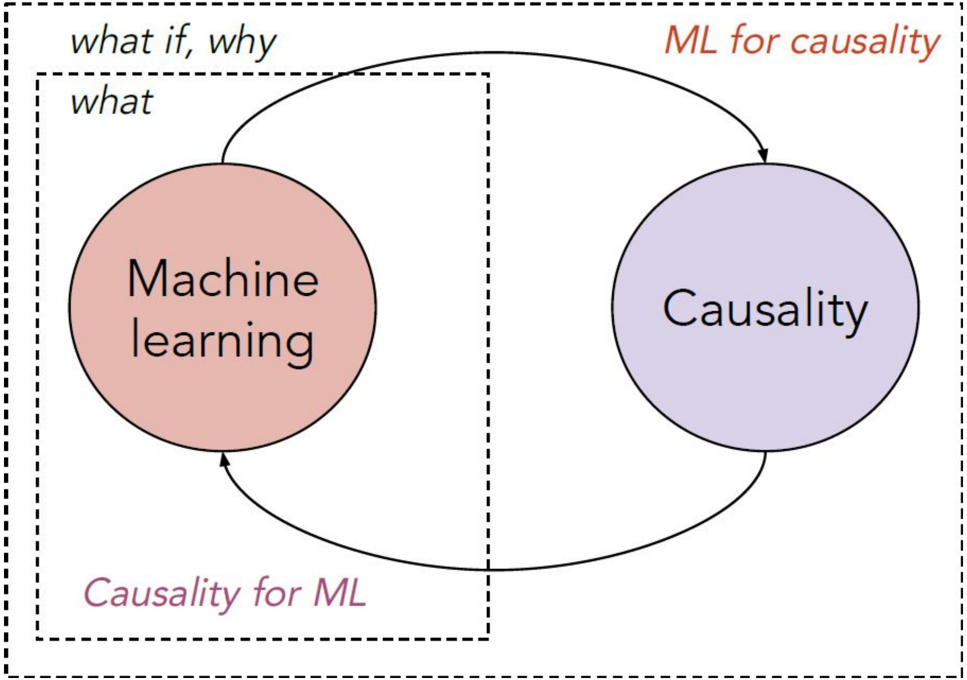
Causal ML for Earth and Environmental Science

Vassilis Sitokonstantinou

(Casual) Causal Course 2025

Causal ML


what if, why
ML for causality
what
Machine learning
Causality
Causality for ML

Why ML for causality

We use ML for estimating (heterogeneous) treatment effects when all potential confounders/controls are observed

- but are either too many (**high-dimensional**) for classical statistical approaches

- or their effect on the treatment and outcome **cannot be satisfactorily modeled by parametric functions**

ML for causality

- **Causal discovery**:

  Identifies the qualitative elements of a system (e.g., relationships between environmental factors, climate, and human activity).

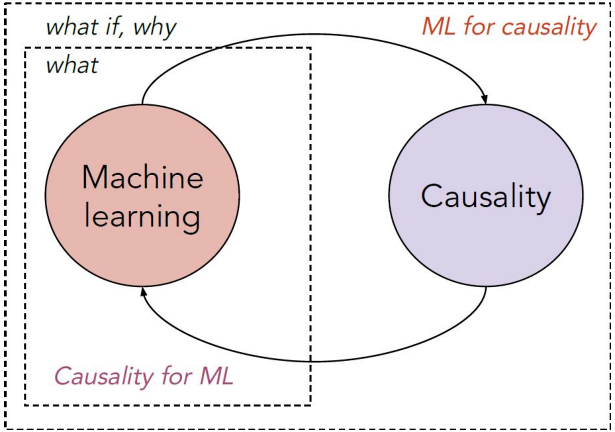
- **Causal effect estimation**:

  Quantifies the impact of one variable (e.g., land management) on another (e.g., productivity), accounting for confounding variables (e.g., climate)

Causality for ML

Predictive tasks: Use the patterns in a training dataset to model associative relationships to predict outcomes in a test dataset.

WHY: Helps improve the robustness and generalization of ML models by focusing on causal relationships rather than mere associations.



Causality and Invariance

Invariance is a desirable property for many ML systems:

- A model that is invariant is one that performs well in new circumstances, particularly when the underlying data distribution changes.
- Invariance also provides a route to some causal insights, even when working only with observational data.

The great lie of machine learning

Learning from examples: Supervised learning involves mapping inputs to outputs using labeled data, aiming to minimize a loss function like mean squared error.

Overfitting: An overfit model is one that has learned the idiosyncrasies (the spurious correlations!) of our dataset.

To avoid overfitting

- Regularization
- Cross-validation

WE ARE RELYING ON A BIG ASSUMPTION

The great lie of machine learning

The IID myth:

The assumption is that the data points are independent and identically distributed (i.i.d.).

By independent, we mean that each data point was generated without reference to any of the others, and by

identically distributed, we mean that the underlying distributions in the data generating process are the same for all the data points.

Real world: ‘Eh..No!’

Dangers of spurious correlations

When we train a ML system with the i.i.d. assumption, we are:

implicitly assuming an underlying data generating process for that data. This data generating process defines an **environment**.

Different data generating processes will result in different environments, with different underlying distributions of features and targets.

When we train in one environment and apply in another…the model performs poorly.

Dangers of spurious correlations

(A) **Cow: 0.99**, Pasture: 0.99, Grass: 0.99, No Person: 0.98, Mammal: 0.98

(B) No Person: 0.99, Water: 0.98, Beach: 0.97, Outdoors: 0.97, Seashore: 0.97

(C) No Person: 0.97, **Mammal: 0.96**, Water: 0.94, Beach: 0.94, Two: 0.94

Figure from Recognition in Terra Incognita, where annotations were provided by ClarifAI.com.

When is a correlation spurious?

What is a genuine and what is a spurious correlation: <u>It depends</u>

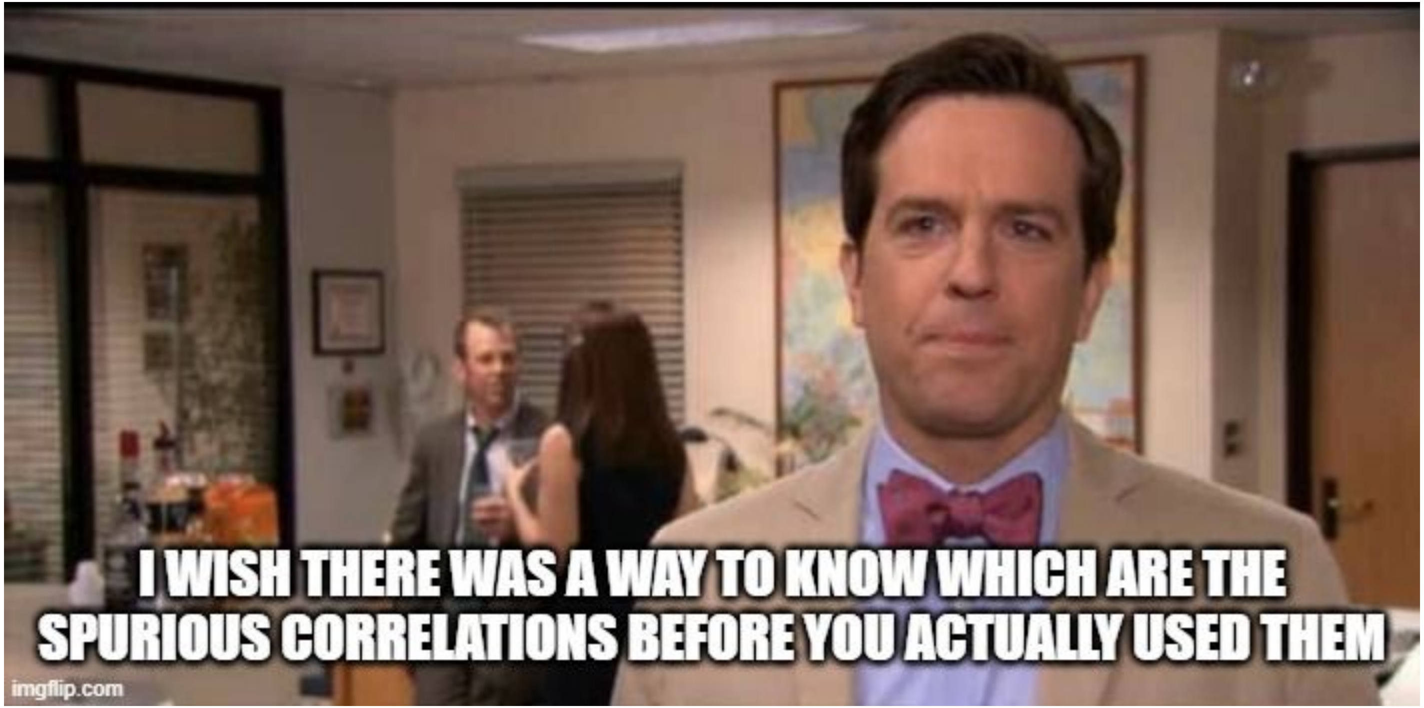
If we want to use the model only in one environment. **USE EVERYTHING**

If we want to use the model on data outside of the training environment: then <u>any correlation that only holds in the training environment is spurious</u>.

A spurious correlation is a correlation that only appears to be true due to a selection effect (such as selecting a training set!).

In the strictest interpretation, **any correlation that does not arise from direct causation could be considered spurious**.

When is a correlation spurious?



Invariance

To be confident of our predictions outside of our training and testing datasets:

- We need a model that is robust to distributional shifts
- Such a model would have learned a representation which ignores dataset-specific correlations and relies on features that affect the target in all environments.

Solution:

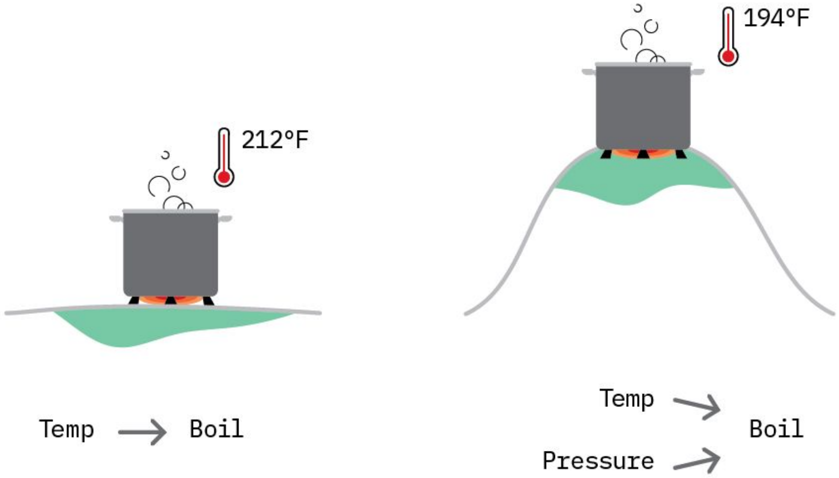
Train with multiple environments - can only interpolate among environments

Invariant models - extrapolate to unseen environments

Material retrieved from https://ff13.fastforwardlabs.com/

Invariance

- Causal relationships are by their nature invariant
- We establish causal relationships by observing them in multiple settings



Invariance

Causality gives us a precise mathematical language to describe invariance:

Environments are defined by interventions in the causal graph.

Each intervention changes the data generating process: correlations between variables in the graph may be different.

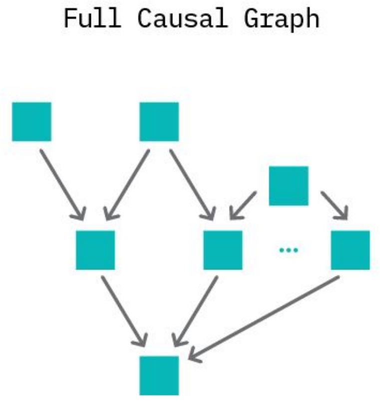
Direct causal relationships are invariant relationships regardless of any interventions.

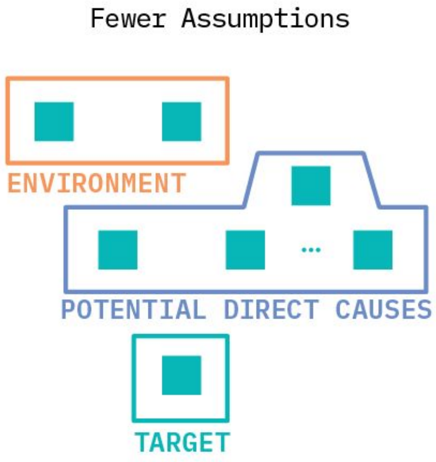
It may be that an intervention restricts the values that the causal variables take, but the relationship itself is not changed.

Changing the arguments to a function does not change the function itself.

Invariant causal prediction

Invariant causal prediction (ICP) addresses the task of invariant prediction explicitly in the framework of structural causal models.





Invariant causal prediction

We fit a model in multiple environments and monitor which features are consistently predictive.

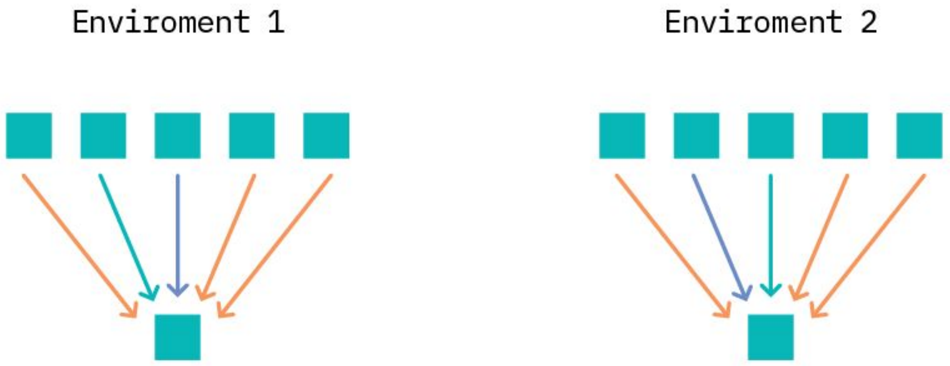
Algorithm:

We test subsets of features by fitting models to the target across environments.

If the model remains unchanged the subset is considered invariant.

We repeat this combinatorially.

Select frequent features.



Invariant Risk Minimization

What happens when we can't or don't want to think with an SCM.

"If both Newton's apple and the planets obey the same equations, chances are that gravitation is a thing." – IRM authors

IRM focusses on out-of-distribution generalization: the performance of a predictive model when faced with a new environment.

"To learn invariances across environments, find a data representation such that the optimal classifier on top of that representation matches for all environments." – IRM authors

Invariant Risk Minimization

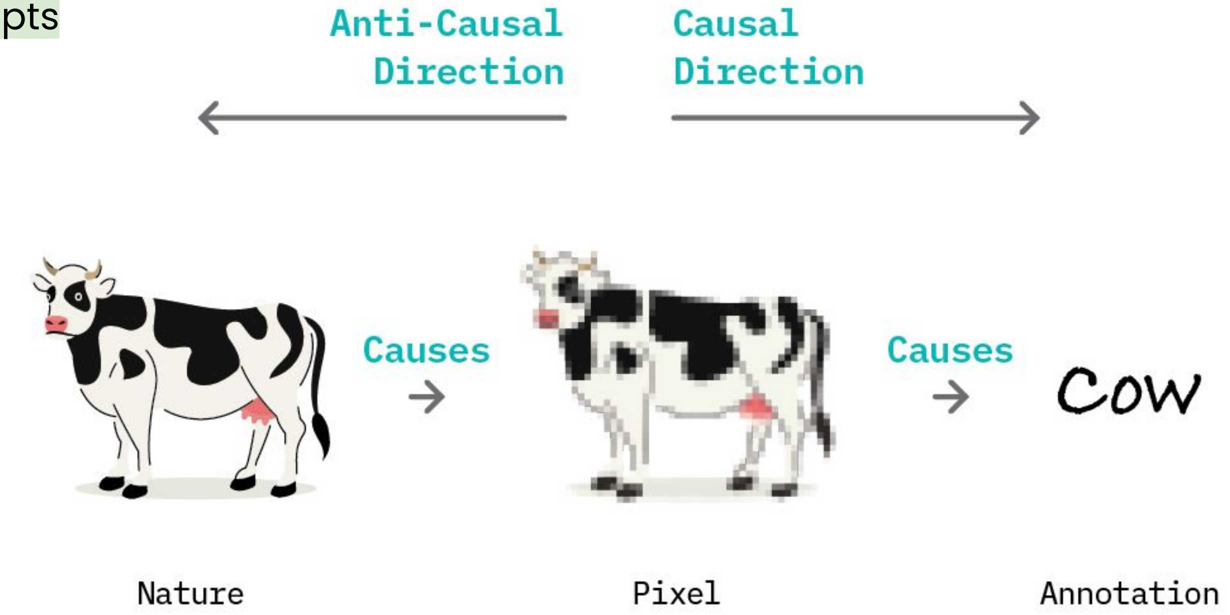
Said differently,

The idea is that there is a latent causal structure behind the problem we’re learning, and the task is to recover a representation that encodes the part of that structure that affects the target.

The causal direction

This is the view taken by IRM, which interprets supervised learning from images as being a causal (rather than anti-causal) problem.

From low features to high level concepts



Not all supervised problems are causal

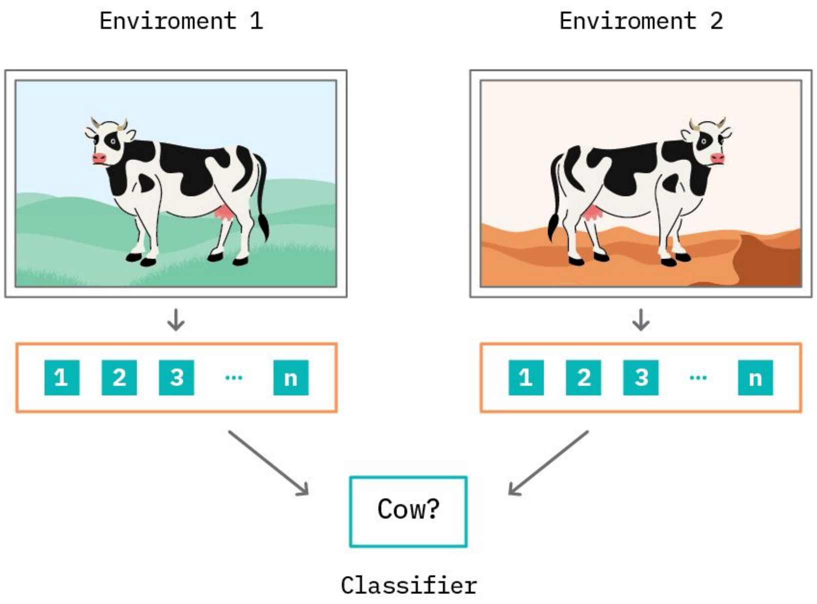
The causal direction

Learning in the causal direction explains some of the success of supervised learning:

there is a chance that it can recover invariant representations without modification.

The modifications that IRM makes to the learning procedure improve the chance by <u>specifically promoting invariance</u>.

IRM



In the IRM setup, we feed the algorithm data from multiple environments, and we must be explicit about which environment a data point belongs to.

Simply providing data from multiple environments is not enough

IRM

The problem of learning the *optimal classifier in multiple environments* is a bi-level constrained optimization problem:

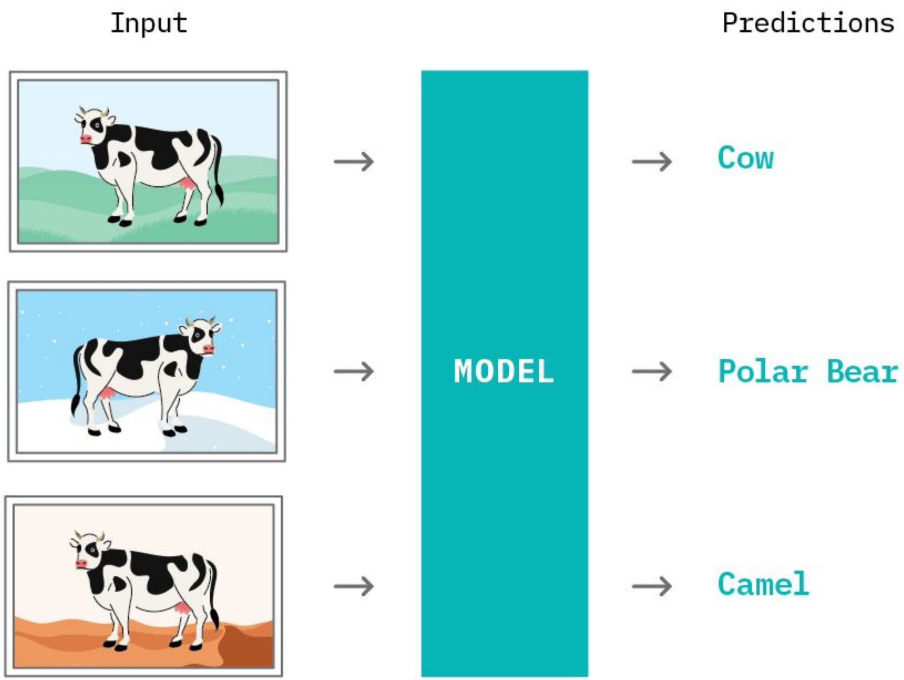
> in which we must simultaneously <u>find the optimal data representation and optimal classifier</u> across multiple separate datasets.

<u>IRM reduces the problem to a single optimization loop</u>, with the trick of using a constant classifier and introducing a new penalty term to the loss

IRM loss = sum over environments (error + penalty)

<u>The error</u> is the usual error (e.g., cross entropy) - calculated on each environment.
<u>The new penalty term</u> measures how much the performance could be improved in each environment with one gradient step - **punish high gradients**

IRM



If we rely on empirical risk minimization, we learn spurious correlations between animals and their environments

Considerations

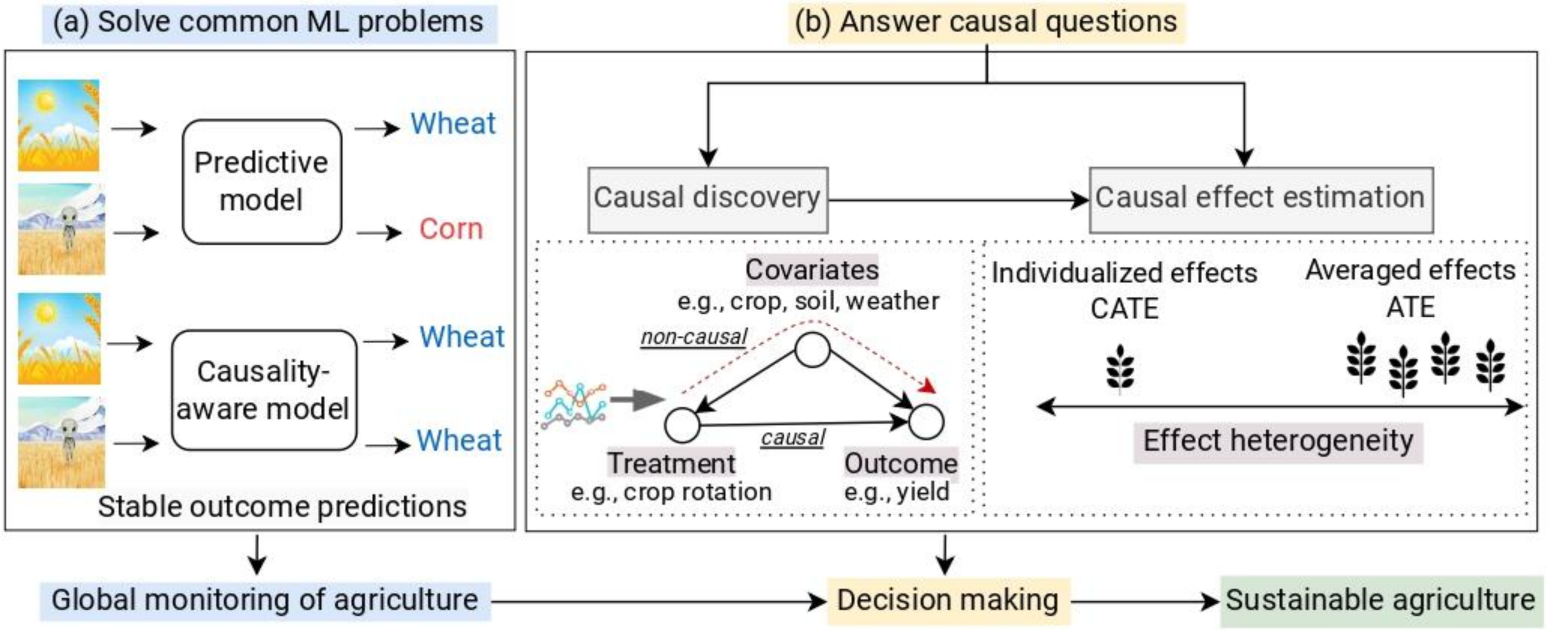
IRM buys us extrapolation powers to new datasets

Access to sufficiently diverse environments - not to learn spurious correlations as invariant

Even when a well-labeled dataset exists, it is seldom accompanied by detailed metadata. **No information on the environment**

The more environments we have and the more diverse they are the better

Causal ML - two perspectives



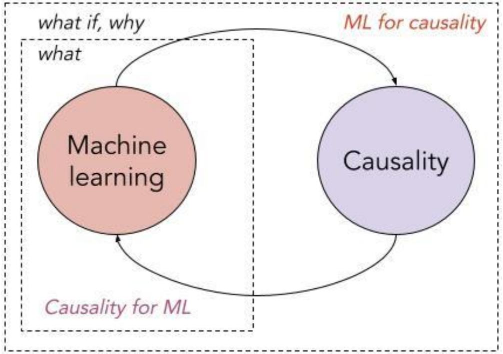
Beyond predictions

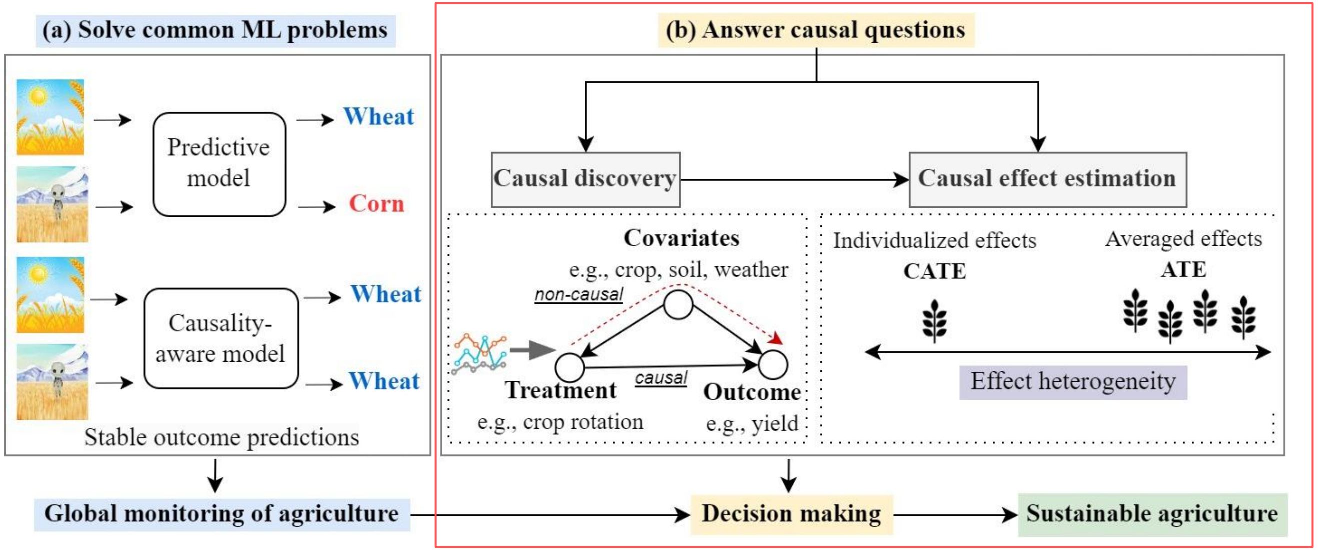
Problems:

Predictive ML struggles with generalization across the globe

Predictive ML cannot answer causal questions

Causal ML





Sitokonstantinou. V., et al., (2024). Causal machine learning for sustainable agroecosystems: Under review

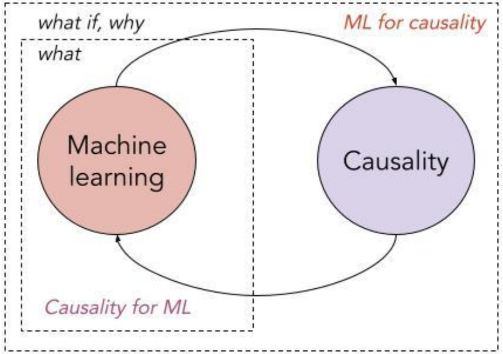
Beyond predictions

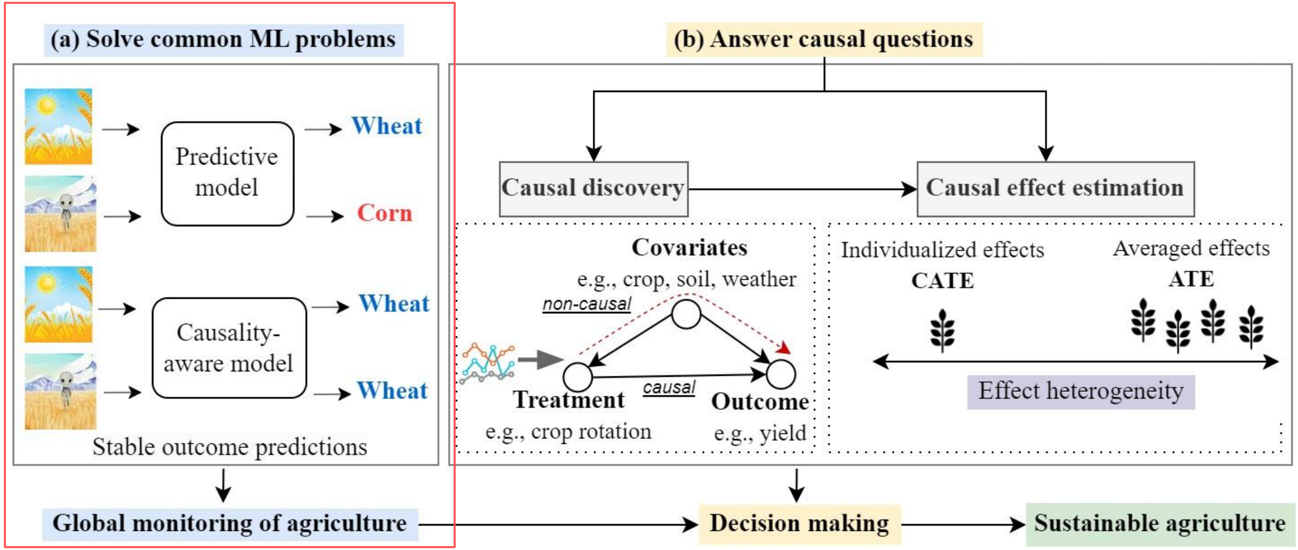
Problems:

Predictive ML struggles with generalization across the globe

Predictive ML cannot answer causal questions

Causal ML





Sitokonstantinou. V., et al., (2024). Causal machine learning for sustainable agroecosystems: Under review

Questions we can answer with causal inference

Causal discovery

<u>What is the relationship between X and Y?</u> Does ENSO cause soil moisture anomalies in Southern and Eastern Africa?

<u>What are the drivers of complex systems?</u> Which are the primary causal factors driving cropland expansion in the Amazon across varying socio-economic contexts? Are there contrasting mechanisms in the food security system for different districts in Somalia?

Questions we can answer with causal inference

Causal effect

What is the impact of climate change? How do increased temperatures impact agricultural productivity?

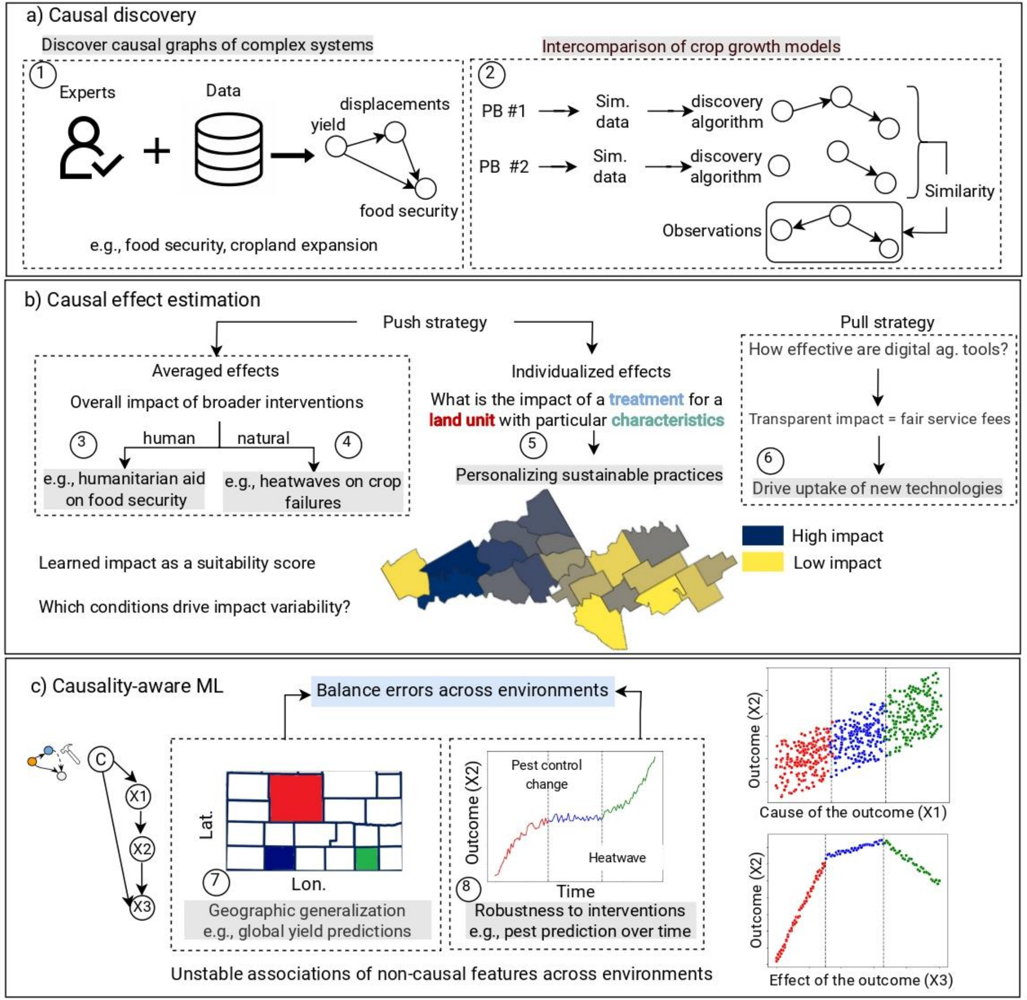
What is the impact of human interventions? , What is the impact of humanitarian aid on food insecurity in Africa?

What is the impact of extreme weather events? , Do dry spells drive crop failures in Europe?

What is the impact of agricultural practices? Which regions benefit more from organic farming and what drives the differences in effects? How will the effect of practices be affected in the future under different climatic conditions?

Is a digital farm recommendation tool effective? Will sowing on the recommended day increase my yield? What would be the benefit of investing in this digital tool?

Applications



ML for causality (push strategy)

Geospatial personalization of agricultural practices

Causal inference
What would happen

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

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)

Diverse policy measures (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"

Geospatial personalization of agricultural practices

EARTH VISION

CVPR JUNE 19-24 2022 NEW ORLEANS • LOUISIANA

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

NEURAL INFORMATION PROCESSING SYSTEMS

Climate Change AI

Our approach

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

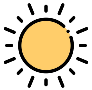
Train causal models estimating the impact of agricultural practices on metrics of interest

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)

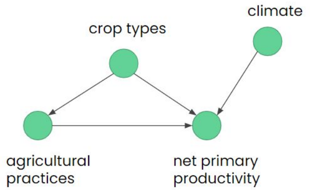
- Flexible framework for CATE estimation
- Robust for spatial data[1]

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

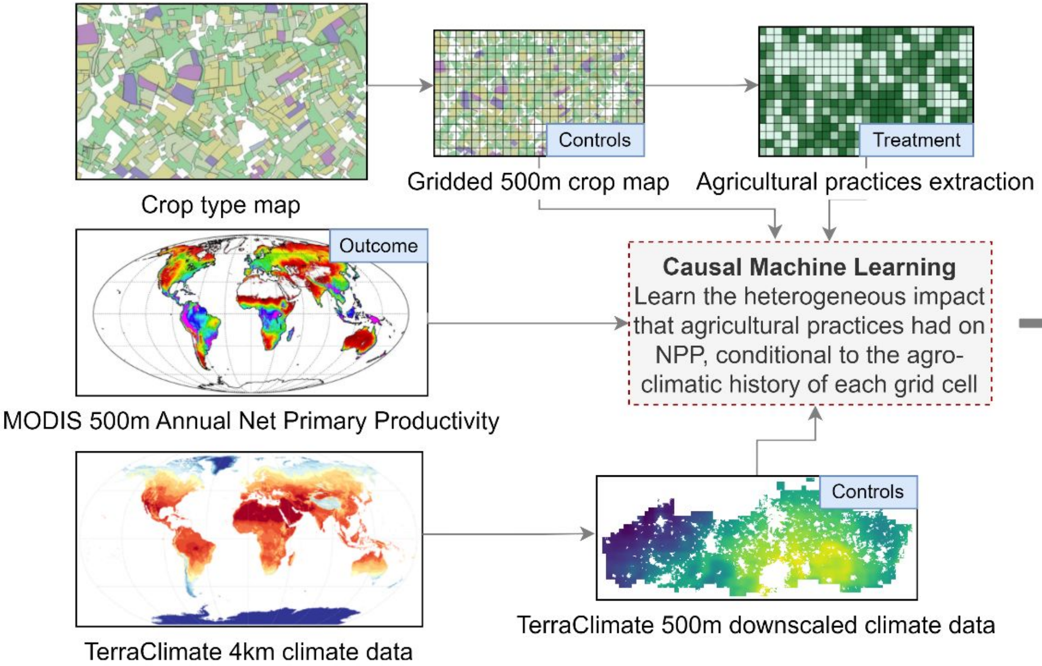
$$\hat{\theta} = \underset{\theta \in \Theta}{\arg\min}\, \mathbb{E}\left[(\tilde{Y} - \theta(X) \cdot \tilde{T})^2\right]$$

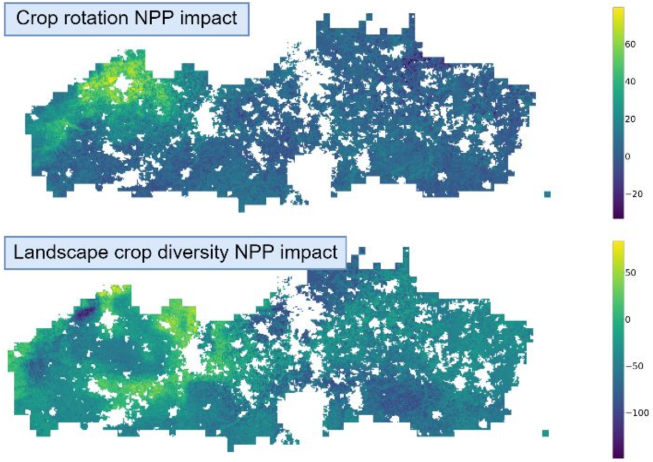
[1] *Approaches to spatial confounding in geostatistics*, Gilbert et al., 2022

Methodology & Results
(Flanders, Belgium, 2010-2020)



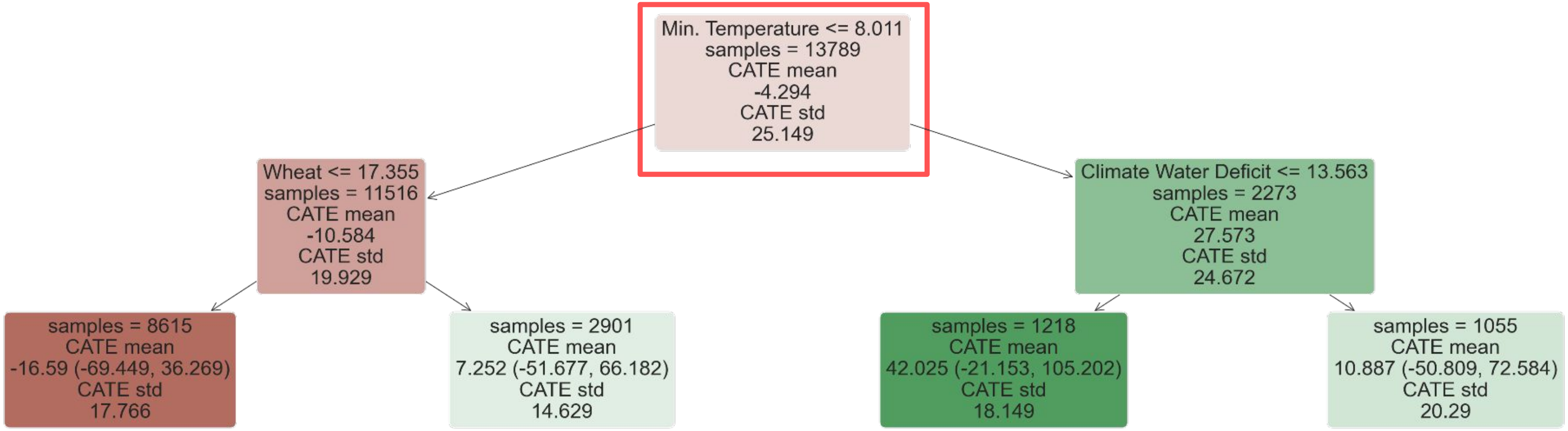
CR ATE: 1.08 (95% CI [-20.35, 22.51])   LCD ATE: -35.73 (95% CI [-58.73, -12.73])





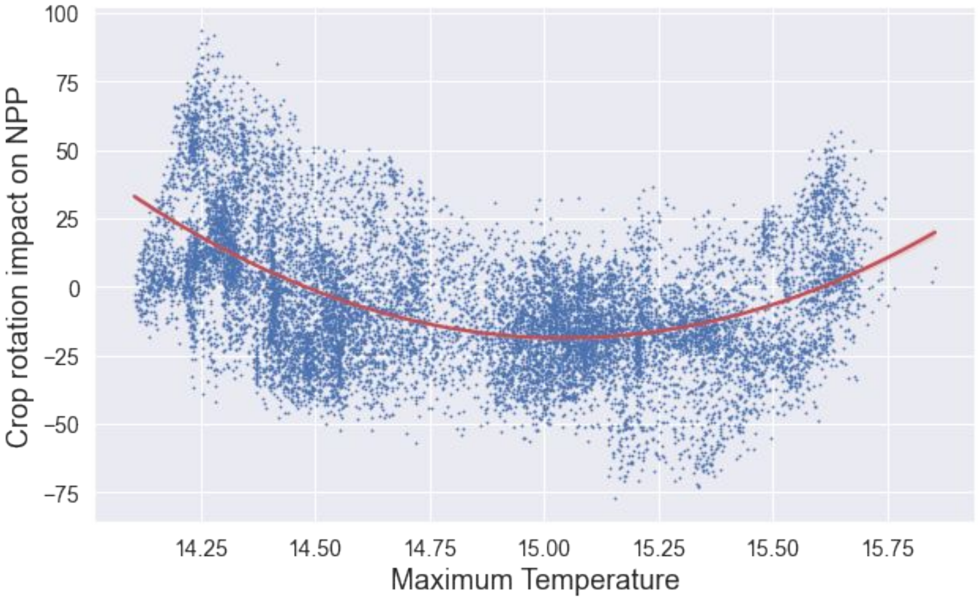
1. The learned impact as an EO-based Land Suitability Score
2. Understand agro-climatic conditions driving impact variability
3. Devise efficient agricultural policies against Climate Change

Interpretability trees for learned CATEs*



*Environmental conditions driving effect heterogeneity for Crop Rotation practice

CATE exploration enables further hypotheses

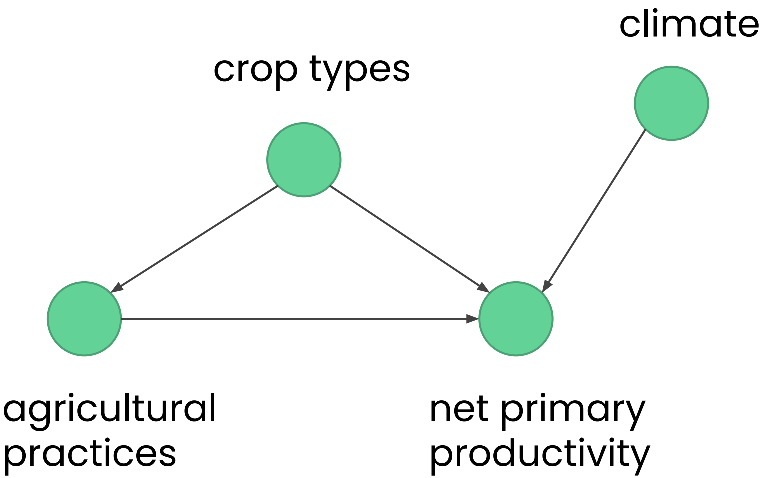


*In a warmer planet, crop rotation might be more beneficial for productivity

*Using future climate projections, how do CATE results change?

*Caution with such conclusions

Limitations & Future Work

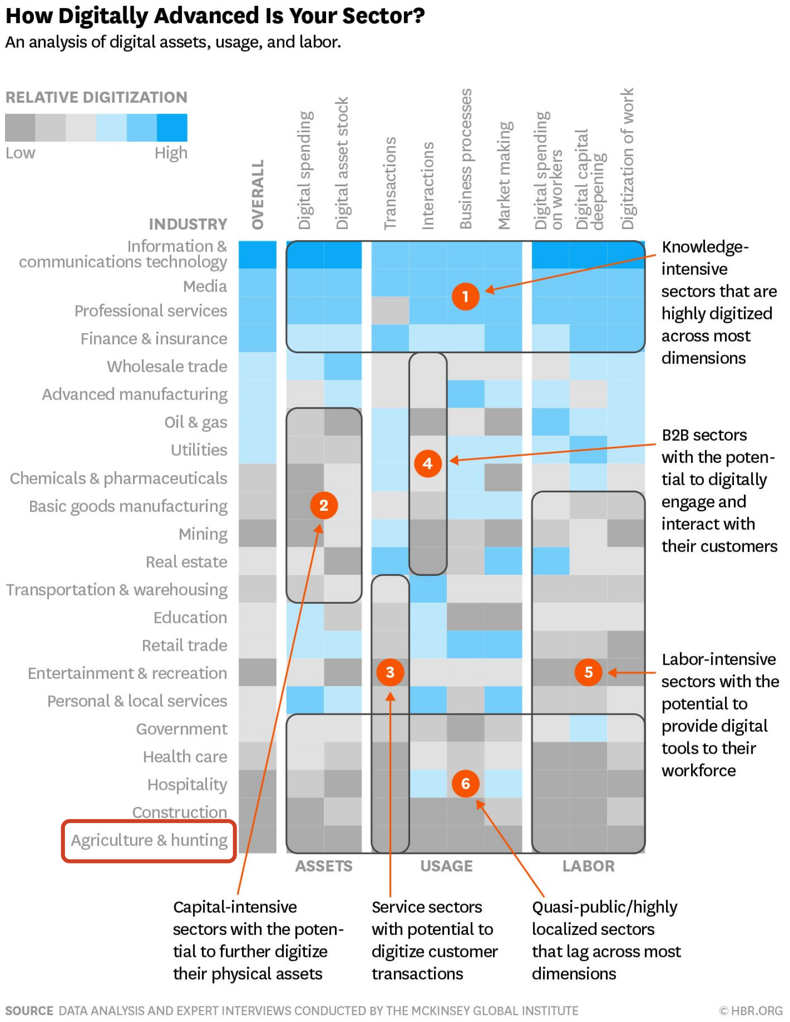


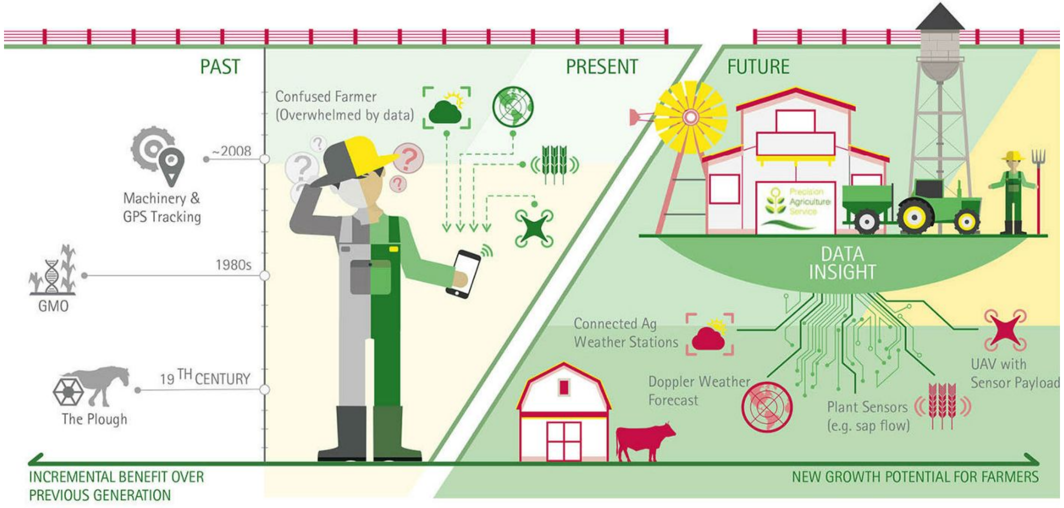
The causal graph assumed is simplistic, some bias and variance in estimates remains:

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





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

Trustworthy climate-smart digital tools

Causal inference
What would happen

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

Causal inference for sustainable agriculture

Trustworthy climate-smart digital tools

Evaluating Digital Agriculture Recommendations with 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
{i.tsoumas, giannarakis, vsito, akoukos, nbartsotas, kontoes}@noa.gr
{ilias.tsoumas, ioannis.athanasiadis}@wur.nl
{dimitra.loka}@elgo.gr

AAAI 2023, AI for Social Impact Track

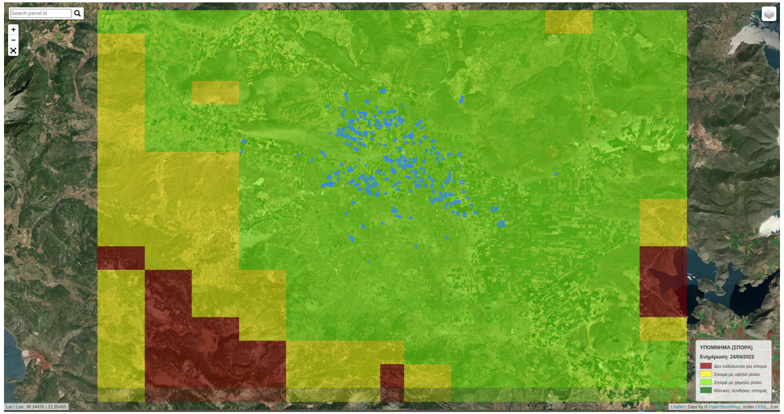
Is today a good day to sow?

Knowledge-based recommendation system for optimal cotton sowing

Collaboration with a farmer’s cooperative (171 cotton fields) in Greece

Common practices, homogeneous fertilizer application, and jointly owned machinery.

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



Trustworthy climate-smart digital tools

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

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

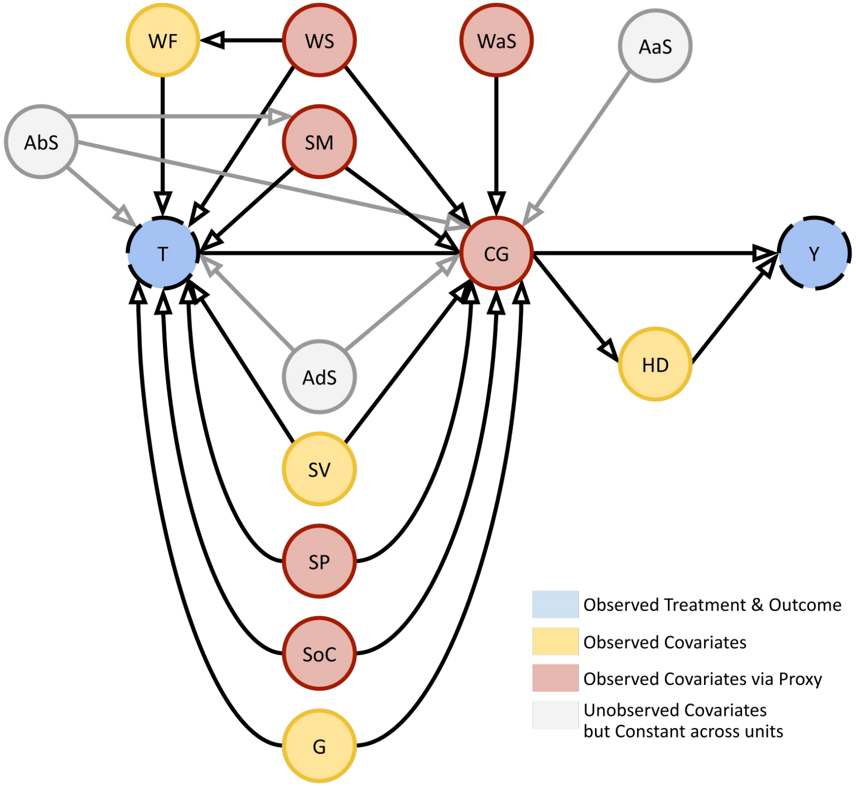
$$\text{ATE} = \mathbb{E}\big[Y|do(T=1)\big] - \mathbb{E}\big[Y|do(T=0)\big]$$

Account for exactly the variables that will allow us to identify the Average Treatment Effect (ATE) of the treatment on the outcome

Unobserved confounding, selection bias, counterfactual yield not observed

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

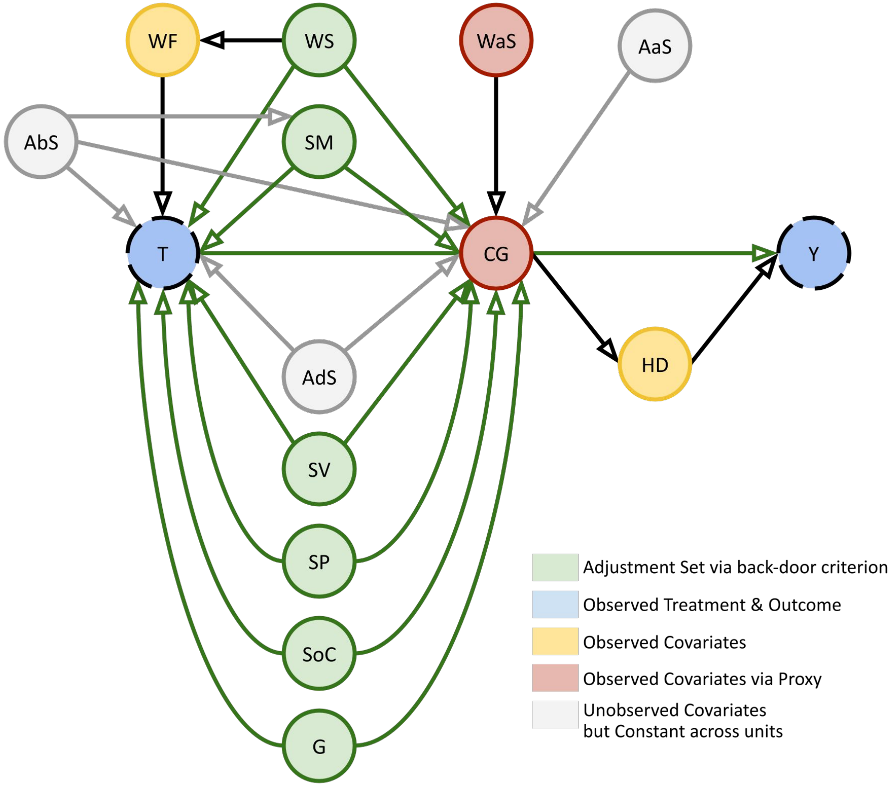
Graph Building



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



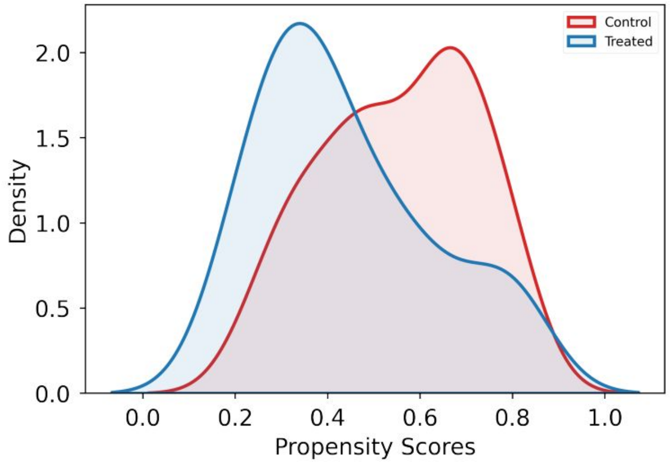
Id	Variable Description	Source
T	Treatment	Recommendation System
WF	Weather forecast	GFS, WRF
WS	Weather on sowing day	Nearest weather station
WaS	Weather after sowing	Nearest weather station
CG	Crop Growth	NDVI via Sentinel-2
SM	Soil Moisture on sowing	NDWI via Sentinel-2
SP	Topsoil physical properties	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:

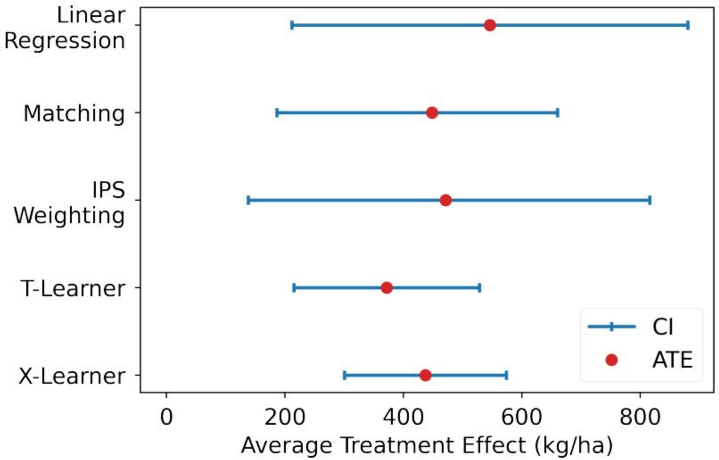
$$Z = \{\mathrm{WS}_{\mathrm{MIN,\ MAX}},\ \mathrm{SOC},\ \mathrm{SM},\ \mathrm{G},\ \mathrm{SP}_{\mathrm{SILT,\ CLAY,\ SAND}},\ \mathrm{ABS},\ \mathrm{ADS},\ \mathrm{SV}_{1\text{-}13}\}$$

Effect Estimation

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

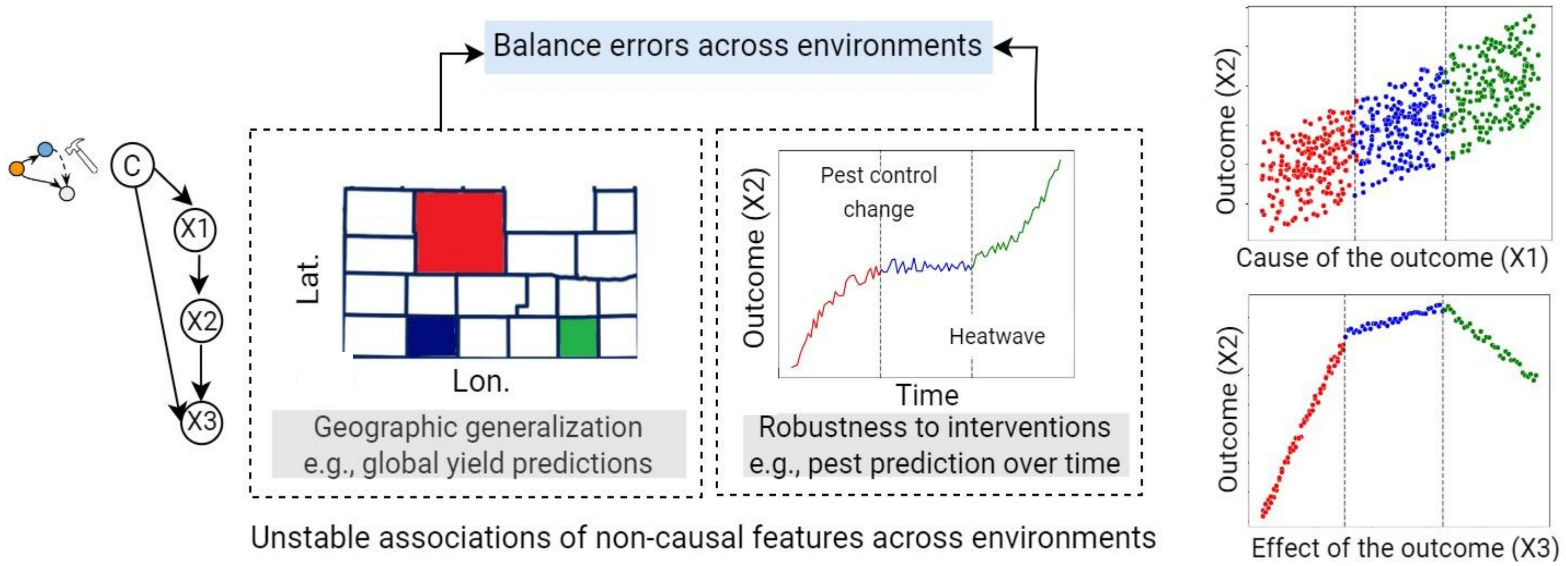


Propensity score P(T=1|Z) distribution and overlap for treatment and control groups



Point ATE estimates and 95% confidence intervals

Causality for ML



Challenges and opportunities

- Domain scientists hold important knowledge that can be unlocked - translated to causal graphs.
- Defining the causal questions with precision: causal inference is always translated to a causal conclusion - we need to be honest and precise
- We are as good as our data: multiple sources, different quality, different scale. Most important: faithfulness to the concept in the graph.
- How do we aggregate data: this will define the causal question we are answering that is possibly not the one we were going for.
- Selection biases, spillover effects etc. - spatial causal inference is a beast
- How do we validate? There is no ground truth for observational causal inference? Refutation tests? Sensitivity analysis? Corroboration with literature (how do we learn something new?)

References

- Sitokonstantinou, V., et al., (2024). Causal machine learning for sustainable agroecosystems: Under review at *Nature Food*

- Sitokonstantinou, V., Koukos, A., Drivas, T., Kontoes, C., and Karathanassi, V. (2022). DataCAP: A Satellite Datacube and Crowdsourced Street-Level Images for the Monitoring of the Common Agricultural Policy. In International Conference on Multimedia Modeling (pp. 473-478). Springer.

- Tsardanidis, I., Koukos, A., Sitokonstantinou, V., Drivas, T., & Kontoes, C. (2024). Cloud gap-filling with deep learning for improved grassland monitoring. *arXiv preprint arXiv:2403.09554*. Under review at Computers and Electronics in Agriculture.

- Rousi, M.*, Sitokonstantinou, V.*, Meditskos, G., Papoutsis, I., Gialampoukidis, I., Koukos, A., ... and Kompatsiaris, I. (2020).Semantically enriched crop type classification and linked earth observation data to support the common agricultural policy monitoring. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14, 529-552.

- Giannarakis, G., Sitokonstantinou, V., Roxanne, L. and Kontoes, C. (2022). Towards assessing agricultural land suitability using causal machine learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW).

- Jo, HW.*, Koukos, A.*, Sitokonstantinou V., Lee, WK. & Kontoes, C. (2022). Towards Global Crop Maps with Transfer Learning. In NeurIPS Workshop Climate Change AI.

- Tsoumas, I.*, Giannarakis, G.*, Sitokonstantinou, V., Koukos, A., Loka, D., Bartsotas, N., ... Athanasiadis, I. (2022). Evaluating Digital Agriculture Recommendations with Causal Inference. In Thirty-Seventh AAAI conference on artificial intelligence. 2023.