsilon$  is the error term. Based on this product effect assumption, two *nuisance estimators*:  $m(\cdot)$ , and  $e^\phi(\cdot)$  are further introduced to partial out the confounding associations. This extends the classical Robinson decomposition (Robinson, 1988) to high-dimensional treatments such as satellite imagery. By mapping the treatment into a low-dimensional embedding, it enables efficient estimation while capturing spatial and spectral features beyond traditional binary or scalar treatments.

#### Theorem 1 (Generalized Robinson Decomposition)

*The confounding impact of  $\mathbf{X}$  on both  $\mathbf{S}$  and  $Y$  can be partialled out through the decomposition:*

$$Y - m(\mathbf{X}) = \psi(\mathbf{X})^\top (\phi(\mathbf{S}) - e^\phi(\mathbf{X})) + \varepsilon. \quad (4)$$

where  $m(\cdot) : \mathbb{R}^K \rightarrow \mathbb{R}$  is the mean outcome model estimating the outcome variable  $Y$  with covariates  $\mathbf{X}$ . And  $e^\phi(\cdot) : \mathbb{R}^d \rightarrow \mathbb{R}^d$  estimates the treatment features of  $\mathbf{S}$  with covariates  $\mathbf{X}$  defined as  $e^\phi(\mathbf{X}) \triangleq \mathbb{E}[\phi(\mathbf{S}) | \mathbf{x}]$ .

Finally, CATE of the counterfactual and observed treatments  $\bar{s}$  and  $s$ , conditioning on the covariates  $\mathbf{x}$ , is re-parameterized as:

$$\tau(\bar{s}, s, \mathbf{x}) = \psi(\mathbf{x})^\top (\phi(\bar{s}) - \phi(s)) \quad (5)$$

which is the inner product of the covariate features vector  $\psi(\mathbf{x})$  and the difference between counterfactual and observed treatment features  $\phi(\bar{s})$ , and  $\phi(s)$ . We illustrate the CATE estimation in Figure 2.

#### 3.1. Two-step training process and model architecture

To derive the nuisance estimators  $\hat{m}(\cdot)$ ,  $\hat{e}^\phi(\cdot)$  and the representation learning models  $\hat{\psi}(\cdot)$  and  $\hat{\phi}(\cdot)$ . The training process of SP consists of two steps: in the first step, the mean outcome model  $\hat{m}(\cdot)$  is trained. And the second step trains the representation learning models  $\hat{\psi}(\cdot)$ ,  $\hat{\phi}(\cdot)$ , and the propensity model  $\hat{e}^\phi(\cdot)$ .

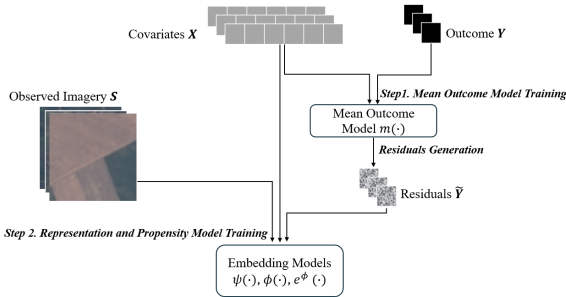
**Step 1.** The objective of this step is to train the mean outcome model  $m(\cdot)$ , which uses covariates  $\mathbf{x}$  to predict the outcome variable  $y$ , that is, to optimize the following objective function:

$$\mathcal{L}_m^{(i)} = \left( y^{(i)} - m(\mathbf{x}^{(i)}) \right)^2 + \Lambda(m) \quad (6)$$

where  $\Lambda(m)$  is the regularization term. For the mean outcome model  $\hat{m}(\cdot)$ , we employ a multilayer perceptron (MLP). Given an input  $\mathbf{x}^{(i)} \in \mathbb{R}^K$ , the network forward pass is defined as:

$$\begin{aligned} h_1 &= \text{Dropout}(\text{ReLU}(W_1 \mathbf{x}^{(i)} + b_1)) \\ h_l &= \text{Dropout}(\text{ReLU}(W_l h_{l-1} + b_l)), l = 2, \dots, L \\ \hat{y}^{(i)} &= W_{L+1} h_L + b_{L+1} \end{aligned} \quad (7)$$

where  $W_1 \in \mathbb{R}^{d_1 \times K}$ ,  $W_l \in \mathbb{R}^{d_l \times d_{l-1}}$  for  $l = 2, \dots, L$ , and  $W_{L+1} \in \mathbb{R}^{1 \times d_L}$  represent the weight matrices, and  $b_k \in \mathbb{R}^{d_l}$  for  $l = 1, \dots, L$  and  $b_{L+1} \in \mathbb{R}$  are the bias vectors.  $L$  denotes the number of hidden layers and  $K$  is the number of covariates.  $d_l$  is the dimension of the  $l$ -th hidden layer, and  $h_l \in \mathbb{R}^{d_l}$  are the hidden representations. Dropout regularization, applied during training after each ReLU activation, is set to prevent overfitting. The final output  $\hat{y}^{(i)} = m(\mathbf{x}^{(i)}) \in \mathbb{R}$  is then the predicted mean outcome, obtained through a linear transformation of the final hidden representation  $h_L$ . And we simply set  $\Lambda(m) = \lambda_m \sum_{l=1}^{L+1} \|W_l\|_F^2$ , an L2 regularization based on the Frobenius norm of the weight matrices in the objective function.



**Figure 3. Model training procedure.**

**Step 2:** The second step trains the representation learning models  $\hat{\psi}(\cdot)$ ,  $\hat{\phi}(\cdot)$  and the propensity model  $\hat{e}^\phi(\cdot)$  with the following objective functions:

$$\begin{aligned} \mathcal{L}_{\psi, \phi, e^\phi} = & \left( \sum_i \tilde{y}^{(i)} - \psi(\mathbf{x}^{(i)})^\top \left( \phi(\mathbf{s}^{(i)}) - e^\phi(\mathbf{x}^{(i)}) \right) \right)^2 \\ & + \Lambda(\psi, \phi, e^\phi), \end{aligned} \quad (8)$$

---

### Algorithm 1 SP Training Algorithm

---

**Input:** Observational data:  $\mathcal{D}_1 = \{y^{(i)}, \mathbf{x}^{(i)}\}_{i=1}^N$ ,  $\mathcal{D}_2 = \{\mathbf{s}^{(i)}, y^{(i)}, \mathbf{x}^{(i)}\}_{i=1}^N$

**Output:** Mean Outcome Model  $\hat{m}(\cdot)$ , Representation Learning Models  $\hat{\psi}(\cdot)$ ,  $\hat{\phi}(\cdot)$ , and  $\hat{e}^\phi(\cdot)$

**Step 1: Training mean outcome model**

Divide input data  $\mathcal{D}_1$  into  $M_1$  random partitions

**for** each batch  $B_m$  **do**

    Compute  $\mathcal{L}_m$  (Equation 6)

    Update  $m(\cdot)$  by minimizing  $\mathcal{L}_m$

**end for**

**Step 2: Training representation learning and propensity models**

Divide input data  $\mathcal{D}_2$  into  $M_2$  random partitions

**for** each batch  $B_m$  **do**

    Compute residuals  $\{\tilde{y}^{(i)} = y^{(i)} - \hat{m}(\mathbf{x}^{(i)})\}_{i \in B_m}$

    Update  $\psi(\cdot)$ ,  $\phi(\cdot)$ ,  $e^\phi(\cdot)$  by optimizing  $\mathcal{L}_{\psi, \phi}$  or  $\mathcal{L}_{e^\phi}$  alternatively

**end for**

---

where  $\tilde{y}^{(i)} = y^{(i)} - \hat{m}(\mathbf{x}^{(i)})$  are residuals using the trained  $\hat{m}(\mathbf{x}^{(i)})$  from step 1, and  $\Lambda$  is again an L2 regularization.

**Alternating optimization procedure.** We adopt this procedure to balance computational efficiency and model performance, avoiding the high computational cost of cross-fitting while maintaining robustness. Given that the mappings  $\hat{\psi}$ ,  $\hat{\phi}$ ,  $\hat{e}^\phi(\cdot)$  are fundamentally different. An alternating optimization approach is employed, that is taking  $j \in \{1, \dots, J\}$  steps to optimize  $\hat{\psi}$ ,  $\hat{\phi}$ , followed by one step for  $e^\phi$ . This is to ensure that the representation learning models and the propensity model are trained on distinct data partitions, at the same time avoid the computational cost of the cross-fitting procedure typically used in the R-learner algorithm.

Since representation learning models and propensity model are all generating a  $d$ -dimensional vector, for simplicity, we denote the model output as  $\hat{z}^{(i)} \in \mathbb{R}^d$  when describing the model architectures. For the satellite imagery representation learning model  $\phi(\cdot)$ , we adopt the following architecture:

$$h_L = \mathcal{M}(\mathbf{s}^{(i)}), \hat{z}^{(i)} = W_L h_L + b_L, \quad (9)$$

where  $\mathcal{M}$  can be any image encoder, and the last layer of the model is to map  $h_L$  to the desired vector space. Here, to demonstrate the flexibility of the overall pipeline, we apply two different encoders, the first one is a Convolutional Neural Network (CNN) architecture

defined as:

$$\begin{aligned}
 h_1 &= \text{MaxPool} \left( \text{ReLU} \left( \text{Conv} \left( s^{(i)} \right) \right) \right), \\
 h_l &= \text{MaxPool} \left( \text{ReLU} \left( \text{Conv} \left( h_{l-1} \right) \right) \right), l = 2, \dots, L - 1 \\
 h_L &= \text{Flatten} \left( \text{AdaptiveAvgPool} \left( h_{L-1} \right) \right).
 \end{aligned}
 \tag{10}$$

And for the second encoder, we employ *Satlas*, a Swin Transformer-based model (Liu et al., 2021) pretrained on satellite imagery (Bastani et al., 2023) to generate  $h_L$ . For models  $\psi(\cdot)$  and  $e^\phi(\cdot)$ , we again employ MLP architecture. The representation learning vector  $\hat{z}^{(i)}$  is generated by the following forward pass:

$$\begin{aligned}
 h_1 &= \text{Dropout} \left( \text{ReLU} \left( W_1 x^{(i)} + b_1 \right) \right) \\
 h_l &= \text{Dropout} \left( \text{ReLU} \left( W_l h_{l-1} + b_l \right) \right), l = 2, \dots, L - 1, \\
 \hat{z}^{(i)} &= W_L h_L + b_L,
 \end{aligned}
 \tag{11}$$

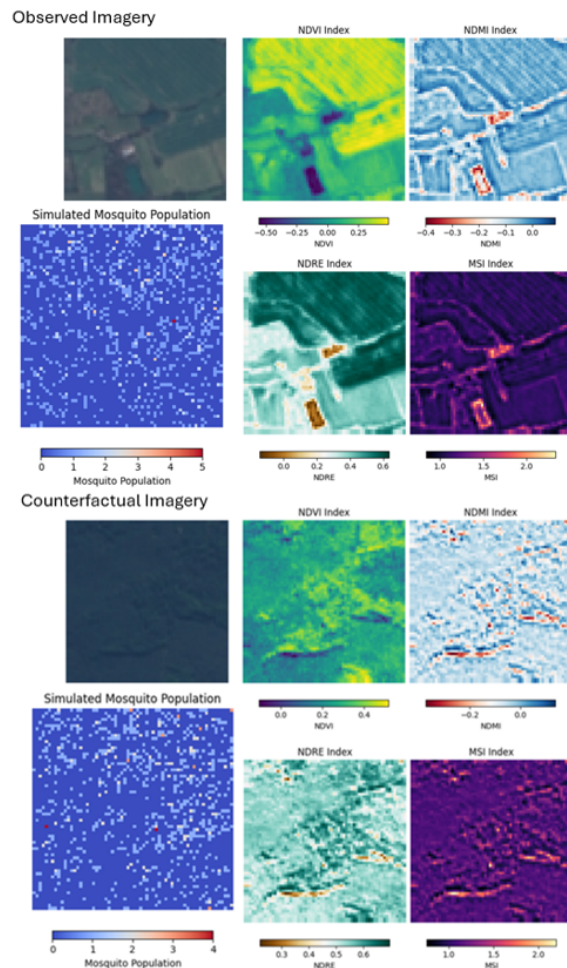
where  $W_l$  and  $b_l$  for  $l = 1, \dots, L$  are the weights and biases, and the final layer maps the last hidden representation  $h_L$  to the latent vector space. MLP is used for  $\hat{m}(\cdot)$  and  $\hat{e}^\phi(\cdot)$  due to its simplicity. CNN and *Satlas* represent canonical and complex encoder architectures. The modular training algorithm of SP allows the integration of more sophisticated models, depending on the context. The overall training process is summarized in Figure 3 and Algorithm 1.

## 4. Experiments

Evaluating causal inference methods is inherently challenging due to the lack of ground truth for the counterfactual outcomes. A widely adopted solution in the research community is the semi-synthetic dataset (Nogueira et al., 2022; Wood-Doughty et al., 2021). In a semi-synthetic dataset, treatments are from a real-world dataset, and covariates and potential outcomes are often simulated based on a specific data-generating mechanism, which is grounded in a scientific, domain-informed procedure. In the evaluation stage, causal inference methods lack access to the underlying data-generation process and must robustly estimate the synthesized causal effects using observational data alone. In this study, we designed experiments for two scenarios of contemporary importance: **Environmental Impact on Mosquito Population** and **Coastal and Ecological Influences on Dark Vessel Prevalence**. We generated corresponding semi-synthetic datasets for each scenario. The first dataset is based on *EuroSAT* (Helber et al., 2019), while the second is derived from *LICS* (O’Sullivan et al., 2024).

### 4.1. Case 1: Environmental Impact on Mosquito Population

Climate change has progressively changed humidity and vegetation across Europe, raising significant concerns about the spread of mosquito populations and mosquito-borne diseases (Becker, 2008; Brugueras et al., 2020). To evaluate whether SP can help address this issue, we repurpose the *EuroSAT* dataset, which was originally proposed for training land-use classification models, to synthesize the experimental data.



**Figure 4. Example of Case 1, including satellite imagery (RGB bands), the corresponding NDVI, NDMI, NDRE, MSI indices, and the simulated mosquito population.**

We take in total 1500 satellite images from five different land-use categories: Forest, Pasture, Annual Crop, Permanent Crop, and Herbaceous Vegetation. For each imagery  $s^{(i)} \in \mathbb{R}^{9 \times 64 \times 64}$ , we randomly generate category-dependent covariates  $x^{(i)} \in \mathbb{R}^7$  consisting of 7 simulated continuous-valued covariates with:

$$x_k^{(i)} \sim \text{Beta}(\rho_k, 10 - \rho), k = 1 \dots K. \quad (12)$$

where  $\rho_k$  is assigned based on the image's category to ensure that the covariates are category-dependent. This design is to imitate the confounding effect  $\mathbf{x}^{(i)}$  on the imagery  $s^{(i)}$ . We further compute 4 pixel-wise indices, denoted by  $\mathbf{z}_{(h,w)}^{(i)} \in \mathbb{R}^4$ . They are Normalized Difference Vegetation Index (NDVI), Normalized Difference Moisture Index (NDMI), Normalized Difference Red Edge Index (NDRE), and the Moisture Stress Index (MSI), directly computed from the satellite imagery  $s^{(i)}$ . Finally, for each pixel  $h, w$  in the  $i$ -th imagery, we sample the mosquito number  $p_{(h,w)}^{(s)}$  from a Poisson distribution:

$$p_{(h,w)}^{(i)} \sim \text{Pois}(\lambda_{h,w}) \quad (13)$$

where  $\lambda_{h,w} = e^{\beta_0 + \beta_1^\top \mathbf{x}^{(i)} + \beta_2^\top \mathbf{z}_{(h,w)}^{(i)}}$ . The coefficients  $\beta_1$  and  $\beta_2$  are sampled from a standard normal distribution. The outcome variable  $y^{(i)}$  is then the total number of mosquitos over all pixels within  $s^{(i)}$ :

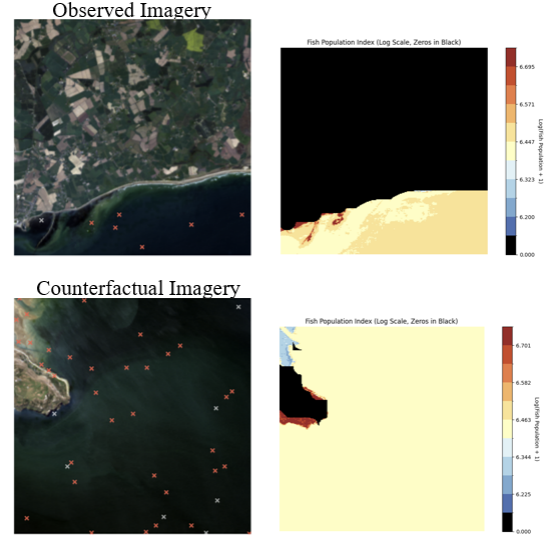
$$y^{(i)} = \sum_{h,w} p_{(h,w)}^{(i)}. \quad (14)$$

The task is to estimate the causal effect of swapping  $s^{(i)}$  for  $\bar{s}^{(i)}$ , which represents a different environment. To achieve this, we randomly sample another  $\bar{s}^{(i)}$  from the dataset and compute the corresponding counterfactual pixel-level indices  $\bar{\mathbf{z}}_{(h,w)}^{(i)}$  with the covariates  $\mathbf{x}^{(i)}$  held the same. The rate parameter used to generate the counterfactual outcome  $\bar{y}^{(i)}$  in Equation (14) is then changed to  $\bar{\lambda}_{h,w} = e^{\beta_0 + \beta_1^\top \mathbf{x}^{(i)} + \beta_2^\top \bar{\mathbf{z}}_{(h,w)}^{(i)}}$  and the corresponding ground truth causal effect is  $\tau^{(i)} = \bar{y}^{(i)} - y^{(i)}$ . An example is shown in Figure 4.

#### 4.2. Case 2: Coastal and Ecological Influences on Dark Vessel Prevalence

Illegal, unreported, and unregulated (IUU) fishing threatens sustainability, equity, and security (Park et al., 2020). We simulate dark vessel prevalence using 1,000 satellite images from the *LICS* dataset, which includes water body masks of the Irish coast.

We derive pixel-wise indices including Sea Surface Temperature (SST), Turbidity (Tu), and Chlorophyll (Ch) from spectral bands. Confounding variables are obtained following a similar approach in causal inference (Thorat et al., 2024) by applying PCA to



**Figure 5. Example of Case 2, including satellite imagery (RGB bands), the simulated fish population based on calculated indices, and locations of vessels (white crosses) and dark vessels (red crosses).**

each image  $s^{(i)} \in \mathbb{R}^{7 \times 256 \times 256}$  and using the top 10 components as covariates  $\mathbf{x}^{(i)}$ , representing factors such as economic incentives or governance structures (Agnew et al., 2009; Le Gallic and Cox, 2006). The fish population is simulated with a widely used approach in population ecology, the logistic growth differential equation (Lotka, 1925). We set the total carrying capacity of each pixel to 1000:

$$F_{(h,w)}^{(i)} = \frac{1000}{1 + \exp(-5 \times (\lambda_{(h,w)}^{(i)} - 0.5))} \quad (15)$$

where  $\lambda_{(h,w)}^{(i)} = \frac{1}{3} [SST_{(h,w)}^{(i)}(1 - |SST_{(h,w)}^{(i)} - 0.5|) \times 2 + Tu_{(h,w)}^{(i)}(1 - |Tu_{(h,w)}^{(i)} - 0.5|) \times 2 + Ch_{(h,w)}^{(i)}]$ .

From the water body mask provided in the *LICS* dataset, we further compute two pixel-level indices: coastal influence ( $C$ ) and distance effect ( $D$ ). Combined with the fish population  $F_{(h,w)}^{(i)}$ , we simulate the probability of vessel presence with:

$$p_{\text{vessel};(h,w)} = \frac{0.005 \times F_{(h,w)}^{(i)}}{10000} + 0.003D_{(h,w)}^{(i)} + 0.002C_{(h,w)}^{(i)}. \quad (16)$$

We use this probability to sample the vessel appearance at locations using with a Bernoulli distribution,  $V_{(h,w)}^{(i)} \sim \text{Bernoulli}(p_{\text{vessel};(h,w)})$  where

$V_{(h,w)}^{(i)} = 1$  indicates the presence of a vessel at pixel  $(h, w)$ . We then use the PCA-generated covariates  $x$  as an overall factor to determine whether the vessel is unreported with  $p_{\text{dark};(h,w)} \sim \text{Beta}(e^{\beta^\top x^{(i)}}, 2)$  where  $\beta$  is again the coefficients sampled from a standard normal distribution across all the observed areas. Finally we get

$$D_{\text{dark};(h,w)}^{(i)} \sim \text{Bernoulli}(p_{\text{dark};(h,w)}) \cdot V_{(h,w)}^{(i)} \quad (17)$$

where  $D_{\text{dark};(h,w)}^{(i)} = 1$  denotes the presence of an unreported, dark vessel at the pixel  $(h, w)$ . The total number of dark vessels in the region is the summation of the dark vessels  $y^{(i)} = \sum_{h,w} D_{\text{dark};(h,w)}^{(i)}$ . The procedure described in the previous case is again used to generate a counterfactual outcome  $\bar{y}^{(i)}$  with a randomly sampled counterfactual satellite imagery  $\bar{s}^{(i)}$  and the ground truth causal effect  $\tau^{(i)} = \bar{y}^{(i)} - y^{(i)}$  is computed. An example can be found in Figure 5.

### 4.3. Baseline methods and Experiment Setup

We include a domain-specific baseline model for each case: Random Forest (RF) for Case 1 due to its effectiveness in modeling mosquito populations (Mudele et al., 2020), and Boosted Regression Trees (BRT) for Case 2 for its capability in modeling fishing activities (Yang et al., 2022). When training and estimating the CATE, we use the explicit features from satellite imagery identical to those used in data generation in Sections 4.1 and 4.2 (e.g., NDVI, SST). This simulates domain experts performing careful feature engineering, giving them an advantage with explicit features in the ground-truth process. This comparison assesses SP, where baselines use privileged knowledge and SP relies on learned latent representations to support causal inference.

We further include the following baseline models inline with causal inference research: **CNN** encodes satellite images using a CNN and concatenates them with covariate features for prediction, serving as a sanity check for causal effect estimation without adjustments; **GraphITE** (Harada and Kashima, 2021) extracts treatment representations via graph neural networks (GNNs) with a Hilbert-Schmidt Independence Criterion-based (HSIC) (Gretton et al., 2005, 2007) regularization to mitigate observational biases. Here we replace the original GNN encoder with a CNN to handle satellite imagery, ensuring compatibility with our data structure; and **NICE** (Thorat et al., 2024) is an approach similar to **GraphITE** but employs VGG (Vedaldi and Zisserman, 2016) and ResNet (He et al., 2016) to extract

representations from images, using Maximum Mean Discrepancy(MMD) regularization instead of HSIC.

We follow a conventional data partition: 70% for training, 10% for validation, and 20% for testing. Model performance is compared using Precision in Estimating Heterogeneous Effects (PEHE, Hill, 2011), defined as  $\frac{1}{n} \sum_i (\tau^{(i)} - \hat{\tau}^{(i)})^2$ , evaluated on the test dataset<sup>1</sup>.

### 4.4. Results

In Tables 1 and 2, we report the mean and standard deviation of PEHE across all seeds for the best hyperparameter configuration. SP-CNN refers to SP with a CNN encoder, while SP-Satlas refers to SP with the pretrained Satlas encoder. SP outperforms the baseline models.

To assess robustness when the unconfoundedness assumption is violated, a sensitivity test is conducted by gradually removing covariates from the training data. The results in Figure 6 show that the performance of SP declines when the number of missing covariates increases. Domain-specific models start outperforming SP when the number of missing covariates reaches 2 in Case 1 and 5 in Case 2.

**Table 1. Evaluation results for Case 1.**

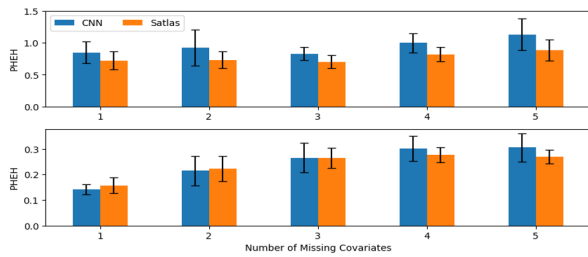
Models	Latent dimension $d$	
	16	32
RF	0.96 (0.09)	
CNN	2.88 (0.81)	2.88 (0.81)
GraphITE	2.88 (0.78)	2.88 (0.77)
NICE-VGG	2.88 (0.77)	2.88 (0.77)
NICE-ResNet	2.88 (0.77)	2.85 (0.81)
SP-CNN	0.93 (0.36)	0.84 (0.35)
SP-Satlas	<b>0.73 (0.30)</b>	0.75 (0.33)

**Table 2. Evaluation results for Case 2.**

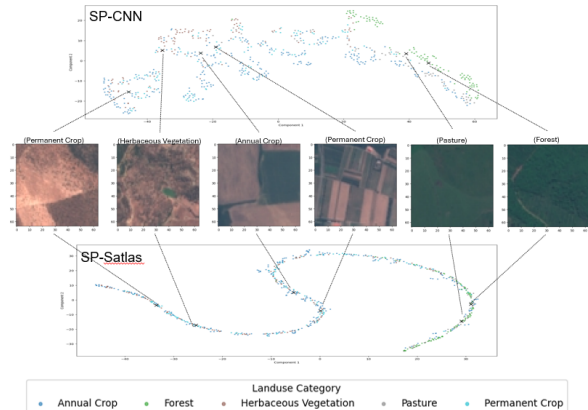
Models	Latent dimension $d$	
	16	32
BRT	0.33 (0.09)	
CNN	1.18 (0.10)	1.18 (0.10)
GraphITE	1.16 (0.12)	1.16 (0.12)
NICE-VGG	1.19 (0.10)	1.19 (0.10)
NICE-ResNet	1.20 (0.11)	1.20 (0.10)
SP-CNN	<b>0.12 (0.05)</b>	0.19 (0.04)
SP-Satlas	0.20 (0.04)	0.22 (0.05)

We further examine the embedding vectors generated by the representation learning model  $\hat{\phi}(\cdot)$

<sup>1</sup>More detailed information on the data synthesis and experiment settings can be found in the online repository <https://github.com/chienluresearch-gif/SP>.



**Figure 6. Results for robust test. Upper: Case 1, Lower: Case 2, showing the mean PEHE and standard deviation bar across all run seeds.**



**Figure 7. Embedding vectors from  $\phi(\cdot)$  in Case 1. Locations from left to right in the embedding space reflect a clear increasing vegetation pattern.**

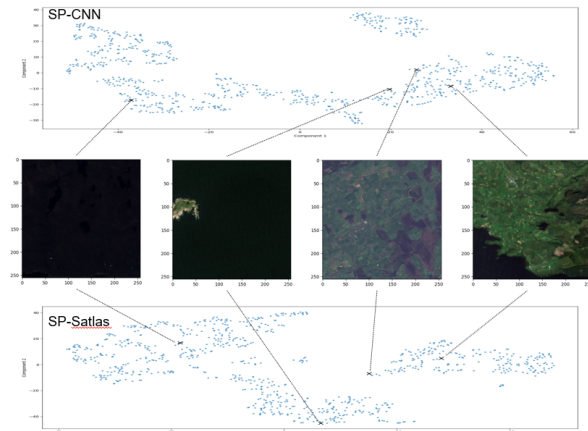
for a qualitative assessment. The visualization via t-SNE is presented in Figures 7 and 8<sup>2</sup>. The relative locations of these vectors clearly reflect environmental patterns. In Case 1, relative locations reveal changes in vegetation, while in Case 2, the relative locations reflect the proportion of land and ocean.

## 5. Real-world Data Analysis: The Forest-Agriculture Trade-off

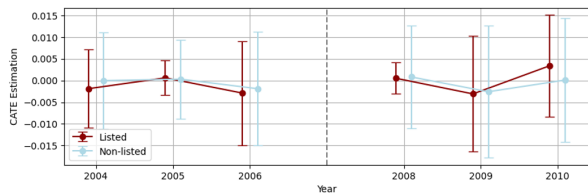
Deforestation is often justified by farmland expansion and increased agricultural productivity. Koch et al. (2019) leveraged the 2008 “Priority List Policy” in the Brazilian Amazon, using synthetic control to construct counterfactual outcomes for listed municipalities from non-listed ones as a donor pool. The results suggest that forest conservation does not reduce crop productivity.

To demonstrate SP’s applicability, we re-analyze the same data using SP model with satellite imagery for 352 municipalities. Unlike synthetic control, which relies

<sup>2</sup>Here we present the training results from one representative run for each case, with other runs showing a similar pattern.



**Figure 8. Embedding vectors from  $\phi(\cdot)$  in Case 2. Locations from left to right in the embedding space reflect the complexity of coastal characteristics.**



**Figure 9. CATE values for listed and non-listed municipalities in the Brazilian Amazon. Dots show the mean; bars show  $\pm 1$  standard deviation.**

on a donor pool, SP requires only observational data. We use the data of 2007 as observations, including 17 economic and socio-political variables as confounders (see Appendix) and Landsat imagery as treatments. The data are split (80% training, 20% validation) for model selection, after which SP is retrained on the full dataset.

Using counterfactual images from years before and after 2007, we estimate the CATE of each municipality to assess the impact of landscape changes on crop production, holding 2007 economic and socio-political covariates constant. The CATE values in Figure 9 are centered around zero for both listed and non-listed municipalities, indicating that landscape changes have no significant impact. This finding is consistent with Koch et al. (2019), who, using a different analytical approach, suggest that broader socio-economic factors are the main drivers.

## 6. Discussions and Conclusions

We introduce SP, a method that integrates satellite imagery into causal inference by treating images as the cause/treatment. Its modular two-step training process allows adaptation to diverse environmental

and spatial contexts. Experiments show that SP outperforms baselines in estimation accuracy and produces embeddings that capture meaningful patterns, offering insights into underlying causal mechanisms. Notably, SP-CNN outperforms SP-Satlas in Case 2, suggesting that more complex structured models do not necessarily improve performance, highlighting the importance of encoder choice for imagery. The real-world data analysis further corroborates prior findings on deforestation and agricultural production, with another data analysis perspective.

One limitation of SP is its reliance on the unconfoundedness assumption; we conducted a sensitivity analysis to assess potential effects of unobserved variables. Addressing unobserved confounding remains an important future direction. Overall, SP is expected to advance GeoAI applications in environmental monitoring and resource management.

## Acknowledgements

This work has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Grant Agreement No. 101002240) and Science Foundation Ireland Adapt Research Centre under Grant No. 13/RC/2106\_P2.

## References

- Agnew, D. J., Pearce, J., Pramod, G., Peatman, T., Watson, R., Beddington, J. R., & Pitcher, T. J. (2009). Estimating the worldwide extent of illegal fishing. *PloS one*, 4(2), e4570.
- Bastani, F., Wolters, P., Gupta, R., Ferdinando, J., & Kembhavi, A. (2023). Satlaspretrain: A large-scale dataset for remote sensing image understanding. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 16772–16782.
- Becker, N. (2008). Influence of climate change on mosquito development and mosquito-borne diseases in europe. *Parasitology Research*, 103(Suppl 1), 19–28.
- Brugueras, S., Fernández-Martínez, B., Martínez-de la Puente, J., Figuerola, J., Porro, T. M., Rius, C., Larrauri, A., & Gómez-Barroso, D. (2020). Environmental drivers, climate change and emergent diseases transmitted by mosquitoes and their vectors in southern europe: A systematic review. *Environmental research*, 191, 110038.
- Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., & Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters.
- Conlin, C. (2024). Using machine learning and daytime satellite imagery to estimate aid's effect on wealth: Comparing china and world bank programs in africa.
- Darwish, D. (2025). Geospatial ai concepts and fundamentals. In *Recent trends in geospatial ai* (pp. 1–26). IGI Global Scientific Publishing.
- de Luna, X., & Johansson, P. (2006). Exogeneity in structural equation models. *Journal of Econometrics*, 132(2), 527–543.
- Fick, S. E., Nauman, T. W., Brungard, C. C., & Duniway, M. C. (2021). Evaluating natural experiments in ecology: Using synthetic controls in assessments of remotely sensed land treatments. *Ecological Applications*, 31(3), e02264.
- Ford, T. E., Colwell, R. R., Rose, J. B., Morse, S. S., Rogers, D. J., & Yates, T. L. (2009). Using satellite images of environmental changes to predict infectious disease outbreaks. *Emerging infectious diseases*, 15(9), 1341.
- Gordon, M., Ayers, M., Stone, E., & Sanford, L. (2023). Remote control: Debiasing remote sensing predictions for causal inference. *Proc. Int. Conf. Learn. Represent. Workshops*, 22.
- Gretton, A., Bousquet, O., Smola, A., & Schölkopf, B. (2005). Measuring statistical dependence with hilbert-schmidt norms. *International conference on algorithmic learning theory*, 63–77.
- Gretton, A., Fukumizu, K., Teo, C., Song, L., Schölkopf, B., & Smola, A. (2007). A kernel statistical test of independence. *Advances in neural information processing systems*, 20.
- Harada, S., & Kashima, H. (2021). Graphite: Estimating individual effects of graph-structured treatments. *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, 659–668.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.
- Heckman, J. J., Ichimura, H., & Todd, P. E. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *The review of economic studies*, 64(4), 605–654.

- Helber, P., Bischke, B., Dengel, A., & Borth, D. (2019). Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(7), 2217–2226.
- Hill, J. L. (2011). Bayesian nonparametric modeling for causal inference. *Journal of Computational and Graphical Statistics*, 20(1), 217–240.
- Imbens, G. W., & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of economic literature*, 47(1), 5–86.
- Jerzak, C. T., Johansson, F., & Daoud, A. (2022). Estimating causal effects under image confounding bias with an application to poverty in africa. *arXiv preprint arXiv:2206.06410*.
- Jerzak, C. T., Johansson, F., & Daoud, A. (2023). Integrating earth observation data into causal inference: Challenges and opportunities. *arXiv preprint arXiv:2301.12985*.
- Kaddour, J., Zhu, Y., Liu, Q., Kusner, M. J., & Silva, R. (2021). Causal effect inference for structured treatments. *Advances in Neural Information Processing Systems*, 34, 24841–24854.
- Koch, N., zu Ermgassen, E. K., Wehkamp, J., Oliveira Filho, F. J., & Schwerhoff, G. (2019). Agricultural productivity and forest conservation: Evidence from the brazilian amazon. *American Journal of Agricultural Economics*, 101(3), 919–940.
- Künzel, S. R., Sekhon, J. S., Bickel, P. J., & Yu, B. (2019). Metalearners for estimating heterogeneous treatment effects using machine learning. *Proceedings of the national academy of sciences*, 116(10), 4156–4165.
- Le Gallic, B., & Cox, A. (2006). An economic analysis of illegal, unreported and unregulated (iuu) fishing: Key drivers and possible solutions. *Marine policy*, 30(6), 689–695.
- Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., & Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. *Proceedings of the IEEE/CVF international conference on computer vision*, 10012–10022.
- Lotka, A. J. (1925). *Elements of physical biology*. Williams & Wilkins.
- Mudele, O., Bayer, F. M., Zanandrez, L. F., Eiras, A. E., & Gamba, P. (2020). Modeling the temporal population distribution of *Ae. aegypti* mosquito using big earth observation data. *Ieee Access*, 8, 14182–14194.
- Nie, X., & Wager, S. (2021). Quasi-oracle estimation of heterogeneous treatment effects. *Biometrika*, 108(2), 299–319.
- Nogueira, A. R., Pugnana, A., Ruggieri, S., Pedreschi, D., & Gama, J. (2022). Methods and tools for causal discovery and causal inference. *Wiley interdisciplinary reviews: data mining and knowledge discovery*, 12(2), e1449.
- O’Sullivan, C., Kashyap, A., Coveney, S., Monteys, X., & Dev, S. (2024). Enhancing coastal water body segmentation with landsat irish coastal segmentation (lics) dataset. *Remote Sensing Applications: Society and Environment*, 101276.
- Park, J., Lee, J., Seto, K., Hochberg, T., Wong, B. A., Miller, N. A., Takasaki, K., Kubota, H., Oozeki, Y., Doshi, S., et al. (2020). Illuminating dark fishing fleets in north korea. *Science advances*, 6(30), eabb1197.
- Pearl, J. (2009). *Causality*. Cambridge university press.
- Robinson, P. M. (1988). Root-n-consistent semiparametric regression. *Econometrica: Journal of the Econometric Society*, 931–954.
- Schölkopf, B. (2022). Causality for machine learning. In *Probabilistic and causal inference: The works of judea pearl* (pp. 765–804).
- Thorat, A., Kolla, R., & Pedanekar, N. (2024). I see, therefore i do: Estimating causal effects for image treatments. *arXiv preprint arXiv:2412.06810*.
- Vedaldi, A., & Zisserman, A. (2016). Vgg convolutional neural networks practical. *Department of Engineering Science, University of Oxford*, 66.
- Witmer, F. D. (2015). Remote sensing of violent conflict: Eyes from above. *International Journal of Remote Sensing*, 36(9), 2326–2352.
- Wood-Doughty, Z., Shpitser, I., & Dredze, M. (2021). Generating synthetic text data to evaluate causal inference methods. *arXiv preprint arXiv:2102.05638*.
- Yang, S., Zhang, H., Fan, W., Shi, H., Fei, Y., & Yuan, S. (2022). Behaviour impact analysis of tuna purse seiners in the western and central pacific based on the brt and gam models. *Frontiers in Marine Science*, 9, 881036.