Structured Pixels: Satellite Imagery as the Cause in Causal Effect Estimation (Appendix)

I. More on semi-synthesized datasets

A. *EuroSAT*

We calculate those indices using *EuroSAT* dataset:

1. Normalized Difference Vegetation Index (NDVI) The NDVI is used to assess vegetation greenness and health.

$$\text{NDVI} = \frac{NIR - Red}{NIR + Red}$$

2. Normalized Difference Moisture Index (NDMI) The NDMI helps estimate vegetation water content.

$$\text{NDMI} = \frac{NIR - SWIR_1}{NIR + SWIR_1}$$

3. Normalized Difference Red Edge Index (NDRE) NDRE is used to assess chlorophyll content and early plant stress.

$$\text{NDRE} = \frac{NIR - RedEdge}{NIR + RedEdge}$$

4. Moisture Stress Index (MSI) Higher MSI values indicate greater moisture stress.

$$\text{MSI} = \frac{SWIR_1}{NIR}$$

- NIR (band 8): Near-Infrared reflectance
- Red (band 4): Red reflectance
- $SWIR_1$ (band 11): Shortwave Infrared 1 reflectance
- $RedEdge$ (band 8A): Red-edge reflectance

Used bands in image encoders. We take the bands 2, 3, 4, 5, 6, 7, 8, 11, 12, in total 9 bands, in order to stay compatible with the Satlas pretrained model.

Covarates. We simulated the category-dependent covariates as follows. Each factor is generated using a Beta distribution with shape parameters defining the distribution.

PermanentCrop:

- Medium livestock density: $\beta(6, 4)$
- High human activity: $\beta(9, 2)$
- High temperature resilience (irrigation/management): $\beta(7, 3)$
- Low seasonal variability (managed environment): $\beta(7, 3)$
- Medium waste accumulation: $\beta(5, 5)$
- Low predator presence (more mosquitoes): $\beta(2, 8)$
- High artificial light at night (ALAN): $\beta(8, 2)$

AnnualCrop:

- High livestock density: $\beta(7, 3)$
- High human activity: $\beta(8, 2)$
- Medium-high temperature resilience: $\beta(4, 6)$
- High seasonal variability (managed environment): $\beta(9, 1)$
- Medium-high waste accumulation: $\beta(6, 4)$
- Low-medium predator presence: $\beta(3, 7)$
- Medium-high artificial light at night (ALAN): $\beta(7, 3)$

Pasture:

- Very high livestock density: $\beta(9, 1)$
- Medium human activity: $\beta(8, 2)$
- Medium temperature resilience: $\beta(4, 6)$
- Medium seasonal variability: $\beta(9, 1)$
- Low-medium waste accumulation: $\beta(4, 6)$
- Medium-high predator presence (fewer mosquitoes): $\beta(6, 4)$
- Low artificial light at night (ALAN): $\beta(3, 7)$

HerbaceousVegetation:

- Medium-low livestock density: $\beta(4, 6)$
- Low human activity: $\beta(3, 7)$
- Low-medium temperature resilience: $\beta(4, 6)$

- Medium-high variability: $\beta(7,3)$
- Low waste accumulation: $\beta(3,7)$
- High predator presence (fewer mosquitoes): $\beta(7,3)$
- Very low artificial light at night (ALAN): $\beta(2,8)$

Forest:

- Very low livestock density: $\beta(1,9)$
- Very low human activity: $\beta(1,9)$
- High temperature resilience: $\beta(8,2)$
- Low seasonal variability: $\beta(2,8)$
- Very low waste accumulation: $\beta(2,8)$
- Very high predator presence (fewest mosquitoes): $\beta(9,1)$
- Very low artificial light at night (ALAN): $\beta(1,9)$

B. LICS

The generation of fish population is inspired by the the form the logistic differential equation:

$$\frac{dP}{dt} = rP\left(1 - \frac{P}{K}\right)$$

with $K = 1000$, we consider a parameterized form inspired by the given $P(t)$. Adapting it to a new context (e.g., a feature map or probability model), we define:

$$F^{(i)}_{(h,w)} = \frac{1000}{1 + \exp(-5 \times (\lambda^{(i)}_{(h,w)} - 0.5))}$$

where:

- $F^{(i)}_{(h,w)}$ represents the quantity at iteration i and position (h, w),
- $\lambda^{(i)}_{(h,w)}$ is a parameter (e.g., a probability or intensity) influencing the growth,
- 1000 is the carrying capacity or maximum value.

This form resembles a sigmoid function, often used in modeling bounded growth or probabilities, tailored to the specific indices and parameters provided.

Consider a binary water mask denoted W, where $W(x, y) = 1$ represents water regions and $W(x, y) = 0$ represents land regions. The following steps describe the computation of coastal influence and distance effects:

1) **Coastline Extraction**: The coastline C is derived by applying an exclusive-or operation (XOR) between the water mask W and its binary erosion:

$$C = W \oplus \text{ndimage.binary\_erosion}(W)$$

Here, ndimage.binary_erosion(W) shrinks the water regions by one pixel, and the XOR operation identifies the boundary pixels where W and its erosion differ.

2) **Coastal Influence**: The coastal influence I is computed by applying a Gaussian filter to the coastline C, interpreted as a floating-point array:

$$I = \text{ndimage.gaussian\_filter}(C_{\text{float}}, \sigma = 2.0)$$

where C_{float} is the type-cast version of C to floating-point values, and $\sigma = 2.0$ defines the standard deviation of the Gaussian kernel, controlling the spatial spread.

3) **Normalization of Coastal Influence**: Let $I_{\text{max}} = \max(I)$ be the maximum value of the coastal influence. The normalized coastal influence I_{norm} is then calculated as:

$$I_{\text{norm}} = \begin{cases} \text{clip}\left(\frac{I}{I_{\text{max}}}, 0, 1\right) & \text{if } I_{\text{max}} > 0, \\ \text{clip}(I, 0, 1) & \text{otherwise,} \end{cases}$$

where clip$(x, 0, 1)$ constrains the values of x to the interval $[0, 1]$.

4) **Distance Transform**: The Euclidean distance transform D is computed from the complement of the coastline ($\sim C$), representing the distance from each pixel to the nearest coastline pixel:

$$D = \text{ndimage.distance\_transform\_edt}(\sim C)$$

Here, $\sim C$ inverts the binary coastline array ($C = 1$ becomes 0, and vice versa).

5) **Distance Effect**: The distance effect E is modeled as an exponential decay based on the distance D:

$$E = \exp\left(-\frac{D}{20.0}\right)$$

where the decay constant 20.0 determines the rate at which influence diminishes with distance from the coast.

II. Settings for the Experiments

We list the hyper parameters and model settings for the experiments in the following tables:

TABLE I
TRAINING PARAMETERS

Case 1 (EuroSAT)	
Seeds	64 128 87 8787 89
Epochs	50 100
Learning Rate	0.0001 0.00001
Batch Size	64 128
Alternative Training Step J	5 10
Case 2 (LICS)	
Seeds	64 128 87 8787 89
Epochs	50 100
Learning Rate	0.0001 0.00001
Batch Size	16 32
Alternative Training Step J	5 10

TABLE II
MODEL PARAMETERS

General Settings	
Feature Dimension	16 32
Hidden Dimentions	[128 64 32], [64 32]
Drop out rate	0.2, 0.5
SP	
Regularization parameter	0.1
NICE	
Regularization parameter	1, 0.5
GraphITE	
Regularization parameter	10 1000

III. More on the chosen baseline models

The baseline models **CNN** and **NICE** are adaptations of canonical causal inference models **TarNET** and **CFRNet** [Shalit et al., 2017]. Specifically **TarNET** is designed for binary treatments $T \in \{0, 1\}$ and learns shared representations $\Phi(X)$ from covariates X, with separate prediction heads h_0 and h_1 for each treatment group:

$$\hat{Y} = h_T(\Phi(X)). \tag{1}$$

CNN predicts outcomes from concatenated covariate and treatment features $[X, T]$ through a single prediction head h:

$$\hat{Y} = h([\Psi(X), \Phi(T)]), \tag{2}$$

This design allows the model to handle high-dimensional treatments $T \in \mathbb{R}^d$, such as satellite images, where separate heads for each treatment value are infeasible.

CFRNet extends TarNET with Integral Probability Metric (IPM) regularization to balance representations across treatment groups:

$$\mathcal{L} = \frac{1}{n} \sum_{i=1}^{n} \ell\big(h_{T_i}(\Phi(X_i)), Y_i\big) + \lambda \cdot \text{IPM}\big(\{\Phi(X_i) \mid T_i = 0\}, \{\Phi(X_i) \mid T_i = 1\}\big), \tag{3}$$

where ℓ is a supervised loss (e.g., mean squared error) and λ controls the regularization strength.

TABLE III
COMPARISON OF CAUSAL INFERENCE MODELS

Model	Treatment Type	Architecture	Regularization	Key Use Case
TarNET	Binary ($T \in \{0, 1\}$)	Shared $\Phi(X)$, separate heads h_0, h_1	None	Binary treatment causal inference
CNN	High-dimensional ($\mathbb{R}^d$)	Single head $h([\Psi(X), \Phi(T)])$	None	High-dimensional treatments (e.g., images)
CFRNet	Binary ($T \in \{0, 1\}$)	Shared $\Phi(X)$, separate heads	IPM (e.g., Wasserstein, MMD)	Balanced representations for binary treatments
NICE	High-dimensional (images)	CNN-based $\Phi(X, T)$, single head	MMD	Image-based causal inference
Graphite	High-dimensional	CNN-based $\Phi(X)$, single head	HSIC	Independence-focused causal inference

Similarly, **NICE** uses MMD as an IPM regularizer and leverages CNN-based feature extractors (e.g., VGG or ResNet) to learn representations $\Phi(X, T)$ from images:

$$\hat{Y} = h(\Phi(X, T)), \quad \mathcal{L} = \frac{1}{n} \sum_{i=1}^{n} \ell(\hat{Y}_i, Y_i) + \lambda \cdot \text{MMD}(\Phi(X, T)). \tag{4}$$

Graphite is conceptually similar to NICE, but employs the Hilbert–Schmidt Independence Criterion (HSIC) as a regularizer, encouraging independence between learned covariate representations and treatment assignments:

$$\mathcal{L} = \frac{1}{n} \sum_{i=1}^{n} \ell(\hat{Y}_i, Y_i) + \lambda \cdot \text{HSIC}(\Phi(X), T). \tag{5}$$

Summary

1) **Treatment Handling**: TarNET and CFRNet are limited to binary treatments due to separate heads. CNN, NICE, and Graphite handle high-dimensional or continuous treatments, suitable for complex data like images.
2) **Regularization**: TarNET and CNN lack explicit regularization, risking confounding bias. CFRNet uses IPM, NICE uses MMD, and Graphite uses HSIC to mitigate bias.
3) **Feature Extraction**: NICE and Graphite leverage encoders such as VGG or ResNET for image inputs, while TarNET and CFRNet typically assume tabular data. CNN is flexible but does not specify a feature extractor.
4) **Applications**: TarNET and CFRNet are ideal for binary treatment scenarios (e.g., medical trials). CNN, NICE, and Graphite suit modern applications with complex data (e.g., images).

IV. Covariates includede in Section 5

We use the following variables, representing economic and politica status of the region. The data is available at https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1093%2Fajae%2Faay110&file=ajaeajaeaay110-sup-0001.zip

- Quantidade_agri – Quantity of agricultural production
- Valor_Agri – Value of agricultural production
- Quantidad_pecu – Quantity of livestock production
- Valor_pecu – Value of livestock production
- municip_area – Area of the municipality
- distance_av – Distance to main access/road
- pop_density2007 – Population density in 2007
- gdppc_l1 – GDP per capita lagged by 1 period
- soy_price_l1 – Price of soy lagged by 1 period
- timber_price_l1 – Price of timber lagged by 1 period
- perc_owners – Percentage of landowners
- tractors_per_farm – Number of tractors per farm
- upi_parea – Area under UPI (individual production units)
- ind_parea – Industrial area
- settle_parea – Settled area
- nibamafines_no_l1 – Number of IBAMA fines lagged by 1 period
- precipitation – Precipitation (rainfall)

References

[Shalit et al., 2017] Shalit, U., Johansson, F. D., and Sontag, D. (2017). Estimating individual treatment effect: generalization bounds and algorithms. In *International conference on machine learning*, pages 3076–3085. PMLR.