CLAM: Causal Spatial Disaggregation

Disaggregating causal effect for localised inference

Gerrit Grossmann*, Sumantrak Mukherjee* and Sebastian Vollmer

From Coarse to Fine Grained

Statistical downscaling / super resolution

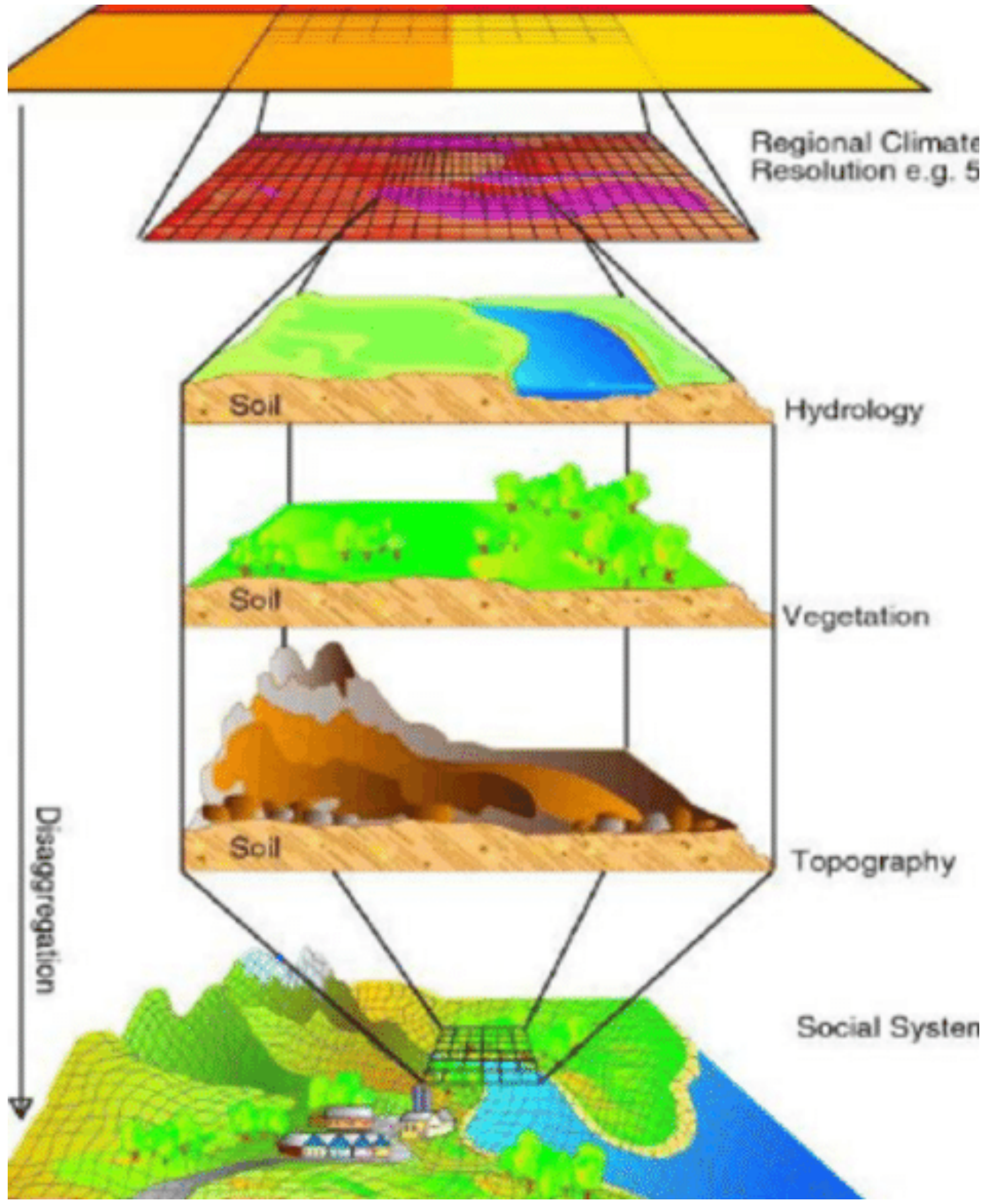
Given state voting outcomes, can we estimate results at the district level?

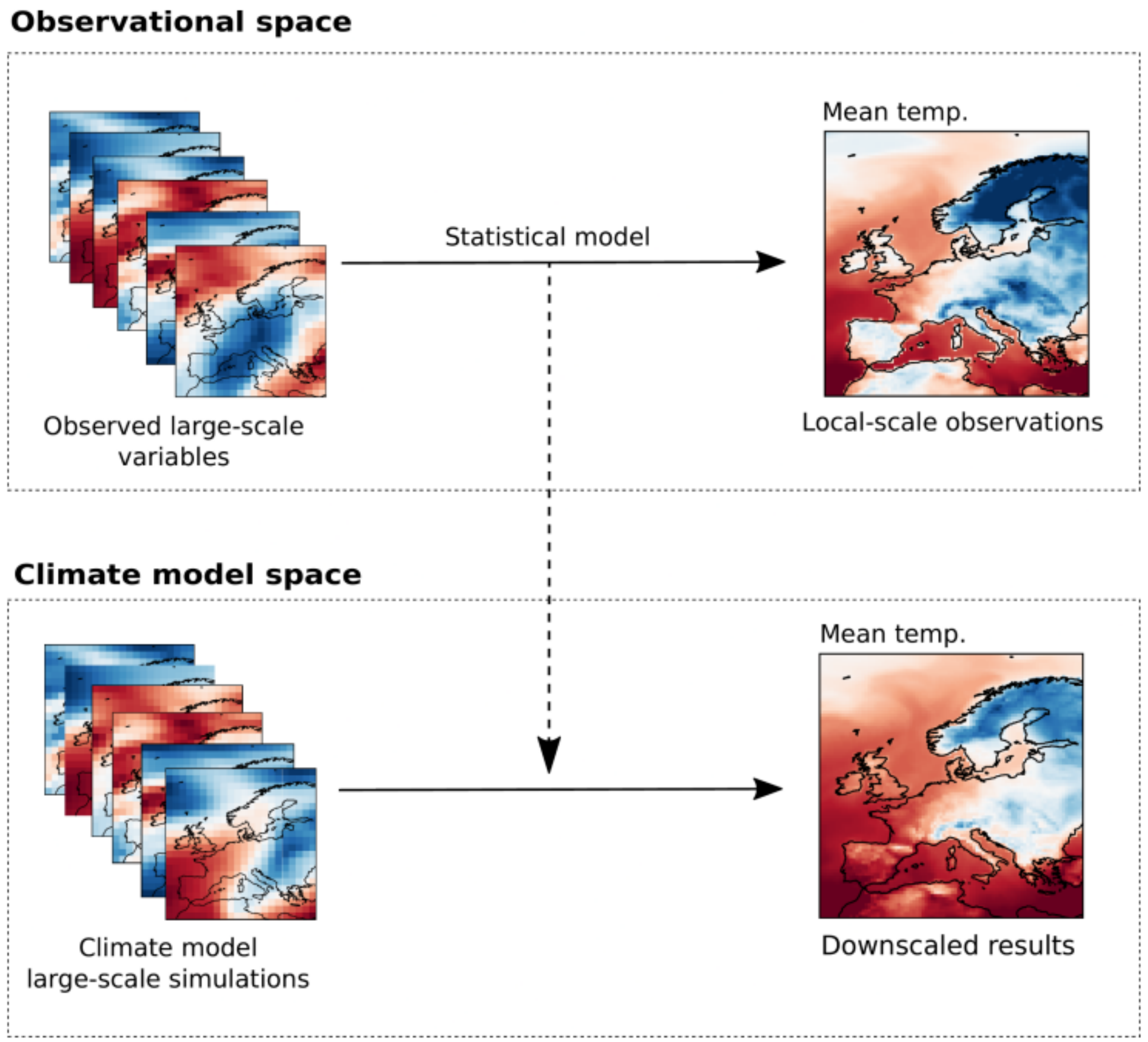
Can daily satellite temperature data estimate hourly surface temperature for specific location?

Given state level education scores, can we estimate school level performance distributions?

Statistical Downscaling

Refining low resolution observations using high resolution information





Regression between coarse predictors and fine scale observations

Matching based methods, using nearest estimate from historical data

Weather Generators and statistical change predictors

Statistical to Causal Downscaling

Learning local causal mechanism from aggregated data

Given regional voting outcomes, can we estimate results at sub-regional level?

How does regional level air quality policies affect street level pollution?

Can daily satellite temperature data estimate hourly surface temperature for specific location?

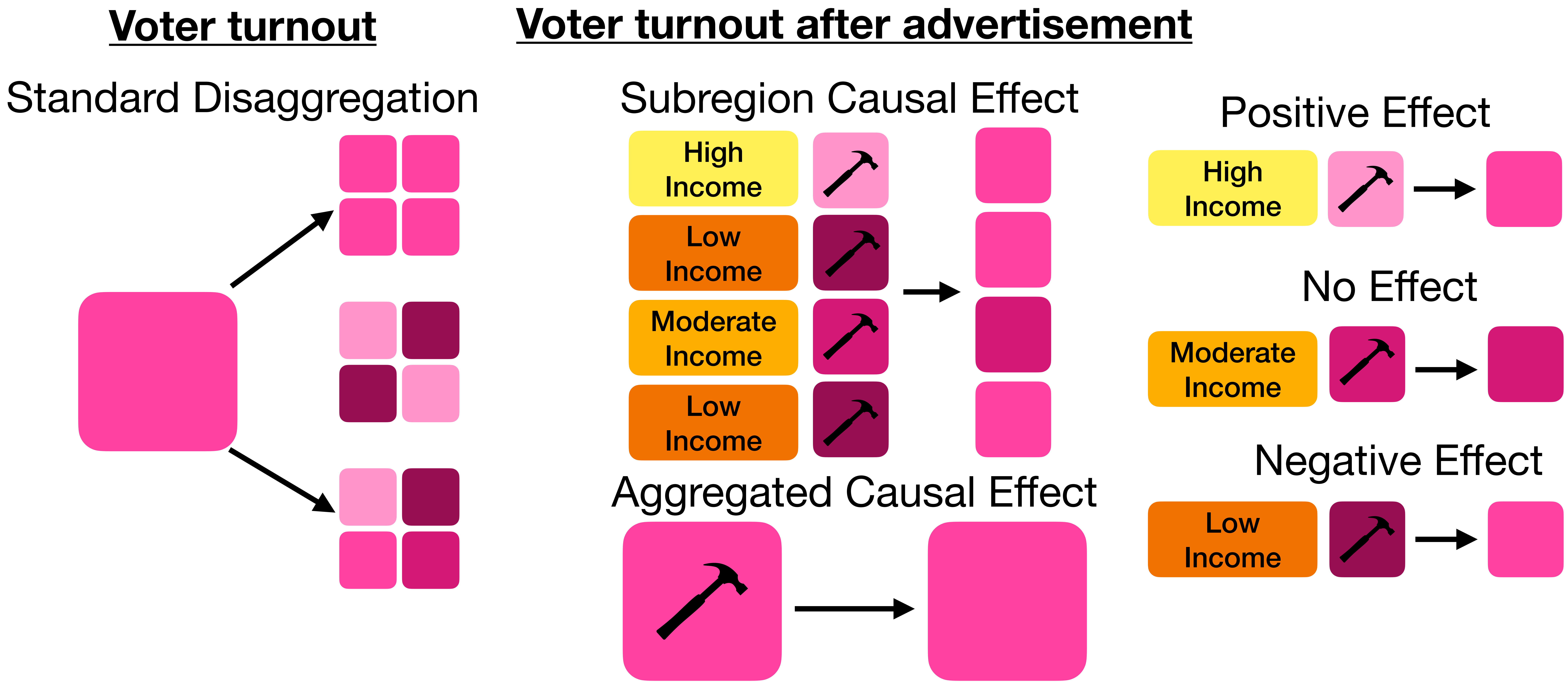
How do national vaccination campaigns affect local infection rates across villages?

Given state level education scores, can we estimate school level performance distributions?

How do district-level education reforms translate into classroom-level learning outcomes?

Causal Disaggregation

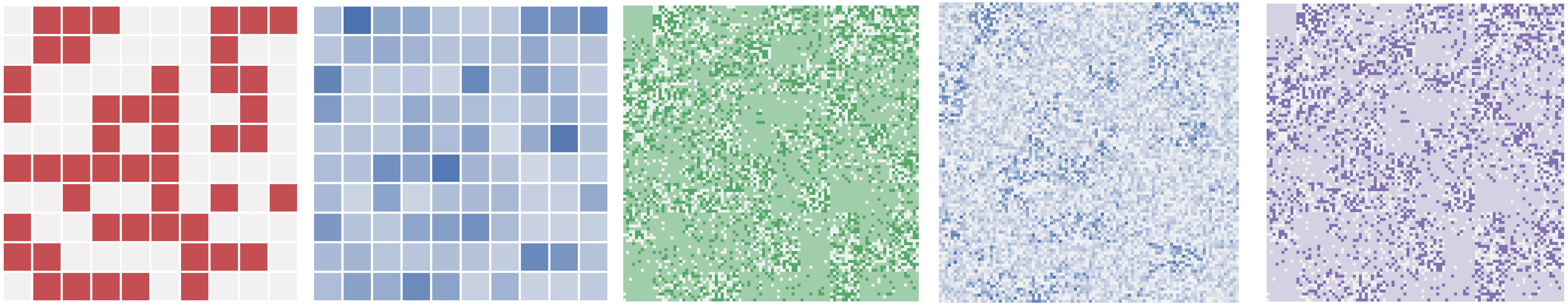
Learning the true underlying causal effects



Sub-regional outcomes from regional policy

How does political campaigning affect politician performance?

Intervention Locations | Regional Outcomes | High-Resolution Context | Sub-regional output | Causal effect output



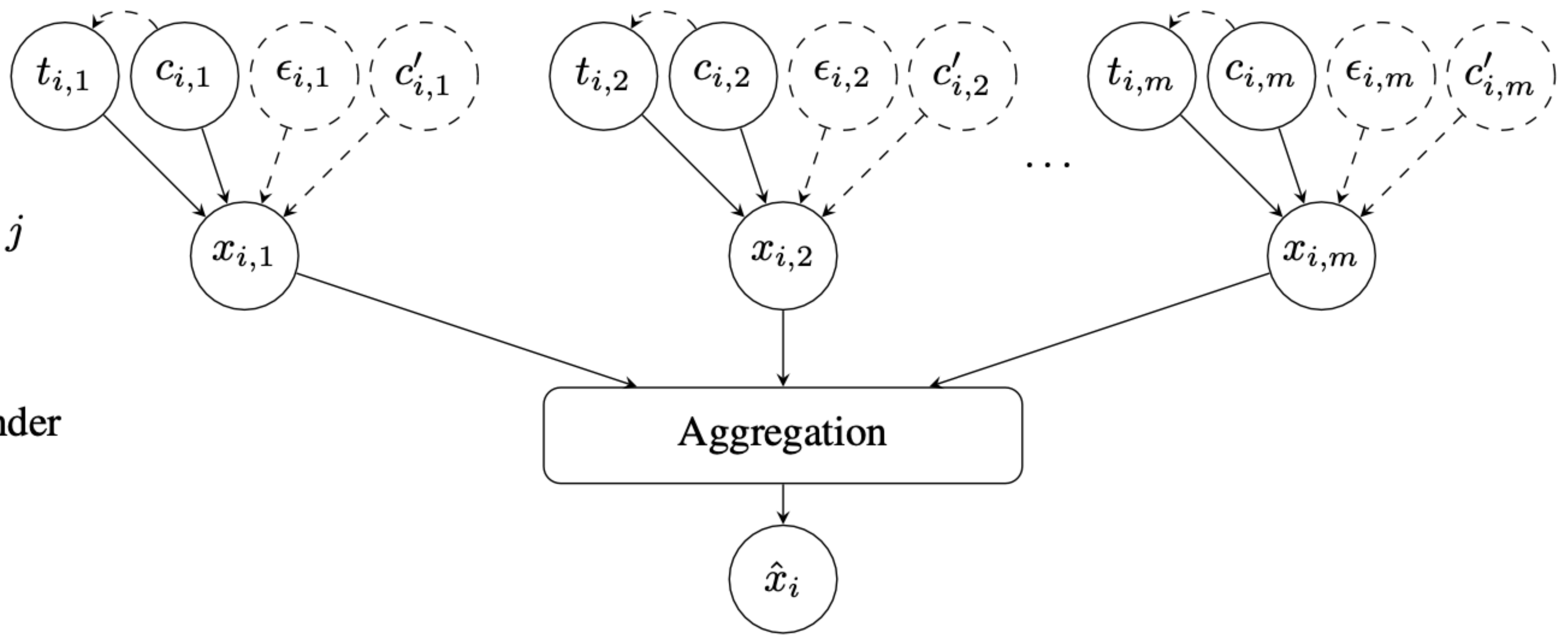
We intervene (political campaigning) in some regions

The subregions of the intervened regions have different effects

We observe (political performance) in all regions

Causal Graph for a Region

Index (i, j) denotes region i, sub-region j
$t_{i,j} \in \{0, 1\}$: intervention indicator
$c_{i,j} \in \mathbb{R}$: context vector
$\epsilon_{i,j} \in \mathbb{R}$: noise
$c'_{i,j} \in \mathbb{R}$: unobserved contextual confounder
$x_{i,j} \in \mathbb{R}$: sub-region outcome
$\hat{x}_i \in \mathbb{R}$: aggregated regional outcome



All functional relationships are shared among the subregions

The outcome of the treatment depends on subregional context

Aggregation can be something like mean, max or sum

Motivating Example

Drought and food prices

Different kinds of subregions	☀		Population	Aggregated Price Increase
Irrigated Farmland	2%	×	5M	
Rain-Fed Agriculture	30%	×	3M	7.3%
Coastal Region	8%	×	2M	
Urban Region	3%	×	10M	

In this example, we look at the effect of drought on food prices in different subregions. Some regions are affected severely while other regions relatively less.

Examples of Intervention and Covariates

Policies are often implemented on an aggregated level

Education Spending

Emission Policies

Public Transit Spending

Vaccination Drives

Healthcare spending

Disaster Relief

Household Survey Data

Income and Wealth

Demographic Data

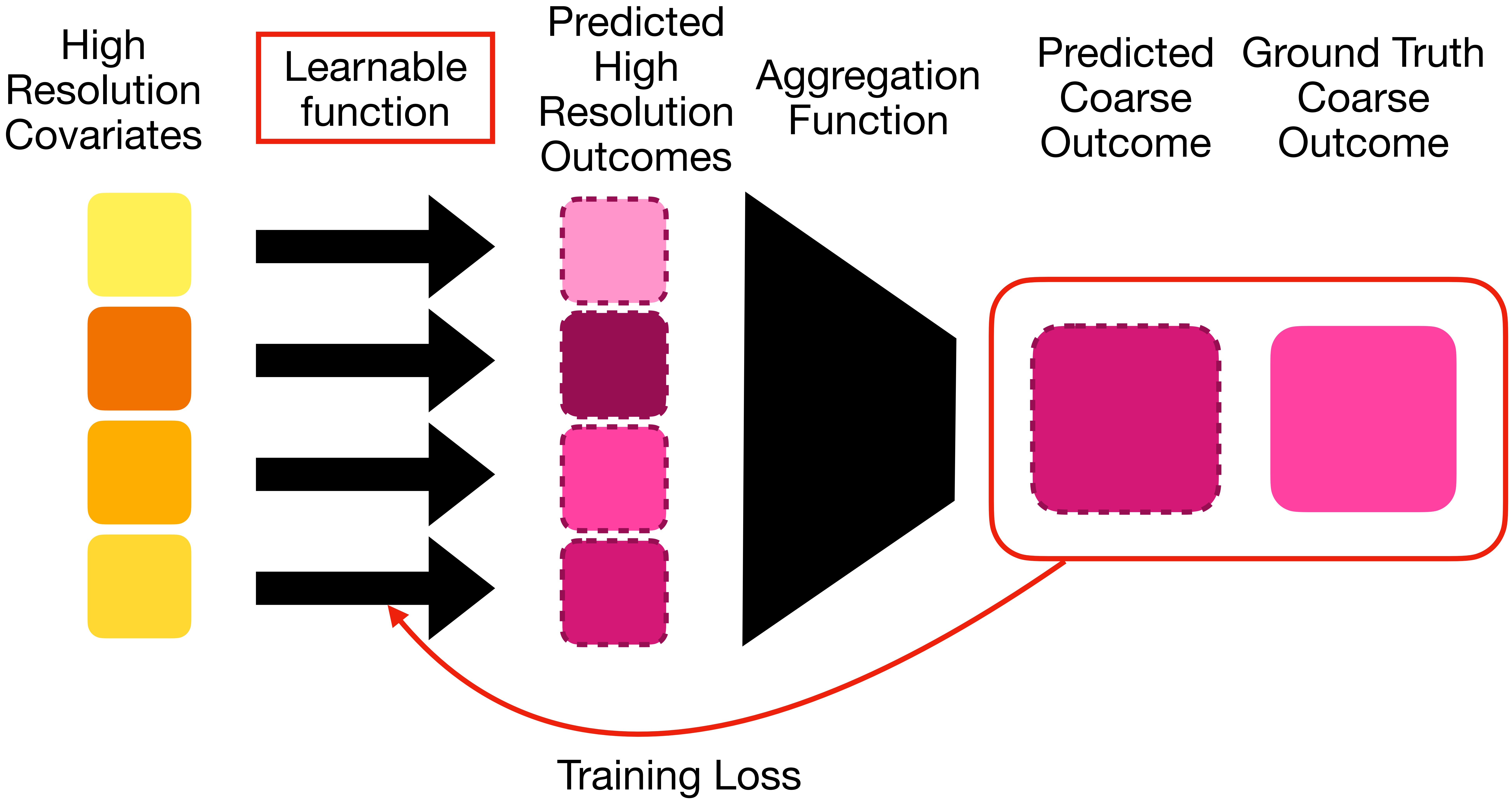
Terrain and Landscape

Internet Surveys

Satellite Images

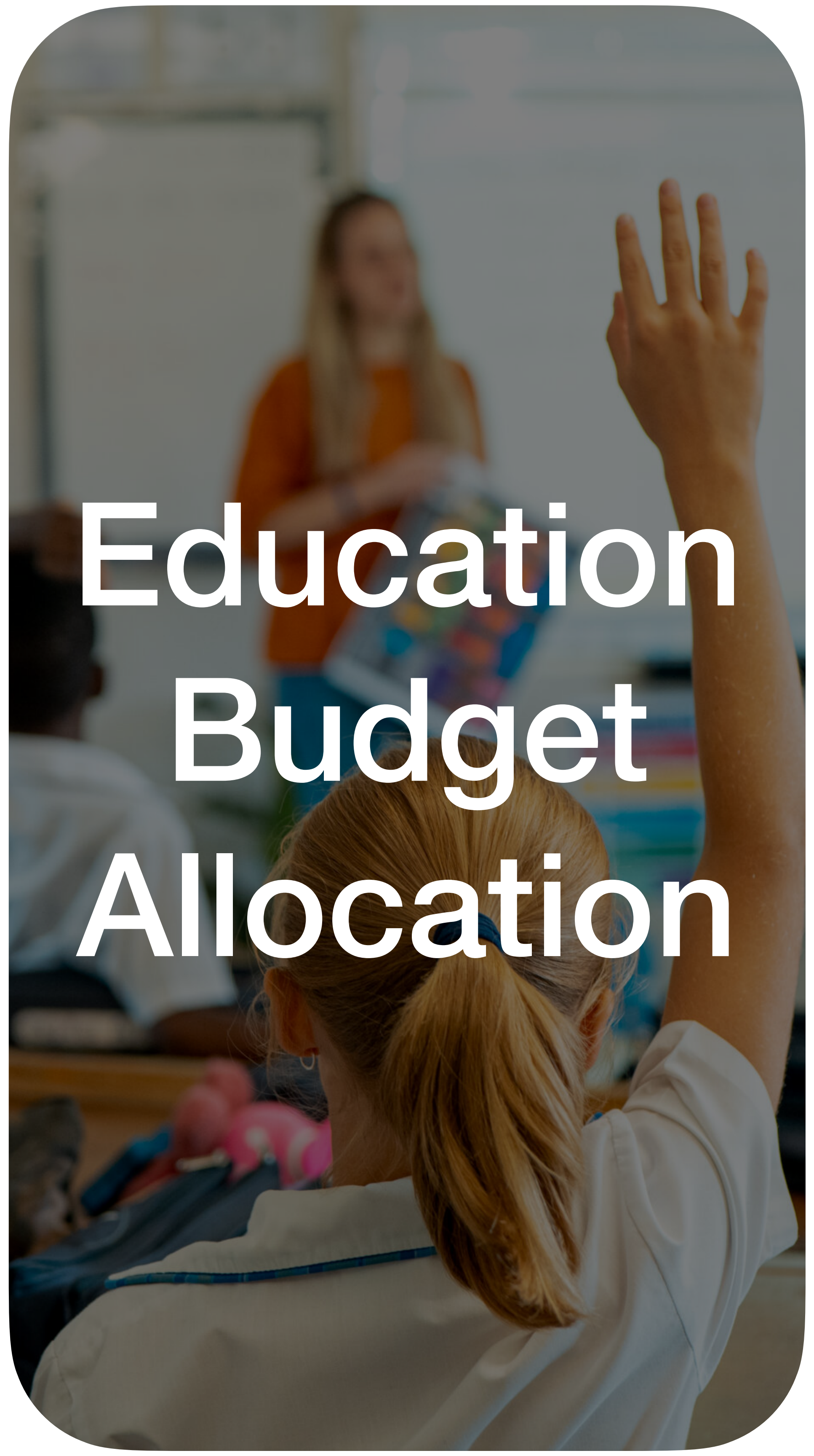
The Idea Behind Our Method

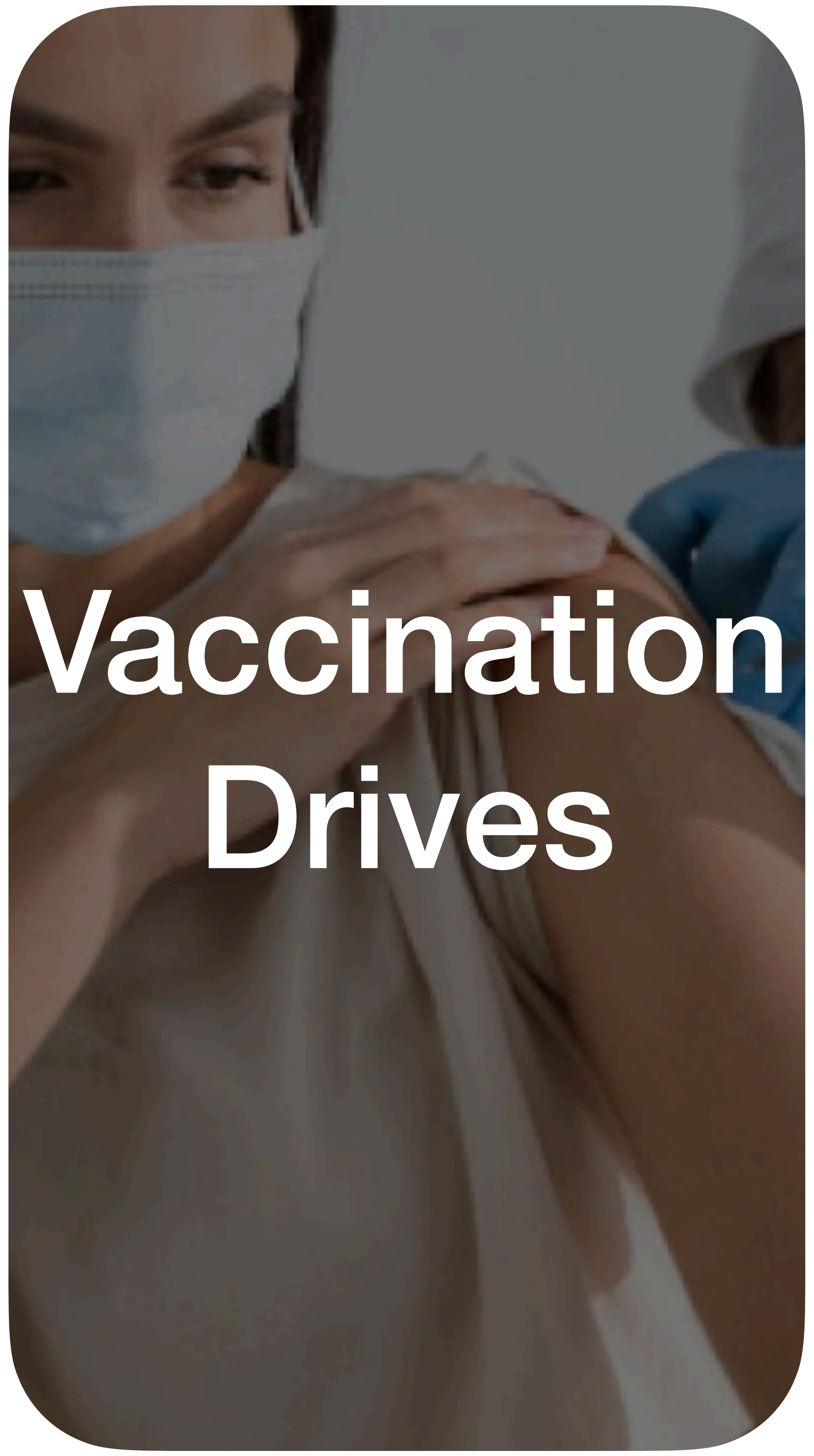
Predict high resolution and then aggregate

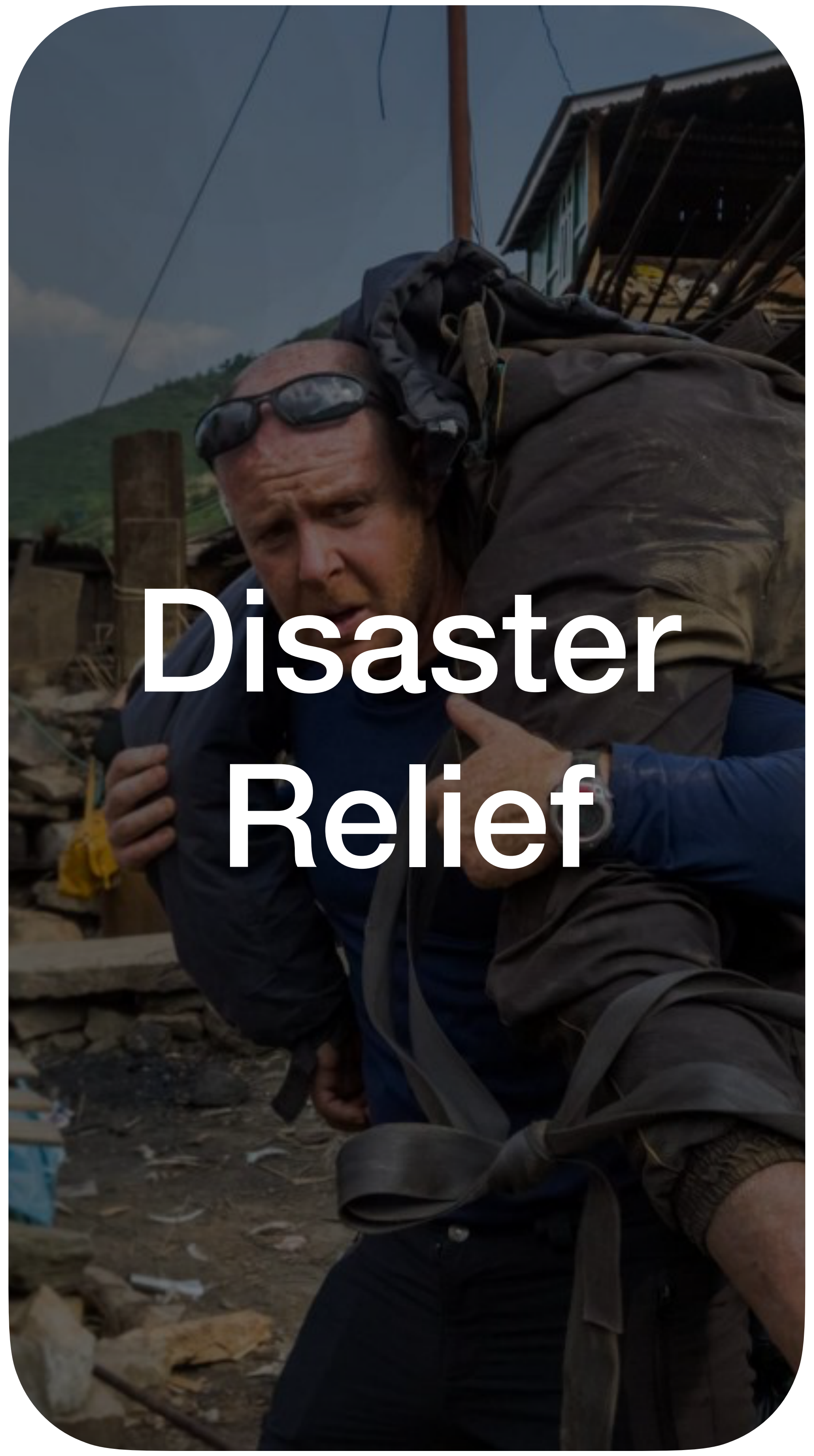


Applications of Causal Deabstraction

Allocating resources for the greatest impact







Political Campaigning

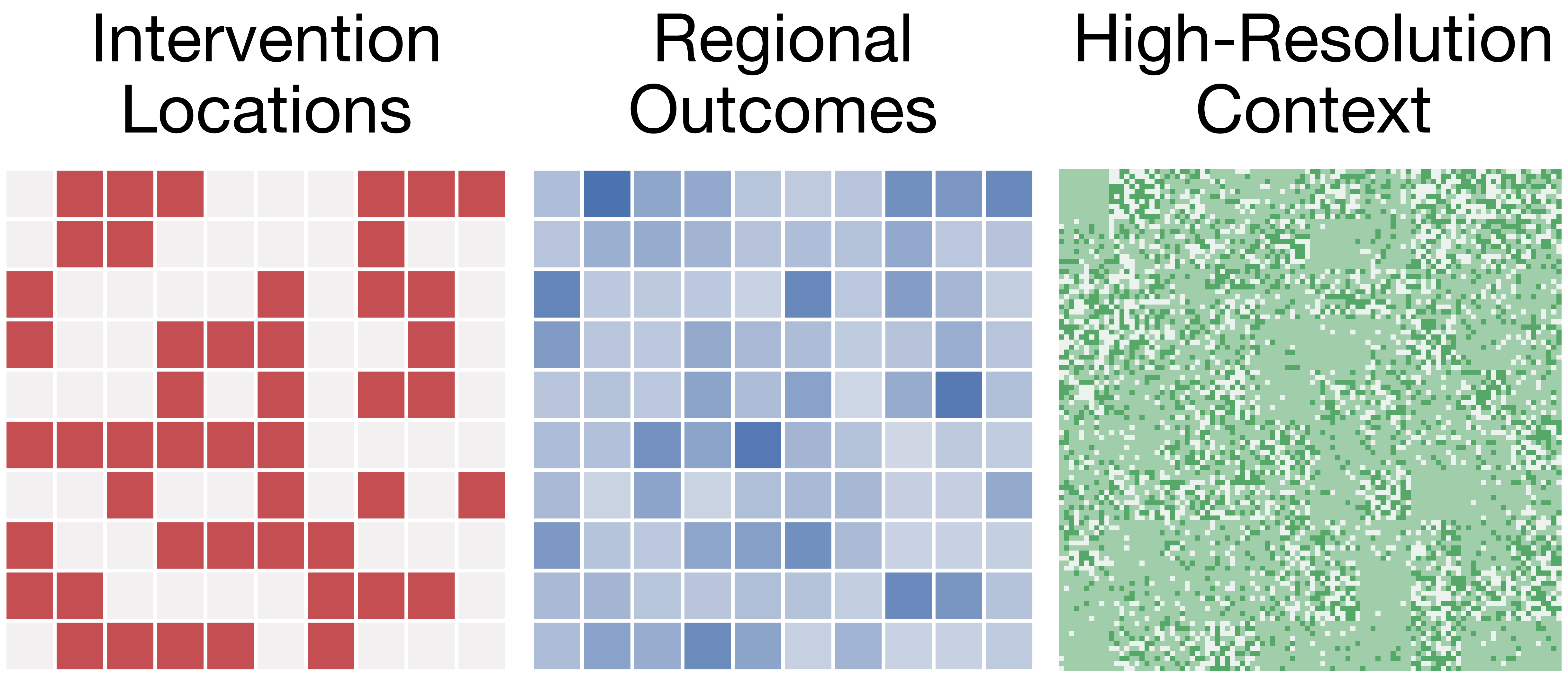
Experiment 1 : How does campaigning affect politician performance

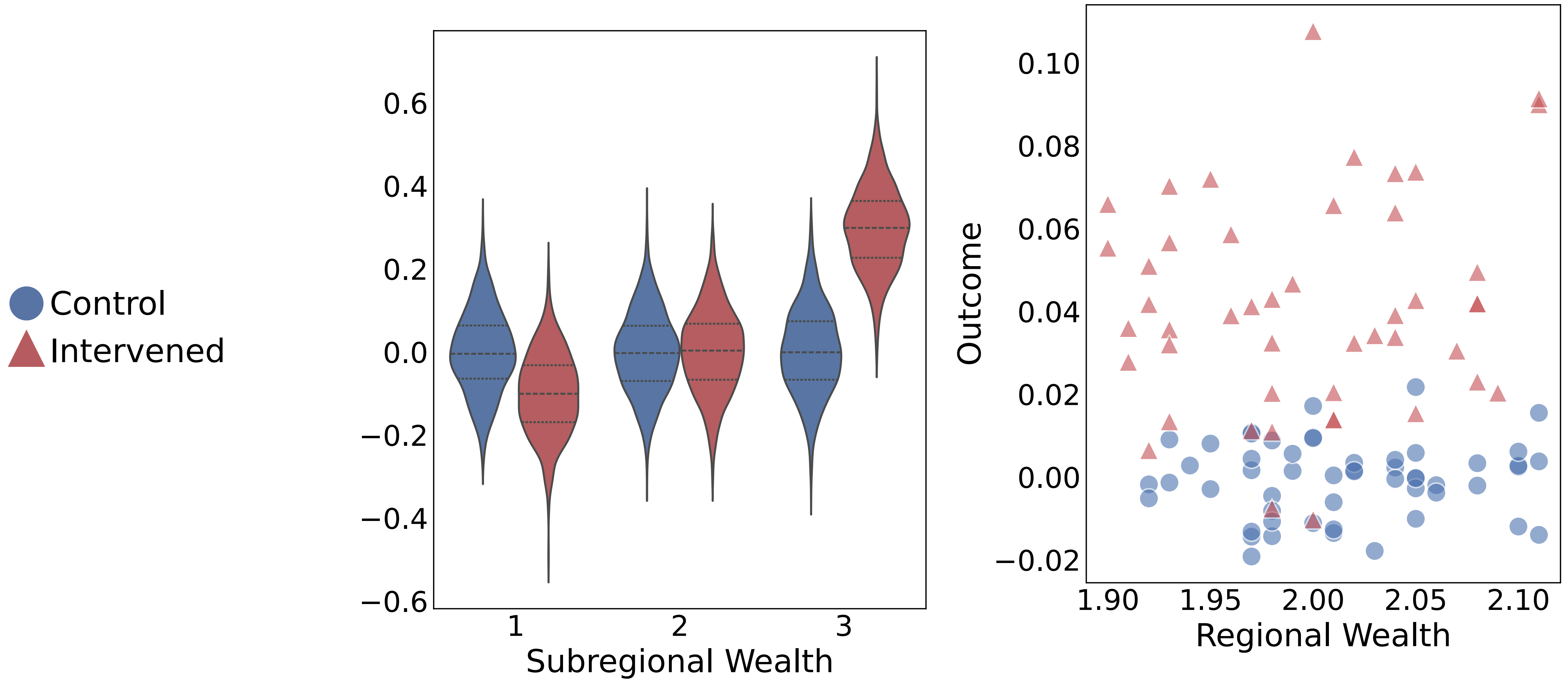
Binary Treatment to Regions
region

Context : 3 wealth levels {0,1,2}

Aggregation : Summing over all subregions

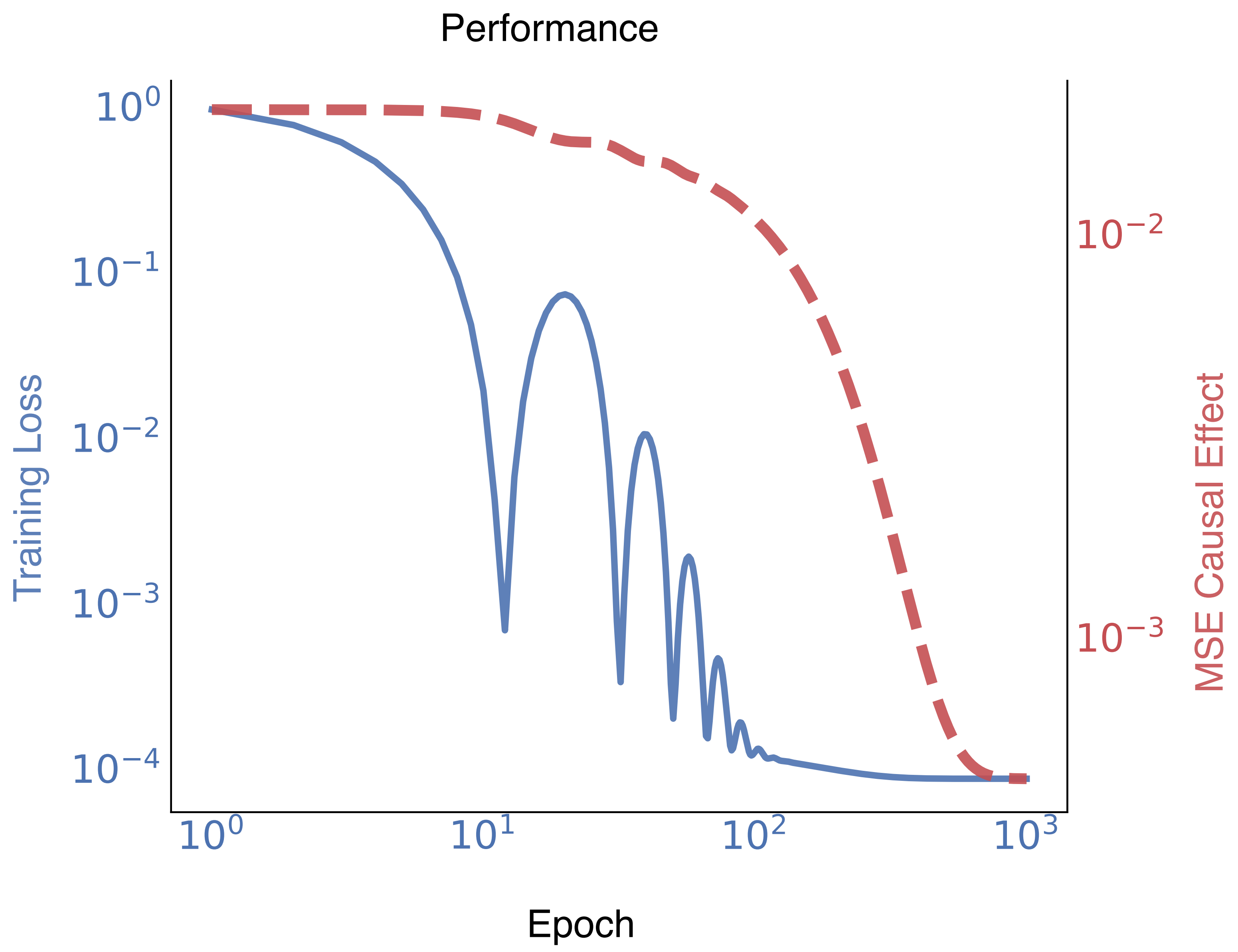
10x10 : Regions
10x10 subregions in each region



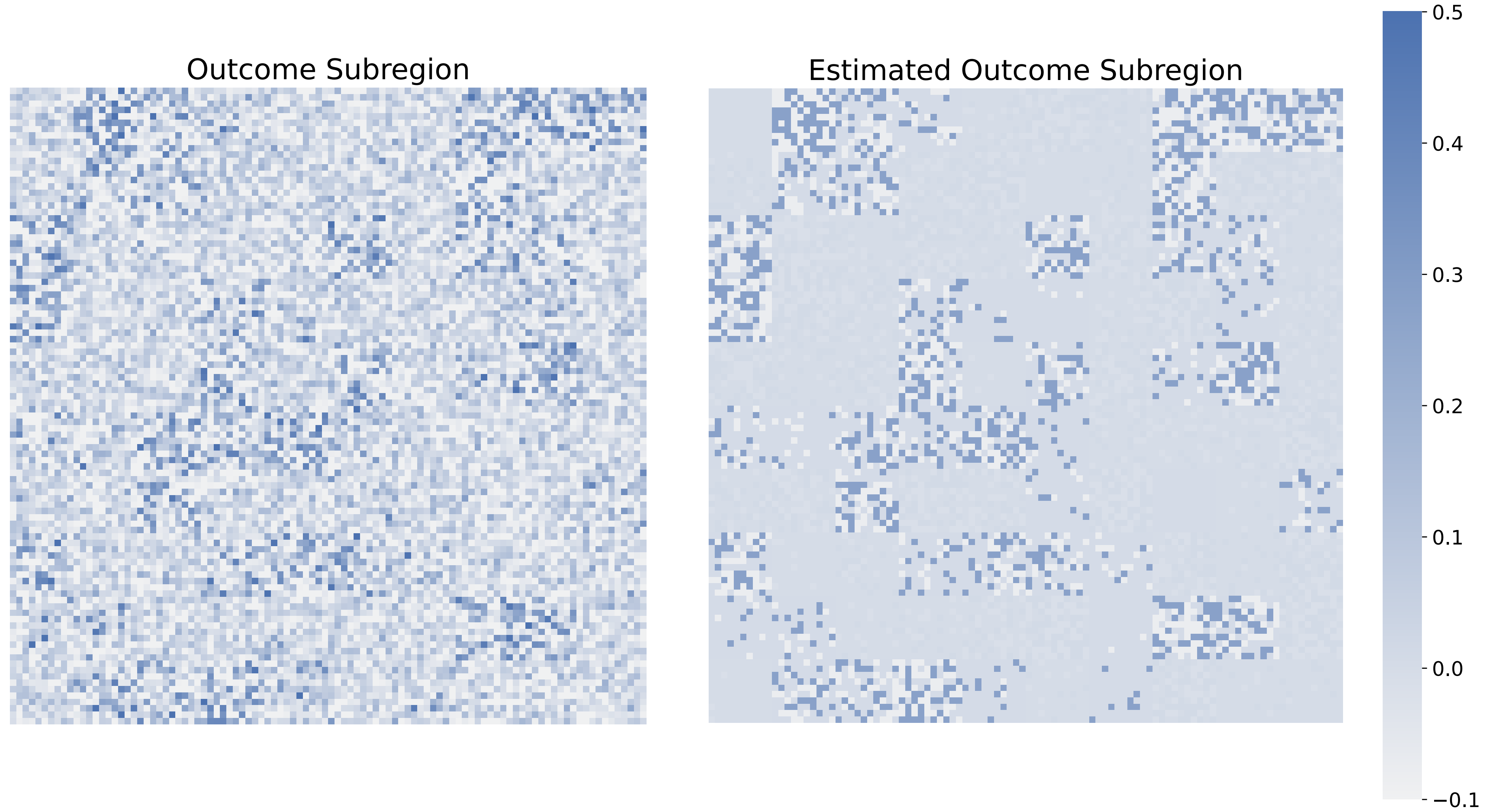


Political Campaigning

Experiment 1 : How does campaigning affect politician performance

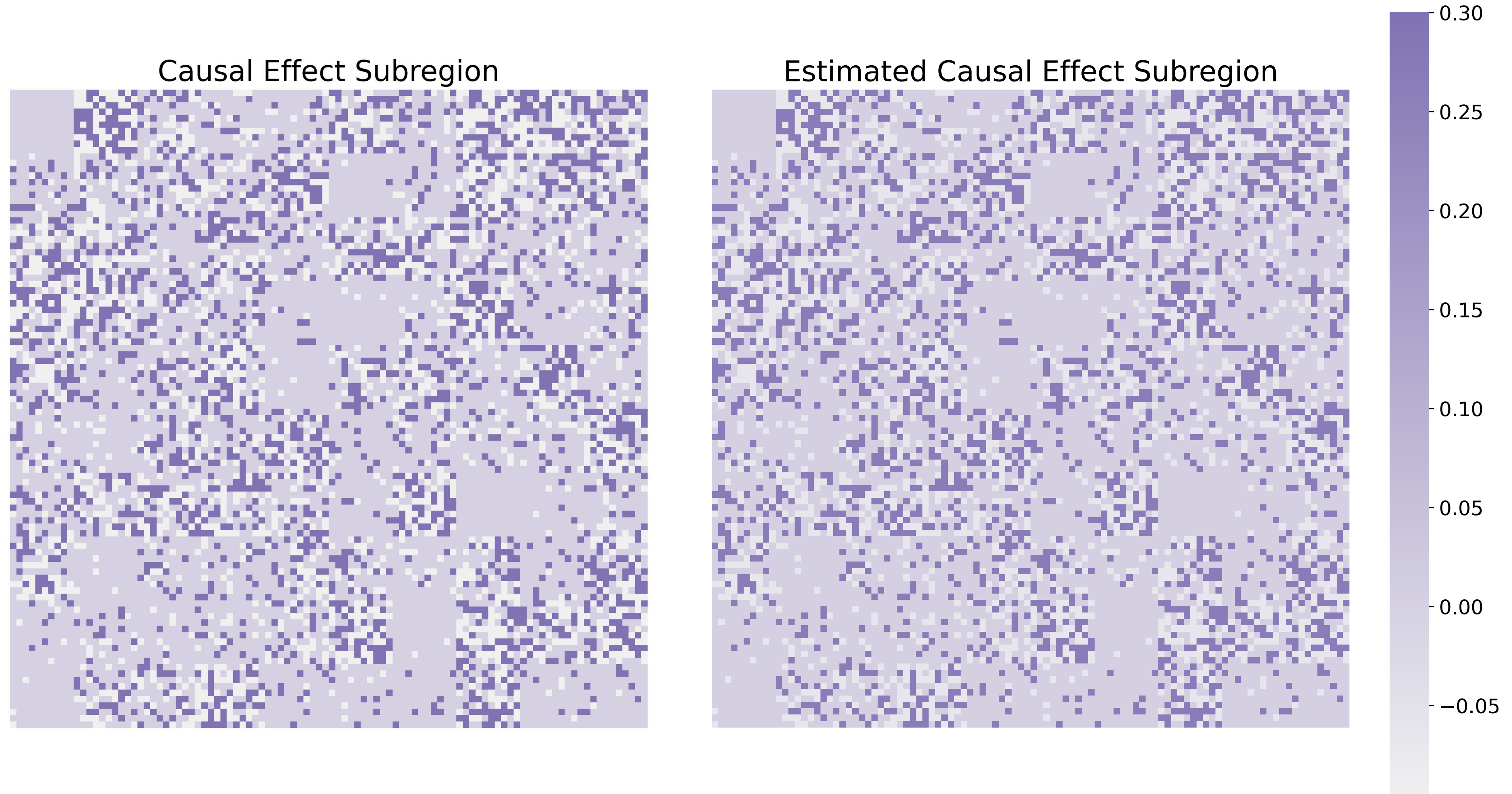


Ground truth vs estimated subregional outcomes



$$f_{\boldsymbol{\theta}}(t_{i,j}, c_{i,j}) = \begin{cases} \theta_1 & \text{if } t_{i,j} = 0 \text{ and } c_{i,j} = 1 \text{ (poor)}, \\ \theta_2 & \text{if } t_{i,j} = 0 \text{ and } c_{i,j} = 2 \text{ (middle class)}, \\ \theta_3 & \text{if } t_{i,j} = 0 \text{ and } c_{i,j} = 3 \text{ (rich)}, \\ \theta_4 & \text{if } t_{i,j} = 1 \text{ and } c_{i,j} = 1 \text{ (poor)}, \\ \theta_5 & \text{if } t_{i,j} = 1 \text{ and } c_{i,j} = 2 \text{ (middle class)}, \\ \theta_6 & \text{if } t_{i,j} = 1 \text{ and } c_{i,j} = 3 \text{ (rich)}. \end{cases}$$

Ground truth vs estimated subregional causal effect

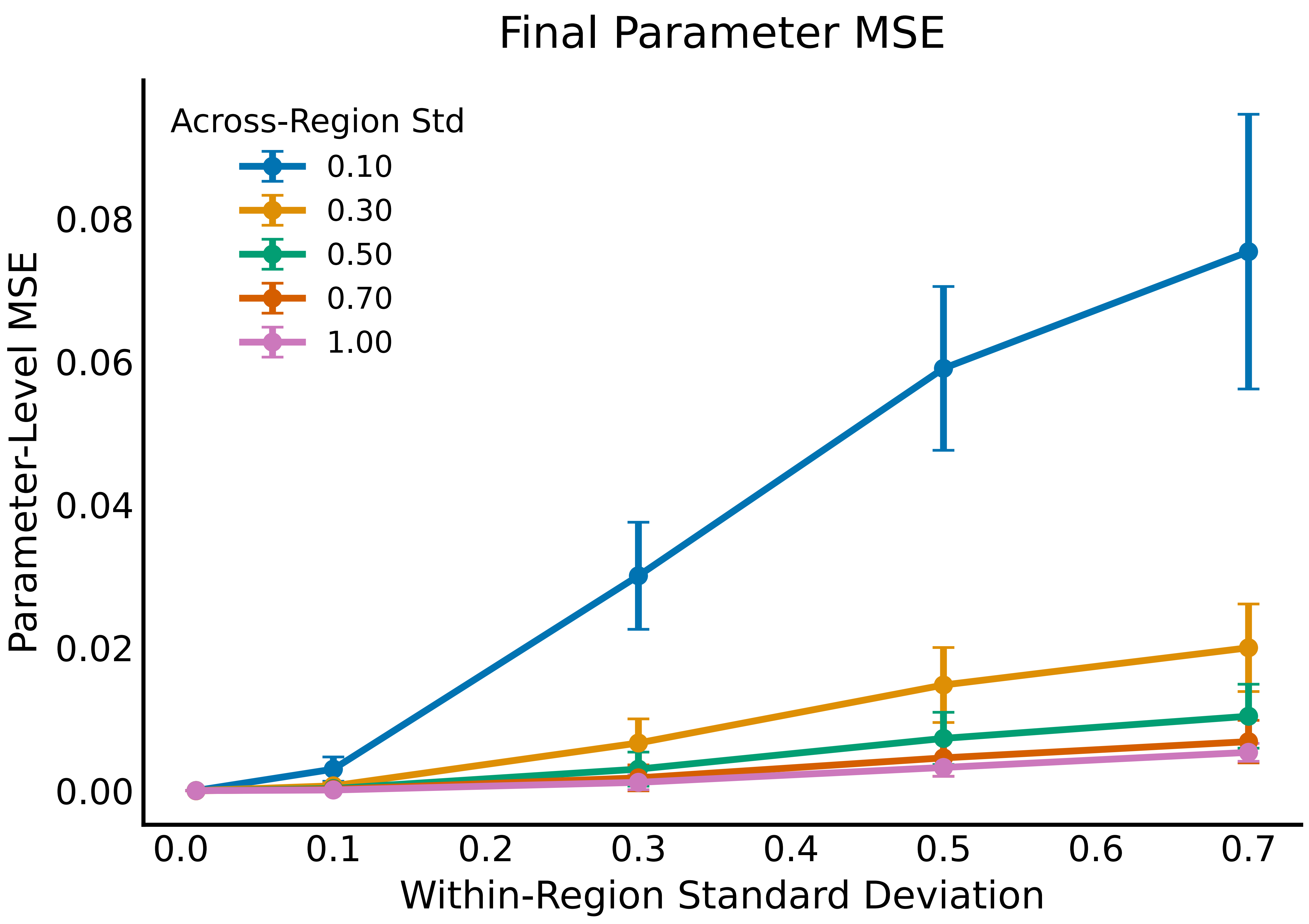


Ablation Study on Experiment 1

Effect of varying intra-region heterogeneity

Low variability across regions - results in underdetermination of the inverse problem

High dimensional contextual covariates at the subregion level can help



Unknown Intervention Locations

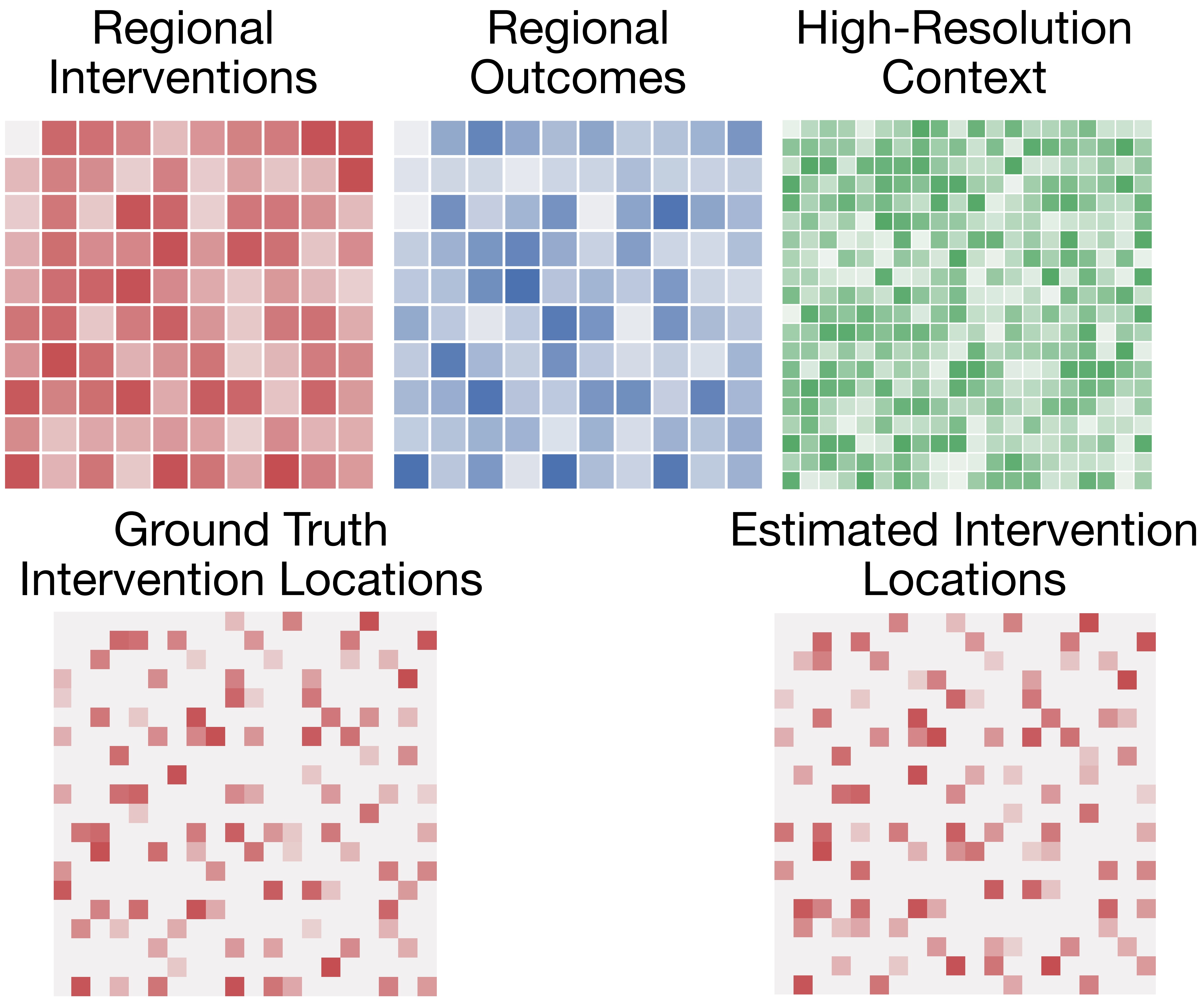
Experiment 2: High school funding vs educational outcomes

Real valued treatment to each region

Real valued context (socio-economic status) between (0,1)

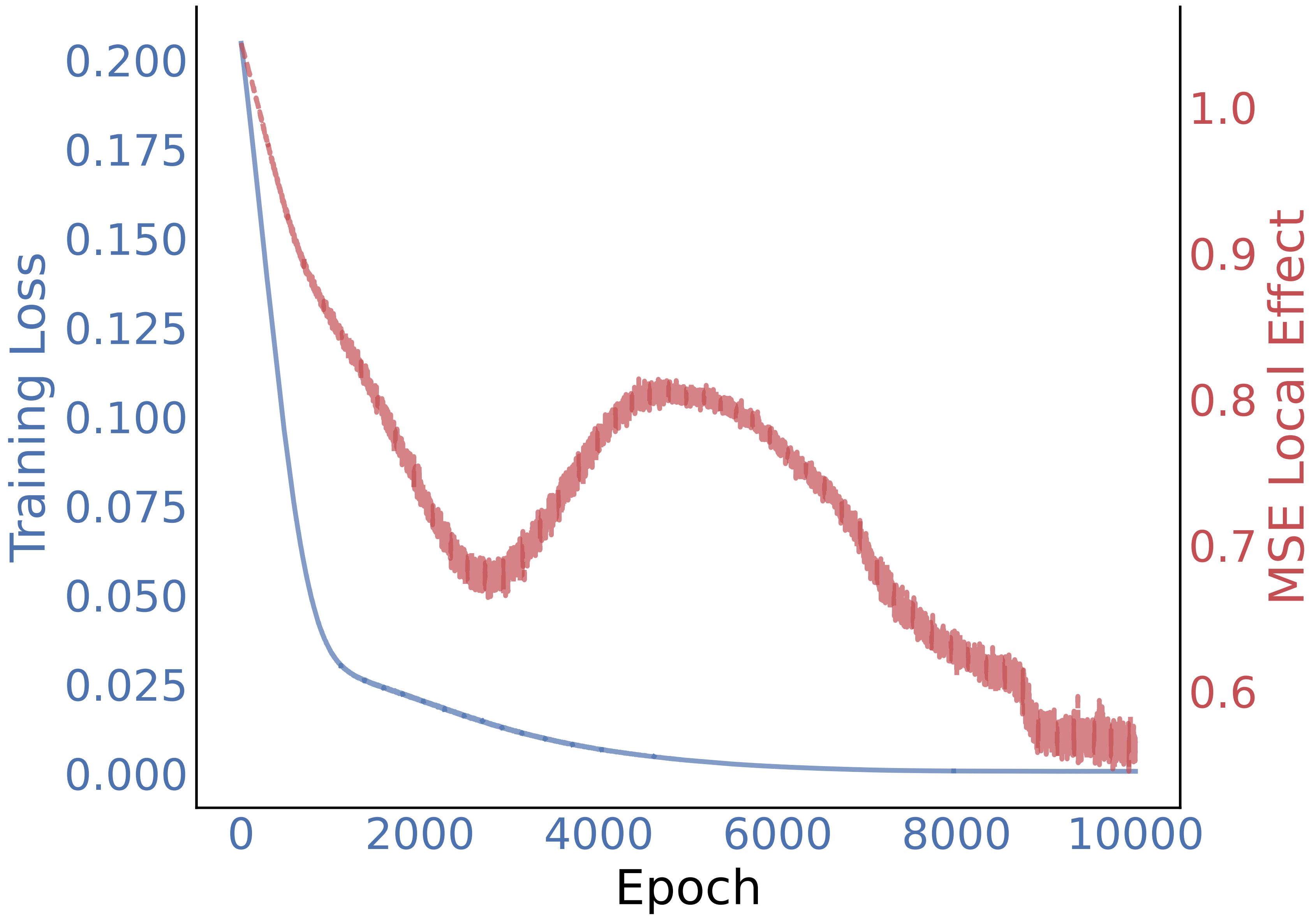
Mean over subregional outcomes

10x10 : Regions
4x4 subregions in each region



Unknown Intervention Locations

Experiment 2: How does campaigning affect politician performance

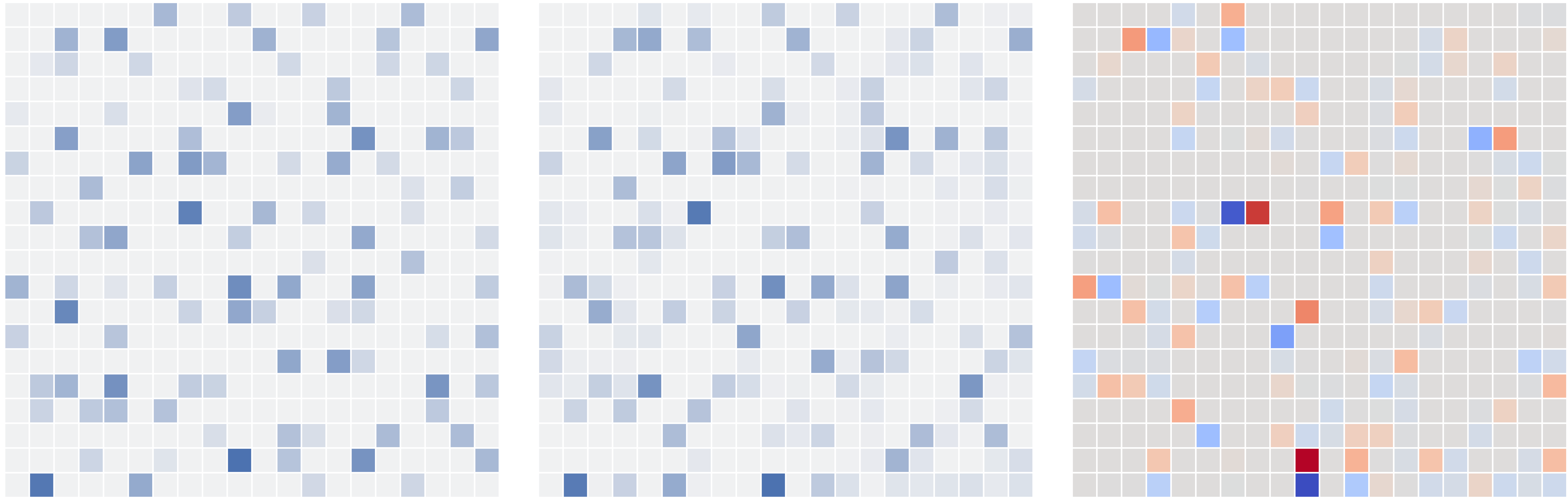


$$f_\theta(t_{i,j}, c_{i,j}) = (\theta_{\text{shift}} - c_{i,j}) \cdot \theta_{\text{scale}} \cdot t_{i,j}$$

Ground truth Causal Effect

Estimated Causal Effect

Estimation Error



- 400 free parameters 4x100 matrix
- $\theta_{\text{shift}}, \theta_{\text{scale}}$

Temperature Control softmax over each row
For intervention

Mean aggregate subregions and match effect to regional effect

Hidden Confounding

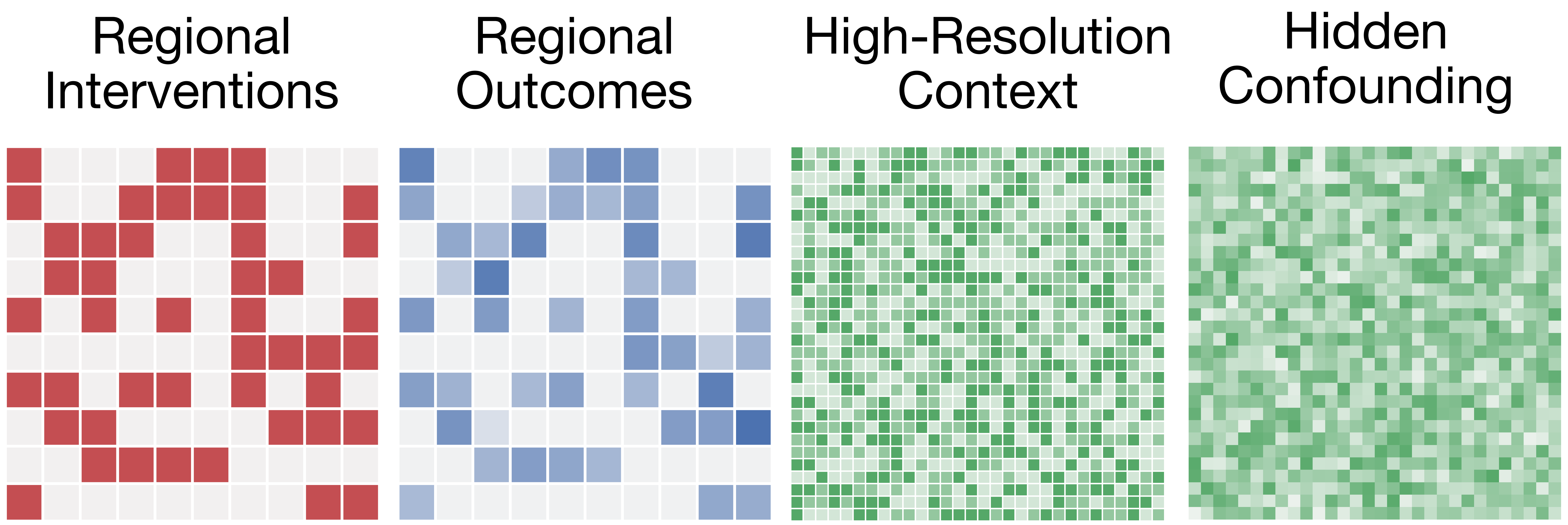
Experiment 3: Extreme heat on educational outcomes

Binary treatment (heatwave or not) to each region for all weeks

Context: High(3), medium(2), low(1) parental education level - slowly evolving over time

Hidden confounding: Vegetation (0 or 1) static over time

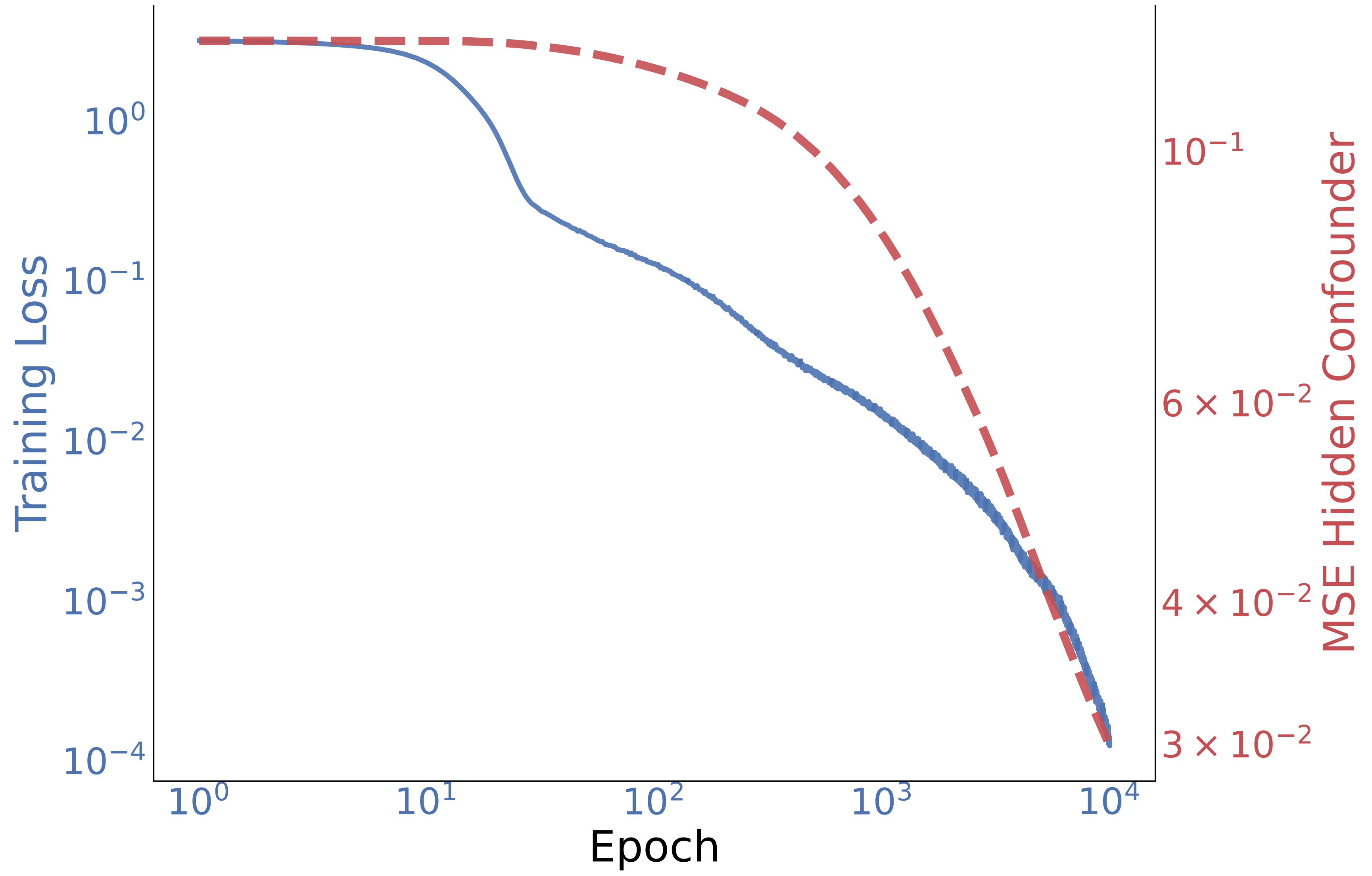
10x10 : Regions
3x3 subregions in each region

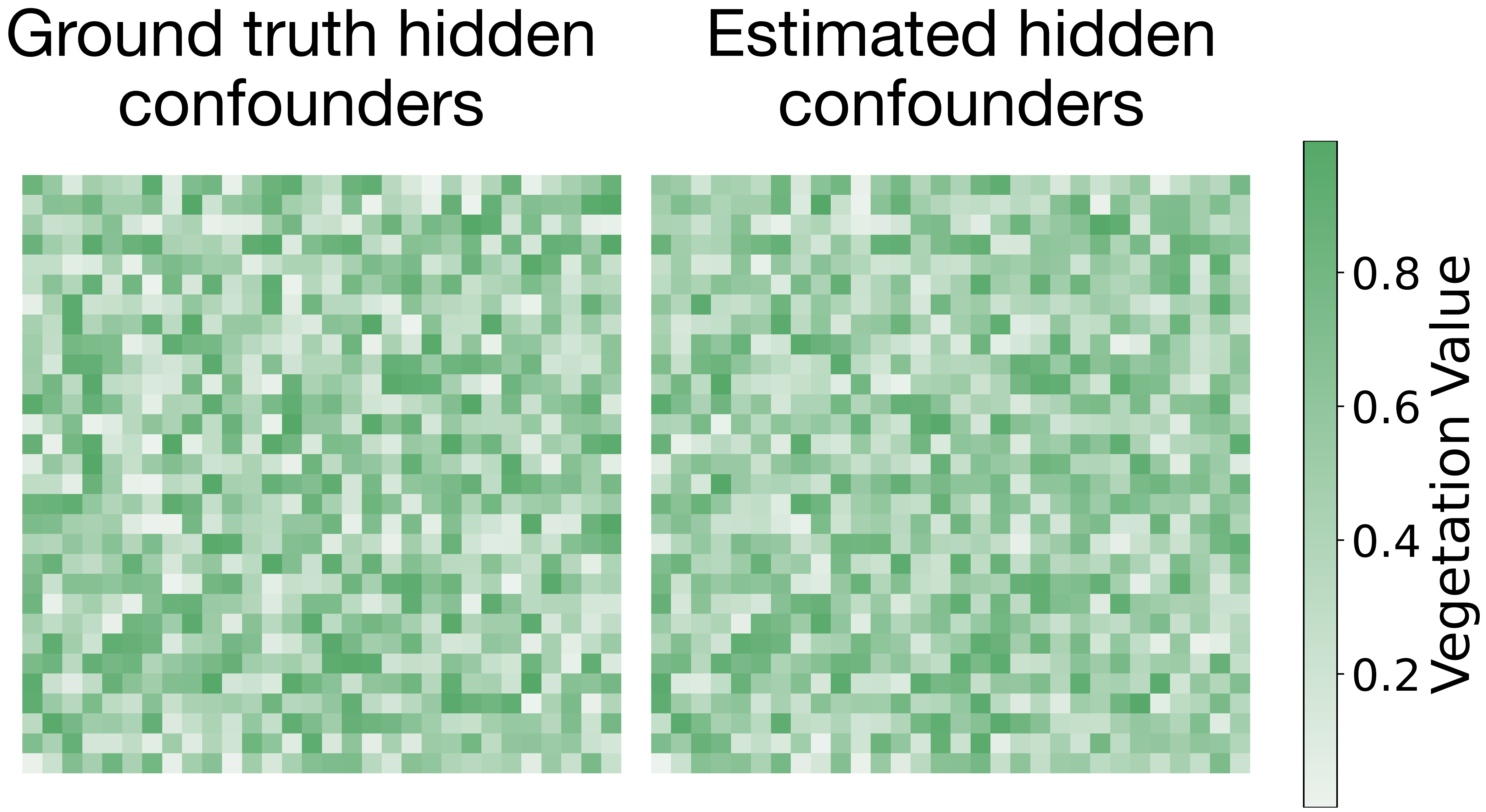


$$f_{\boldsymbol{\theta}}\left(t_{i,j}^{(w)}, c_{i,j}^{(w)}, u_{i,j}\right) = \begin{cases} 0 & \text{if } t_{i,j}^{(w)} = 0, \\ \left(10 \cdot \mathbb{1}[c_{i,j}^{(w)} = 1] + 5 \cdot \mathbb{1}[c_{i,j}^{(w)} = 2] \right. & \\ \quad \left. + \ \mathbb{1}[c_{i,j}^{(w)} = 3]\right) \cdot (1 - u_{i,j}) & \text{if } t_{i,j}^{(w)} = 1. \end{cases}$$

Hidden Confounding

Estimation of hidden confounders





- Demographic data
- Heatwave or not
- Hidden Confounder estimates

Trainable parameters are θ (neural net params) and $\hat{U}$ (confounding)

Mean aggregation to the regional level and MSE loss

Learning the Aggregation Function

Experiment 4: Driving ban vs air quality

Binary Treatment : Driving ban implemented or not

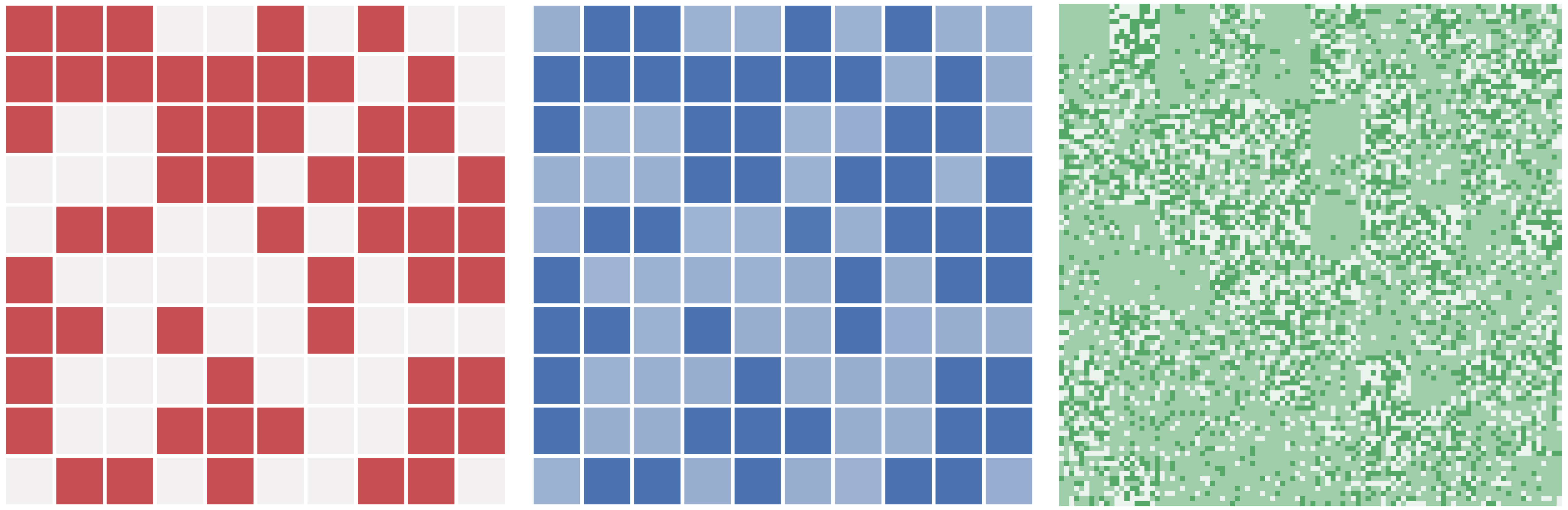
High resolution context : Vegetation in the region

$$f_{\boldsymbol{\theta}}(t_{i,j}, c_{i,j}) = \begin{cases} 0.0 & \text{if } t_{i,j} = 0, \\ \theta_{\text{base}} + \theta_{\text{veg}} \cdot c_{i,j} & \text{if } t_{i,j} = 1, \end{cases}$$

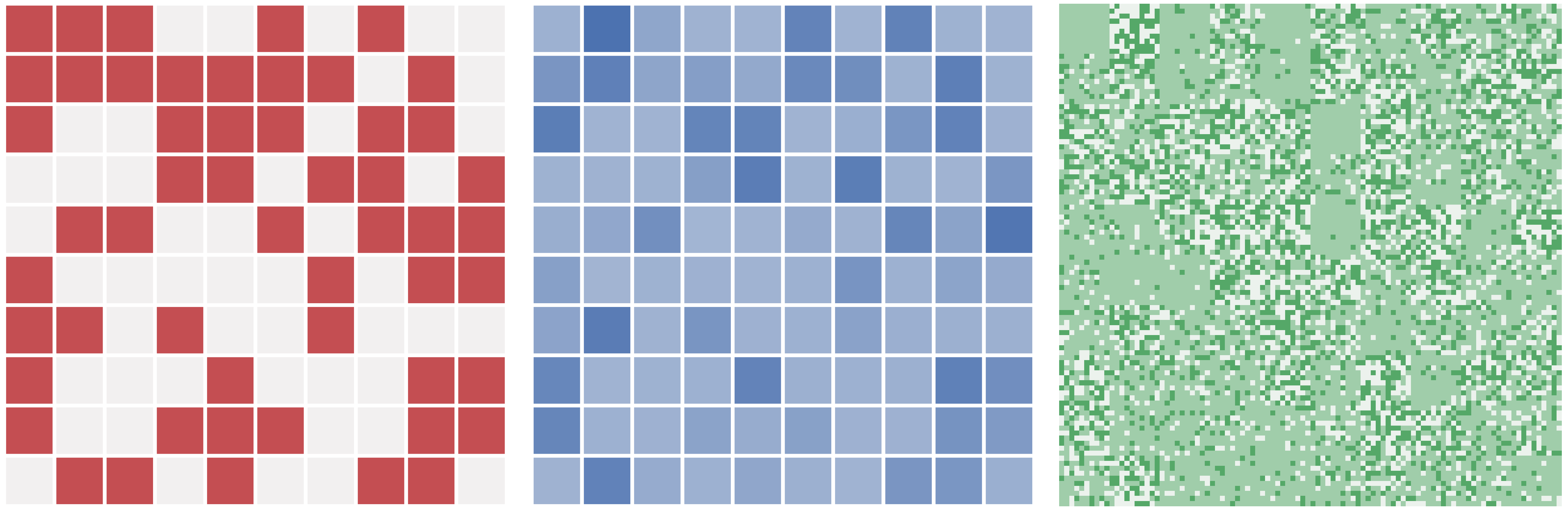
$$p_{i,j}(\tau) = \frac{\exp(\hat{x}_{i,j}/\tau)}{\sum_{k=1}^{M} \exp(\hat{x}_{i,k}/\tau)}$$

$$x_i = \sum_{j=1}^{M} p_{i,j}(\tau) \cdot \hat{x}_{i,j},$$

Max Aggregation ($\beta = 0.1$)



Mean Aggregation ($\beta = 6$)



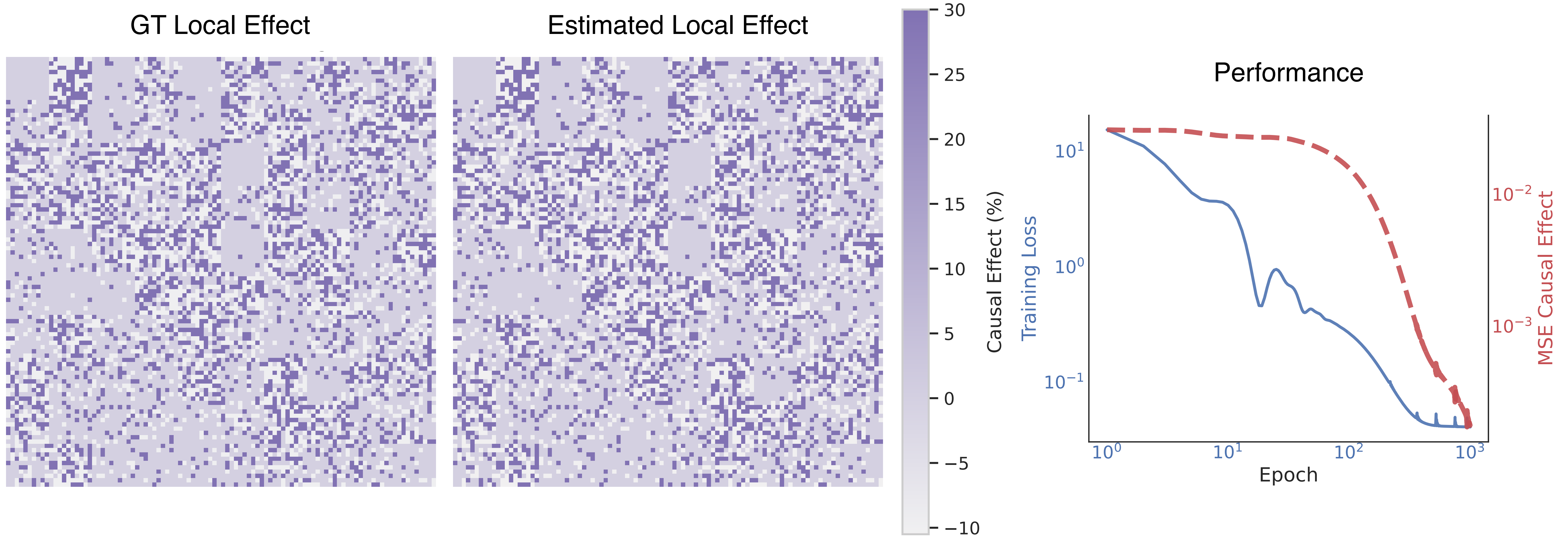
Learning the Aggregation Function

Experiment 4: Driving ban vs air quality

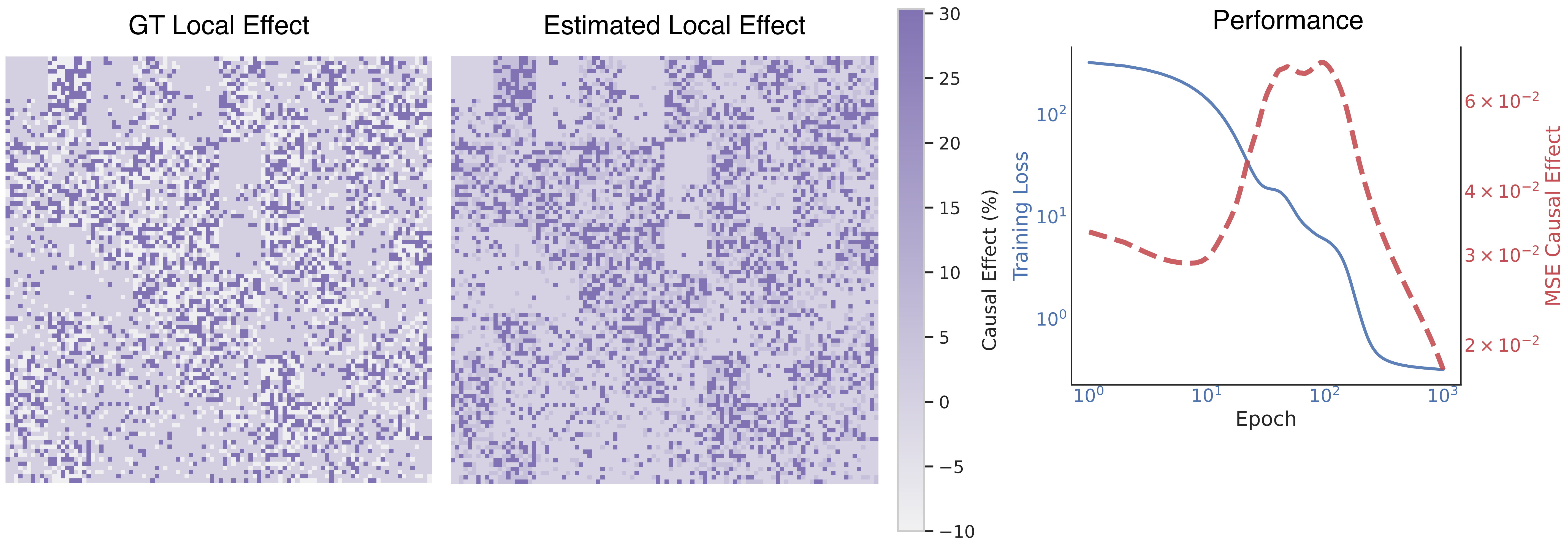
Learnable parameters are $\theta_{\text{base}}, \theta_{\text{veg}}$ and τ which controls the temperature of the aggregation

We see for max aggregation retrieval of causal effect for lower-vegetation regions is poor as expected

Mean Aggregation



Max Aggregation



Covariate Dependent Intervention Allocation

Experiment 5: School funding dependent on context

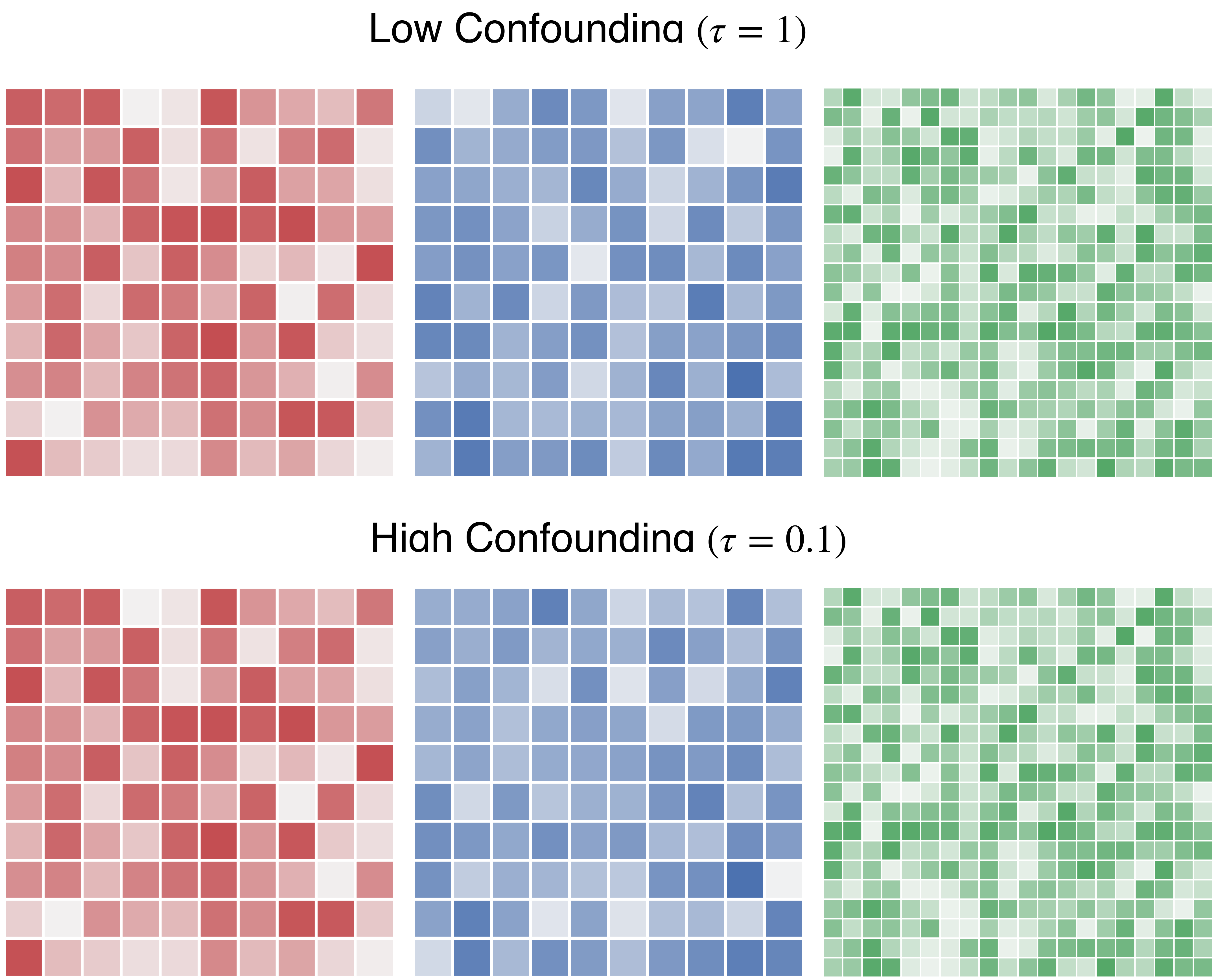
Binary Treatment to each region, with exactly one subregion recieving treatment

Real valued b/w (0,1) high dimensional sub-region context

$$p_{i,j} = \frac{\exp(\ell_{i,j})}{\sum_{k=1}^{M} \exp(\ell_{i,k})}.$$

$$j^{\star} \sim \text{Categorical}(p_{i,1:M})$$

And treat selected region



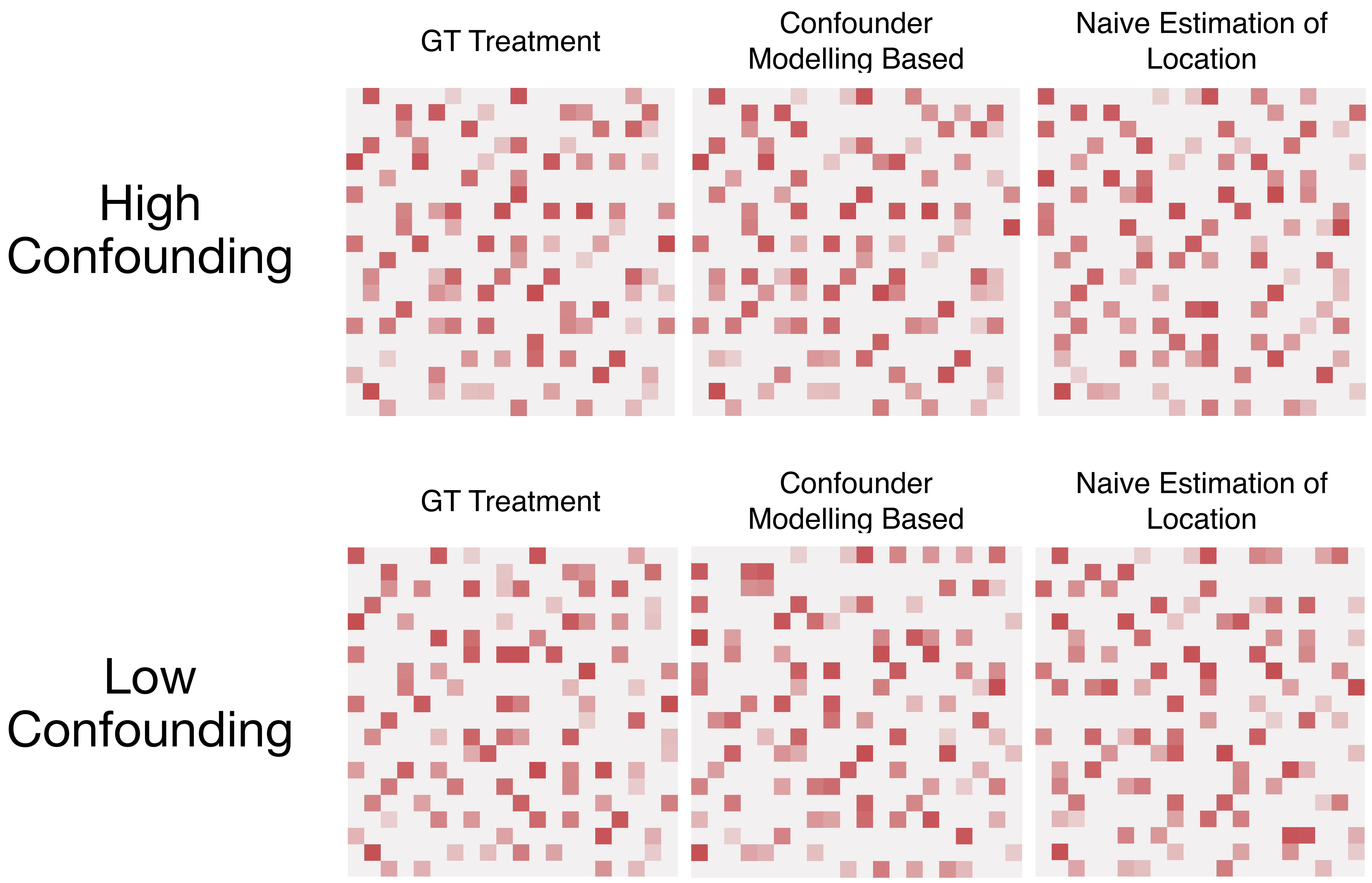
Covariate Dependent Interventions

Experiment 5: School funding dependent on context

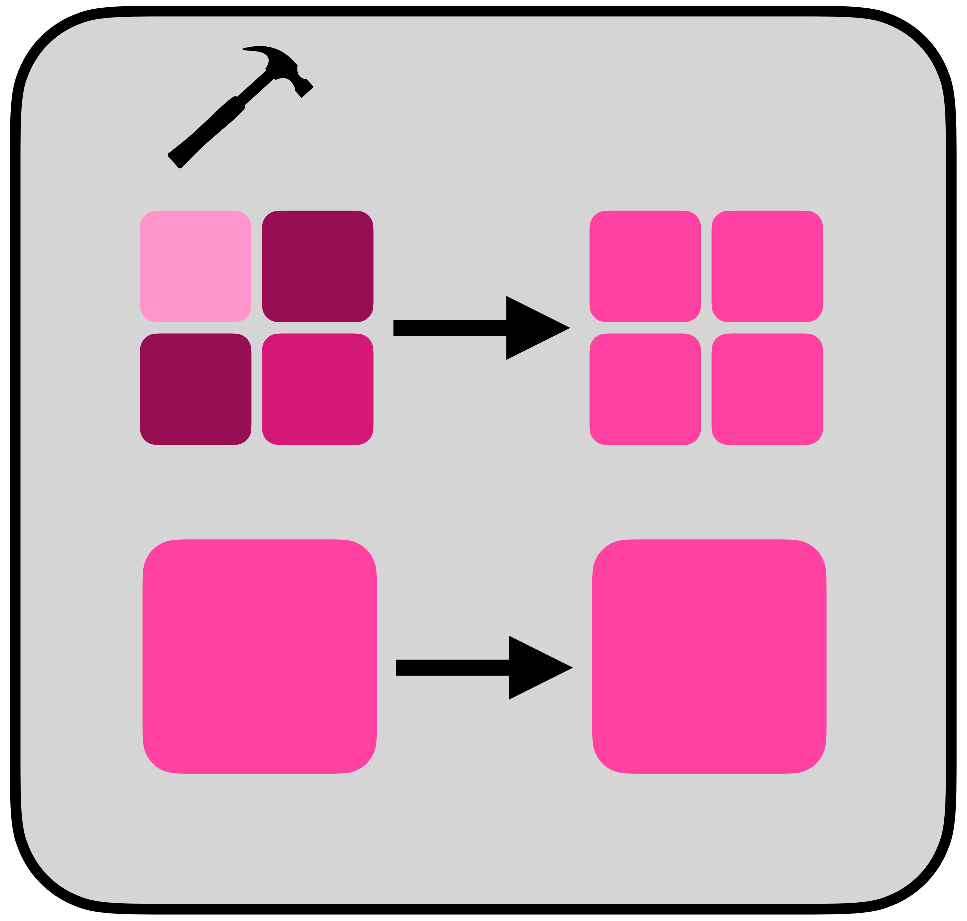
Learnable parameters are θ_{base}, θ_{soc} and τ_{conf} which controls the amount of confounding

We see that the retrieval of the location for high confounding is much better when we model confounding

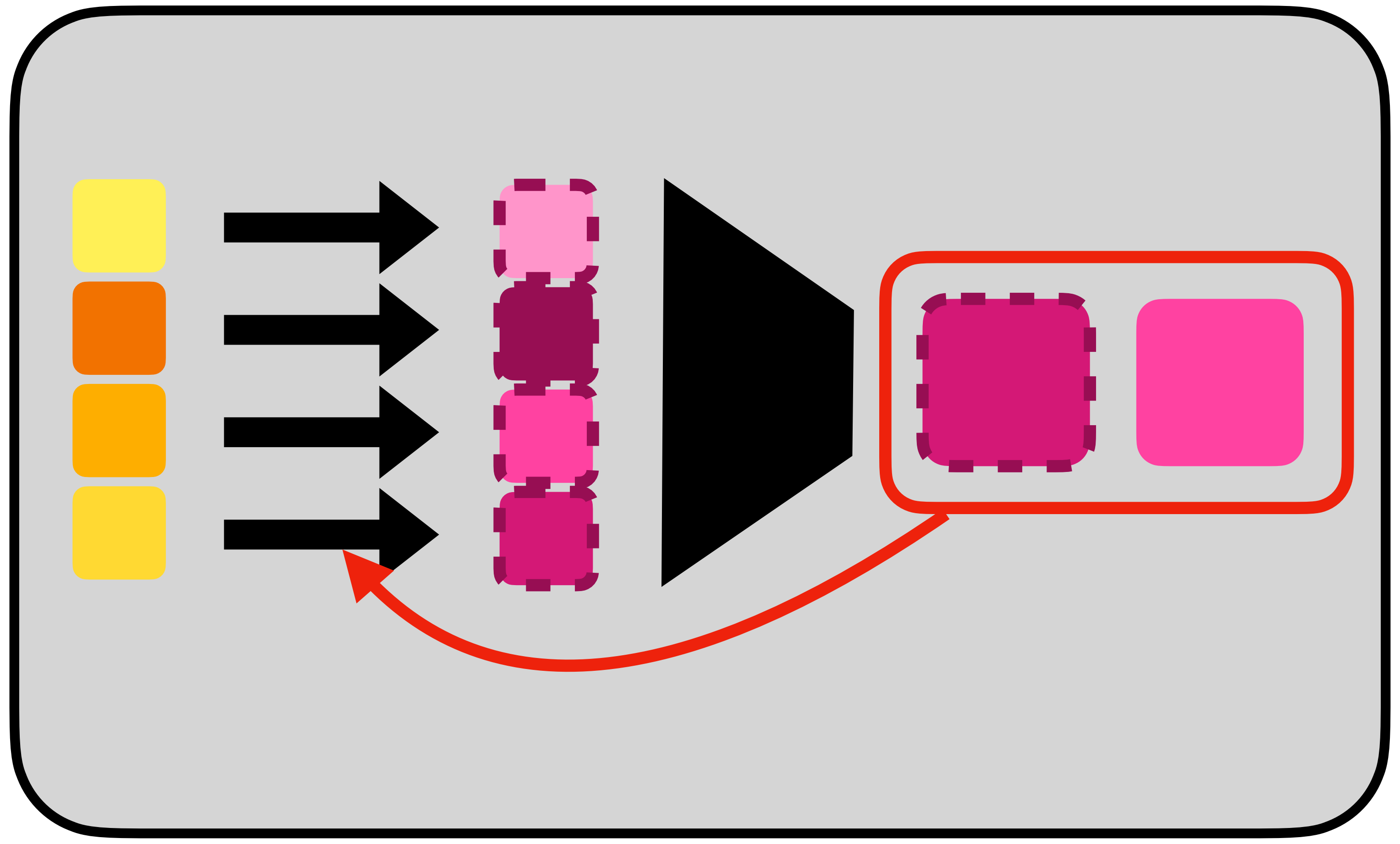
We see that the retrieval of the location for low confounding is similar in both methods



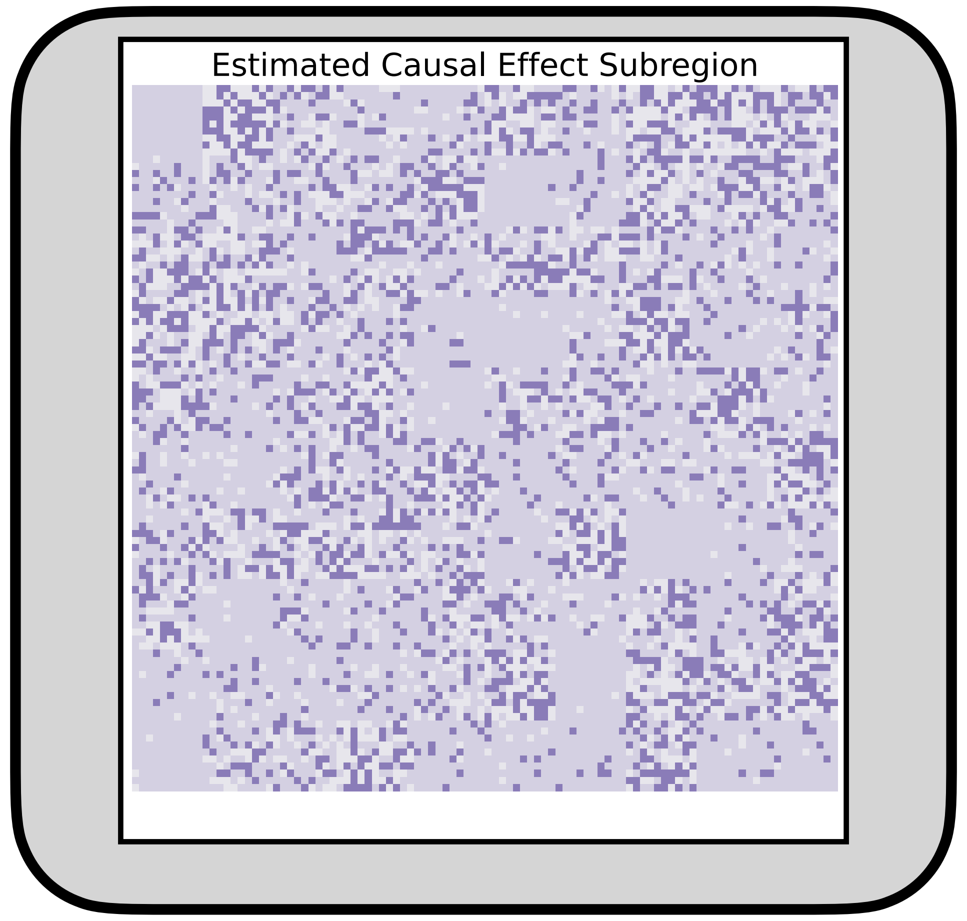
Summary



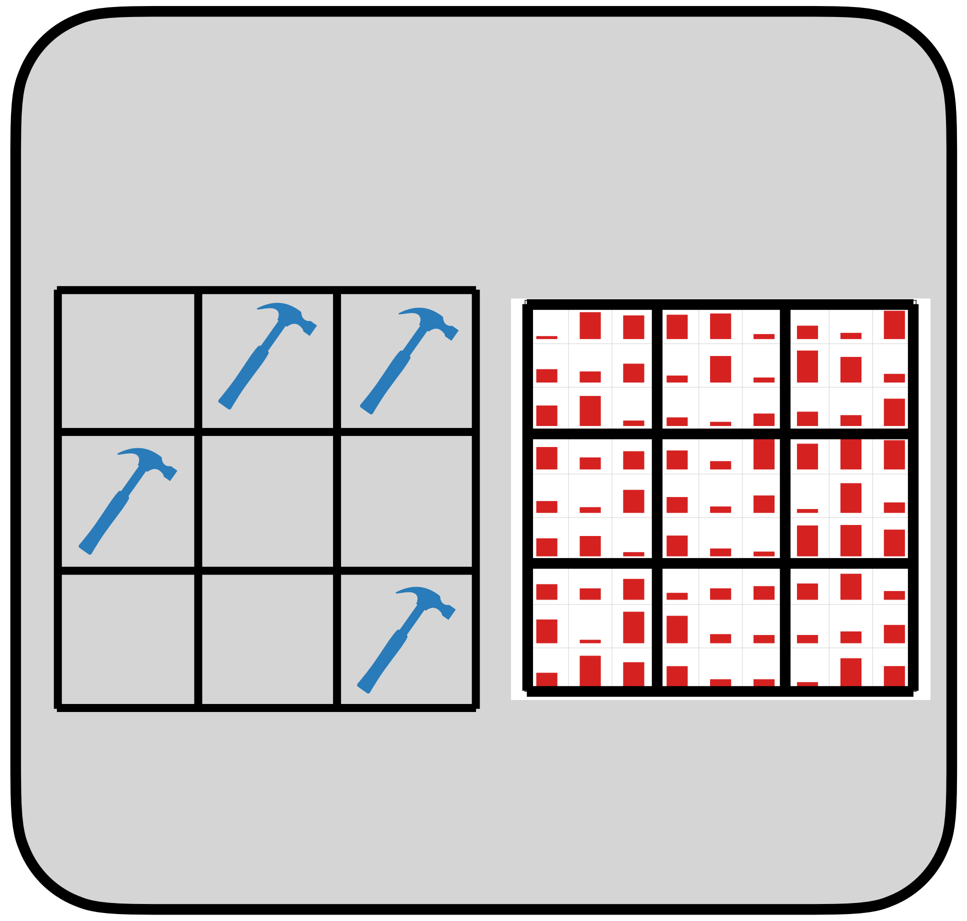
Aggregated causal data
-incorrect conclusions



Learn true causal effects- using
high resolution covariates



Predict localised causal
effects of policy



Generate counterfactuals
- plan better policies

Feedback and Questions?