Causal inference for sustainable and resilient agriculture

Vasileios Sitokonstantinou

Postdoc Researcher, Beyond Center, National Observatory of Athens, Greece

Joint work with Georgios Giannarakis, Ilias Tsoumas, Roxanne S. Lorilla, Alkiviadis Koukos and Charalampos Kontoes

Agriculture: Modern Challenges

Ever-increasing deamand for agricultural products

Climate change, Environmental degradation

“Sustainable Intensification”

A diverse world

Differences in climate, soil, land use

Different environmental responses to interventions carried out by farmers

Effect Heterogeneity

(no one-size-fits-all solution)

What's next? (EU)

The new common agricultural policy: 2023-27

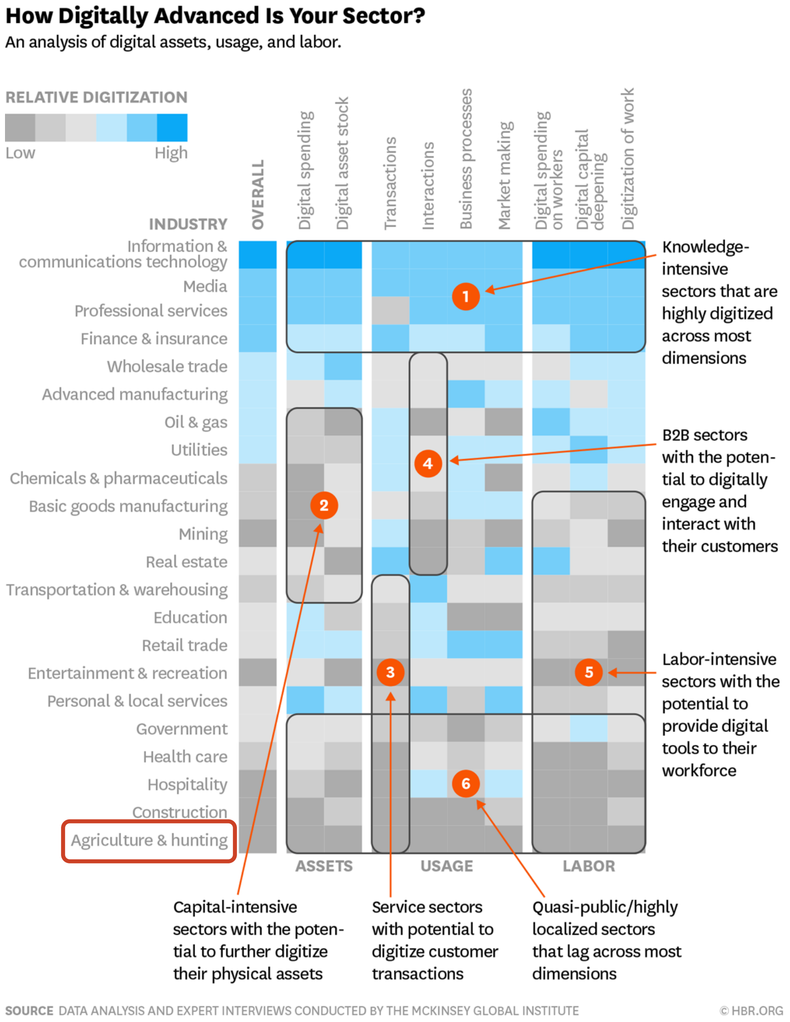
The new common agricultural policy will be key to securing the future of agriculture and forestry, as well as achieving the objectives of the European Green Deal.

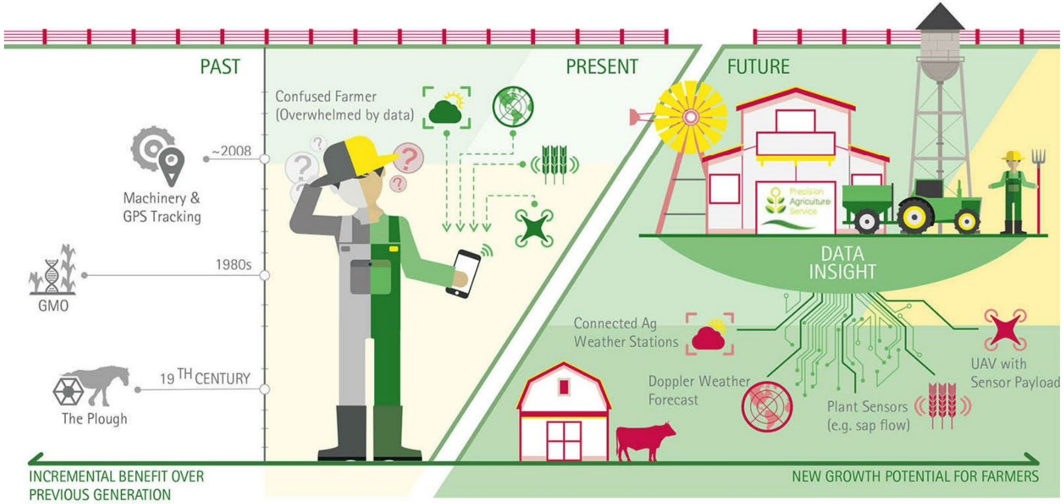
On 2 December, 2021, the agreement on reform of the common agricultural policy (CAP) was formally adopted. The new legislation, which is due to begin in 2023, paves the way for a fairer, greener and more performance-based CAP.

It will seek to ensure a sustainable future for European farmers, provide more targeted support to smaller farms, and allow greater flexibility for EU countries to adapt measures to local conditions.

- Targeted support
- Flexibility to adapt measures to local conditions

Geospatial "personalization"





smart farming technologies could drive to the application of required sustainable agriculture practices

but limited adoption

Farmer needs:

- ↦ actionable advice
- ↦ evidence about effectiveness & benefits

Causal inference for sustainable agriculture

Geospatial personalization of agricultural practices

Trustworthy climate-smart digital tools

Causal inference
What would happen

How much will net primary productivity be affected if we increase crop rotation by 5% in Flanders?

How much will my yield increase if I sow this week instead of next week?

Geospatial personalization of agricultural practices

Towards assessing agricultural land suitability with causal machine learning

Georgios Giannarakis[1] Vasileios Sitokonstantinou[1,2] Roxanne Suzette Lorilla[1]
Charalampos Kontoes[1]
[1]BEYOND Center, IAASARS, National Observatory of Athens, Greece
[2]Remote Sensing Laboratory, National Technical University of Athens, Greece
{giannarakis, vsito, rslorilla, kontoes}@noa.gr

Personalizing Sustainable Agriculture with Causal Machine Learning

Georgios Giannarakis
BEYOND Centre, IAASARS
National Observatory of Athens
giannarakis@noa.gr

Vasileios Sitokonstantinou
BEYOND Centre, IAASARS
National Observatory of Athens
vsito@noa.gr

Roxanne Suzette Lorilla
BEYOND Centre, IAASARS
National Observatory of Athens
rslorilla@noa.gr

Charalampos Kontoes
BEYOND Centre, IAASARS
National Observatory of Athens
kontoes@noa.gr

Tackling Climate Change with Machine Learning: workshop at NeurIPS 2022

Our approach

Treating ALSA as a geospatial impact assessment problem leveraging EO and other large scale geospatial data

Train causal models estimating the impact of agricultural practices on metrics of interest

Propose the estimated impact as a land suitability score

Guide agricultural policy making by prioritizing high-gain practices per land unit

Conditional Average Treatment Effects (CATEs)

What is the impact of a treatment for a unit with particular characteristics?

Agricultural Practice  Land unit  Agro-environmental info 

$$\theta(x) = \mathbb{E}[Y(1) - Y(0) | X = x]$$

Potential NPP when practice is applied

Potential NPP when practice is not applied

Outcome Y: ecosystem services, soil organic carbon, **net primary productivity**

Double ML (Chernozukov et. al, 2016)

- Flexible framework for CATE estimation
- Robust for spatial data [1]

$$Y = \theta(X) \cdot T + g(X) + \varepsilon$$

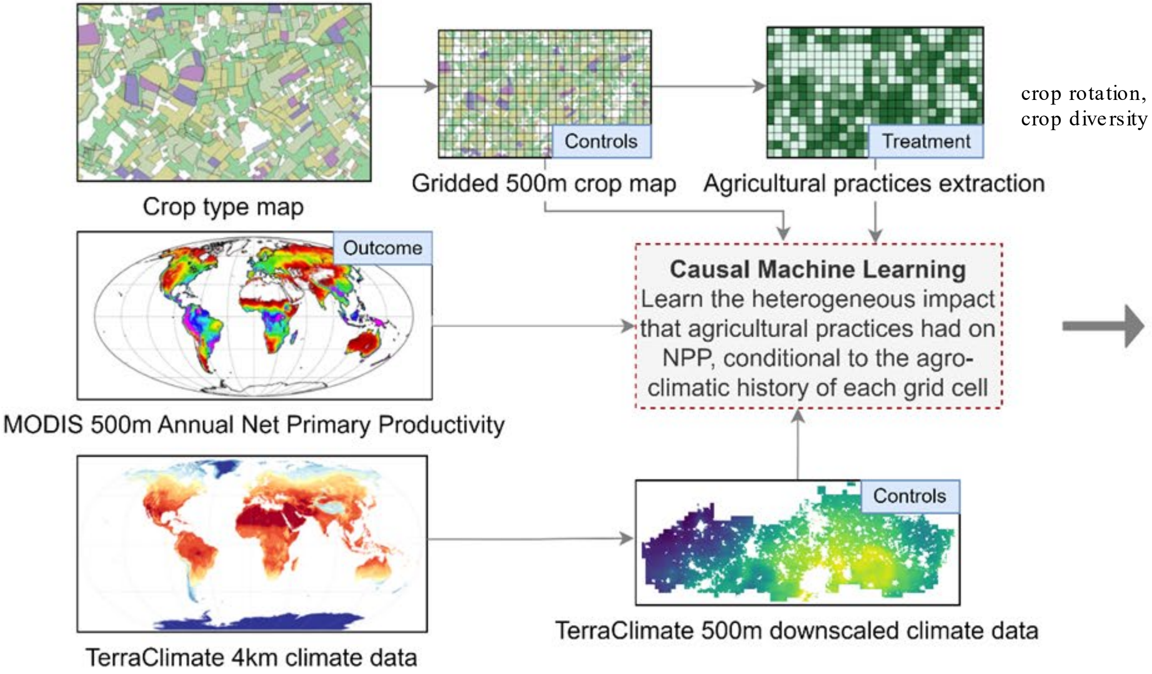
$$T = f(X) + \eta$$

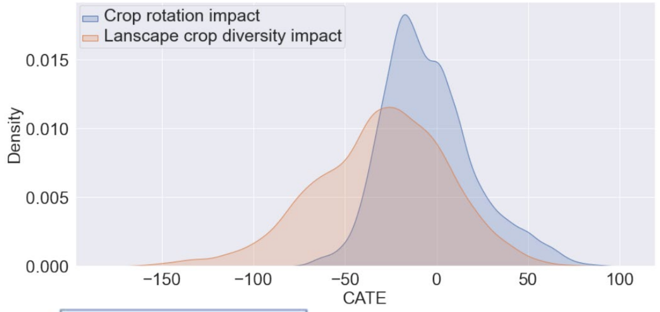
$$\hat{\theta} = \arg\min_{\theta \in \Theta} \mathbb{E}\left[(\tilde{Y} - \theta(X) \cdot \tilde{T})^2\right]$$

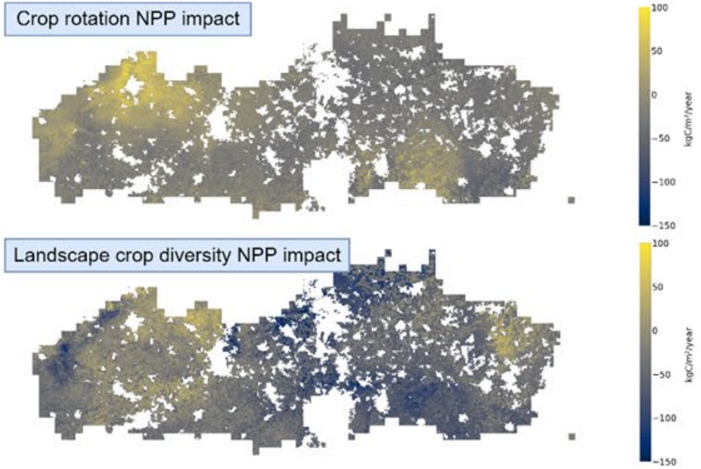
[1] *Approaches to spatial confounding in geostatistics*, Gilbert et al., 2022

Methodology & Results

(Flanders, Belgium, 2010-2020)



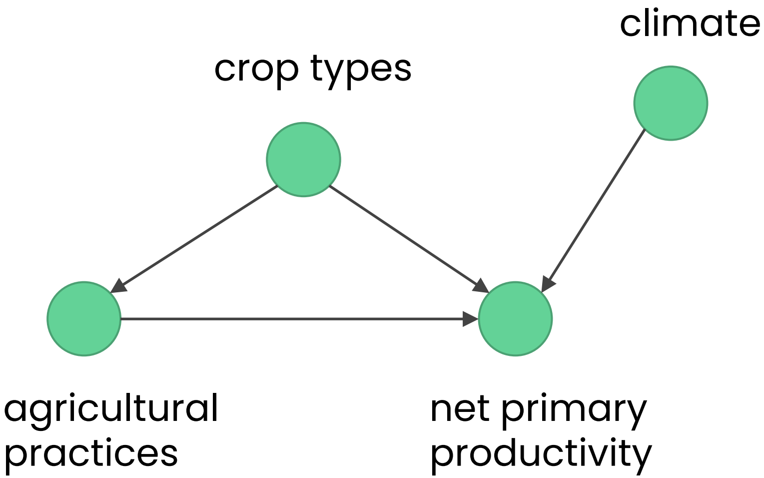




- The learned impact as a Land Suitability Score
- Understand agro-climatic conditions driving impact variability

CR ATE: 1.08 (95% CI [-20.35, 22.51])   LCD ATE: -35.73 (95% CI [-58.73, -12.73])

Limitations & Future Work



The causal graph assumed is simplistic, some bias and variance in estimates remains:

more domain knowledge & causal discovery methods should be incorporated

The Double ML effect estimates are not evaluated in the typical sense as ground truth is not observed:

robustness checks and sensitivity analyses should be performed. Agriculture: harness existing knowledge of field experiments

Causal inference for sustainable agriculture

Trustworthy climate-smart digital tools

Evaluating Digital Tools for Sustainable Agriculture using Causal Inference

Ilias Tsoumas*[1,2] **Georgios Giannarakis***[1] **Vasileios Sitokonstantinou**[1]
Alkiviadis Koukos[1] **Dimitra Loka**[3] **Nikolaos Bartsotas**[1]
Charalampos Kontoes[1] **Ioannis Athanasiadis**[2]
[1]BEYOND Centre, IAASARS, National Observatory of Athens
[2]Wageningen University and Research
[3]Hellenic Agricultural Organization ELGO DIMITRA

Tackling Climate Change with Machine Learning: workshop at NeurIPS 2022

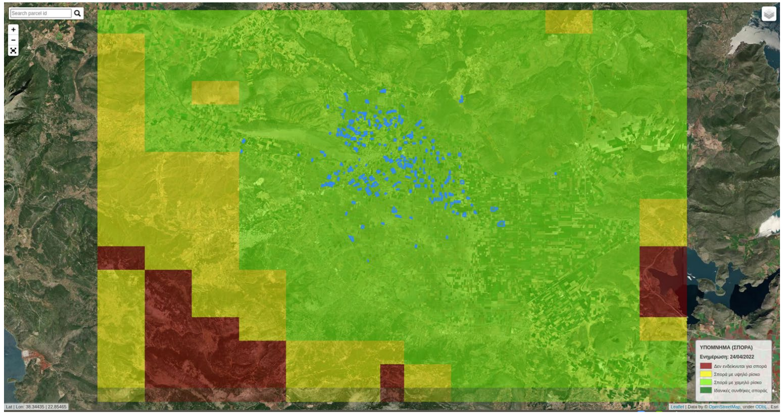
The case of a knowledge-based recommendation system for optimal cotton sowing

Answering on a real need of cotton farmers.
Is today a good day to sow?

Collaboration with a farmer's cooperative (171 cotton fields) in Orchomenos,Viotia-Greece

They have developed highly consolidated routines for interacting with their crops: this includes common practices, homogeneous fertilizer application, and jointly owned machinery.

pilot of sowing map for cotton for cultivation period of 2021 in Orchomenos, GR



but what is the actual impact of the recommended actions?

Our approach

Model the farm system using a causal graph, and identify the effect of sowing on a recommended day on the yield the farmer observed.

Unit	Field
Treatment (T)	The field was sown on a recommended day
Outcome (Y)	Yield observed at the end of season

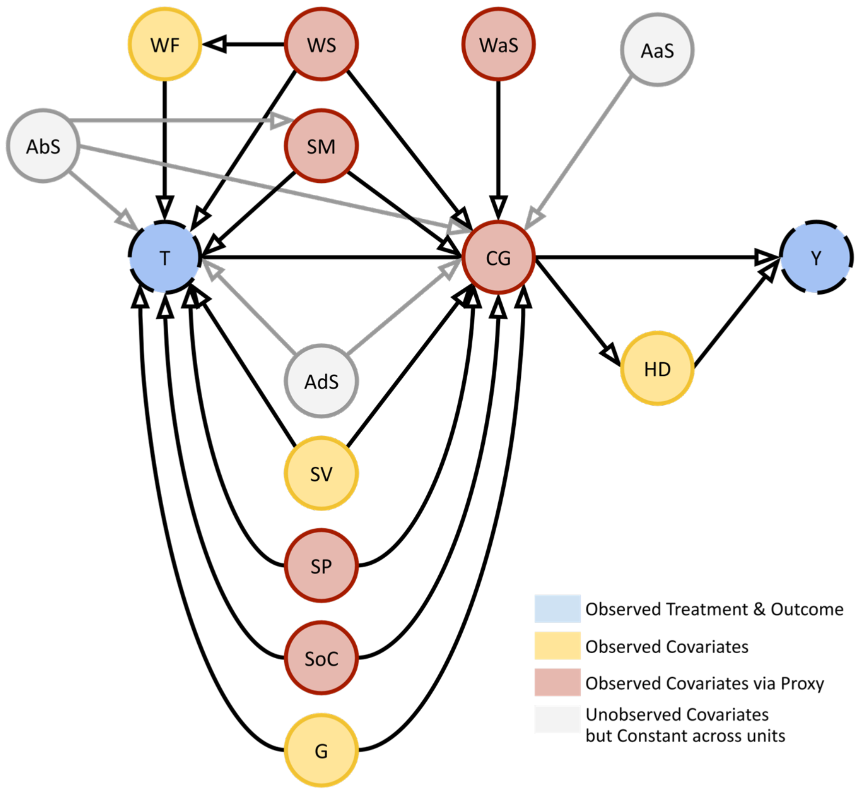
$$\text{ATE} = \mathbb{E}[Y|do(T=1)] - \mathbb{E}[Y|do(T=0)]$$

Our end goal is to account for exactly the variables that will allow us to identify the Average Treatment Effect (ATE) of the treatment on outcome

Unobserved confounding, selection bias, counterfactual yield not observed

Exploit our understanding of the cooperative's modus operandi and harness agricultural knowledge

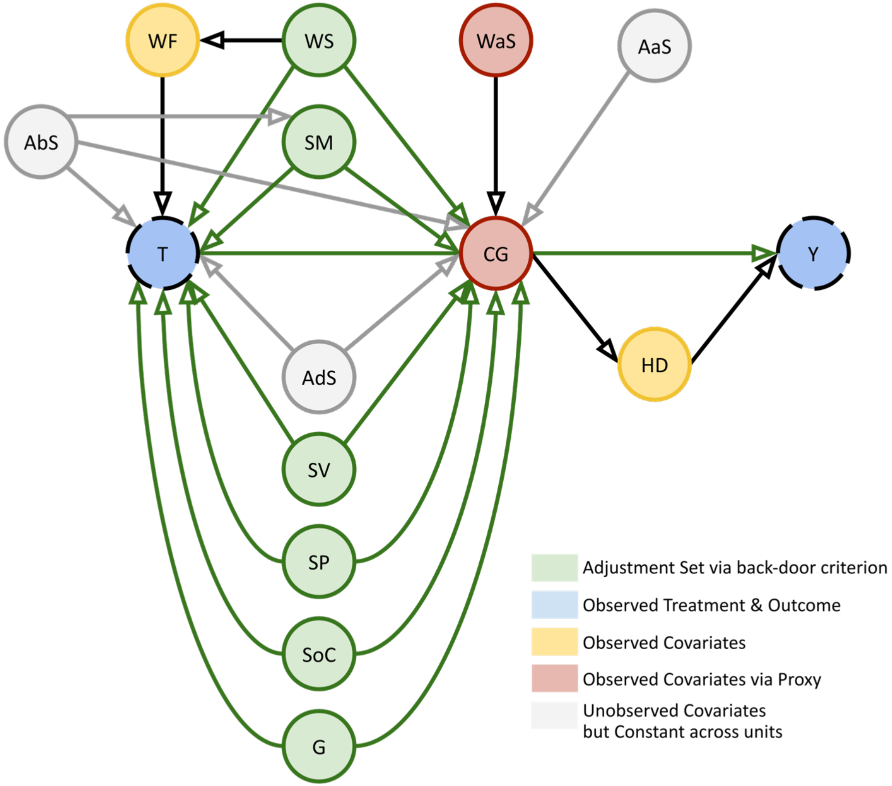
Graph Building



Id	Variable Description	Source
T	Treatment	Recommendation System
WF	Weather forecast	GFS, WRF
WS	Weather on sowing day	Nearest weather station
WaS	Weather after sowing	Nearest weather station
CG	Crop Growth	NDVI via Sentinel-2
SM	Soil Moisture on sowing	NDWI via Sentinel-2
SP	Topsoil physical properties	Map by ESDAC
SoC	Topsoil organic carbon	Map by ESDAC
SV	Seed Variety	Farmers' Cooperative
G	Geometry of field	Farmers' Cooperative
AdS	Practices during sowing	Farmers' Cooperative
AbS	Practices before sowing	Farmers' Cooperative
AaS	Practices after sowing	Farmers' Cooperative
HD	Harvest Date	Farmers' Cooperative
Y	Outcome (Yield)	Farmers' Cooperative

In collaboration with domain experts and by making clear assumptions, we establish a causal graph of the farm system

Effect Identification

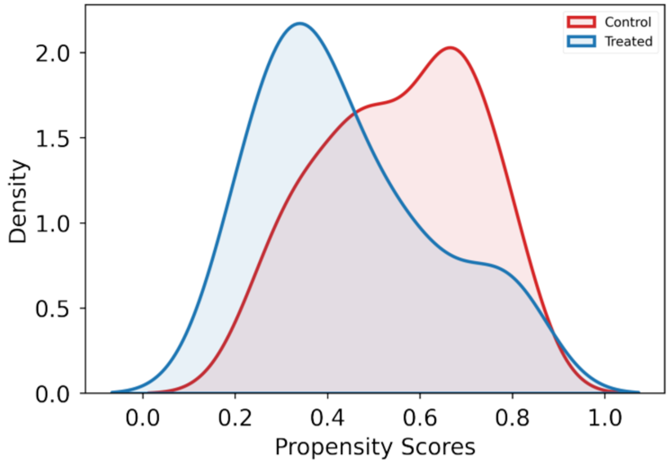


Id	Variable Description	Source
T	Treatment	Recommendation System
WF	Weather forecast	GFS, WRF
WS	Weather on sowing day	Nearest weather station
WaS	Weather after sowing	Nearest weather station
CG	Crop Growth	NDVI via Sentinel-2
SM	Soil Moisture on sowing	NDWI via Sentinel-2
SP	Topsoil physical properties	Map by ESDAC
SoC	Topsoil organic carbon	Map by ESDAC
SV	Seed Variety	Farmers' Cooperative
G	Geometry of field	Farmers' Cooperative
AdS	Practices during sowing	Farmers' Cooperative
AbS	Practices before sowing	Farmers' Cooperative
AaS	Practices after sowing	Farmers' Cooperative
HD	Harvest Date	Farmers' Cooperative
Y	Outcome (Yield)	Farmers' Cooperative

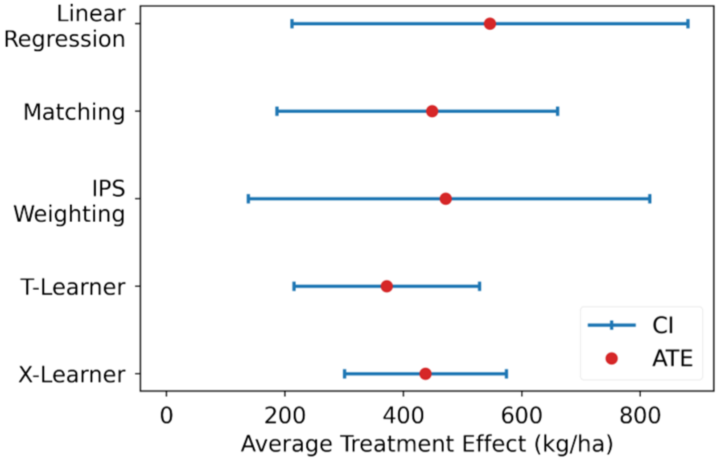
Applying the back-door criterion, the following set of variables was found to be sufficient for effect identification:

$$Z = \{\mathrm{WS}_{\mathrm{MIN,\ MAX}},\ \mathrm{SOC},\ \mathrm{SM},\ \mathrm{G},\ \mathrm{SP}_{\mathrm{SILT,\ CLAY,\ SAND}},\ \mathrm{ABS},\ \mathrm{ADS},\ \mathrm{SV}_{1\text{-}13}\}$$

Effect Estimation



Propensity score P(T=1|Z) distribution and overlap for treatment and control groups



Point ATE estimates and 95% confidence intervals

Results & Refutations

	Causal Effect Estimation			Refutations						
				Placebo		RCC		UCC	RRS	
Method	**ATE**	**CI**	**p-value**	**Effect***	**p-value**	**Effect***	**p-value**	**Effect***	**Effect***	**p-value**
Linear Regression	546	(211, 880)	0.0015	-25.74	0.39	546	0.49	85	543	0.45
Matching	448	(186, 760)	0.0060	50.82	0.39	432	0.40	116	438	0.48
IPS weighting	471	(138, 816)	0.0010	38.82	0.40	470	0.40	113	462	0.45
T-Learner (RF)	372	(215, 528)	0.0240	9.26	0.49	373	0.46	-	353	0.42
X-Learner (RF)	437	(300, 574)	0.0050	5.10	0.50	430	0.37	-	409	0.36

All methods indicate a significant, positive ATE of the treatment on yield

Methods successfully passed 4 refutation tests, indicating robust estimates

Sowing on a recommended day drove a yield increase ranging from 372 to 546 cotton kg/ha (12%-17% relative to mean yield)

Benefits, Limitations and Future Work

Evaluating Digital Tools for Sustainable Agriculture using Causal Inference

Assumptions limit external validity

Driving much-needed tech adoption

To cross-check with experiments

Transparent impact, fair service fees

New pilot applications will allow us to practically test the external validity of our results across different seasons, crops and locations.

Thanks! Questions?

vsito@noa.gr

This work has been supported by the EIFFEL project (EU Horizon 2020 - GA No. 101003518) and the e-shape project (EU Horizon 2020 - GA 820852)