# Causal Machine Learning for Sustainable Agriculture

**Vasileios Sitokonstantinou**[1] vassilis.sitokonstantinou@wur.nl
**Emiliano Díaz Salas-Porras**[2] emiliano.diaz@uv.es
**Jordi Cerdà-Bautista**[2] jordi.cerda@uv.es
**Maria Piles**[2] maria.piles@uv.es
**Ioannis Athanasiadis**[1] ioannis.athanasiadis@wur.nl
**Ilias Tsoumas**[1,3] i.tsoumas@noa.gr
**Gustau Camps-Valls**[2] gustau.camps@uv.es

[1]Artificial Intelligence Group, Wageningen University & Research, The Netherlands
[2]Image Processing Laboratory, Universitat de València, Spain
[3]Beyond Center, National Observatory of Athens, Greece

## Abstract

Agricultural systems involve biological, physical, social and economic dimensions. Current farming practices often fail to balance these and contribute to soil degradation, biodiversity loss and greenhouse gas emissions. Understanding and managing these complex systems requires data-driven approaches that can monitor and anticipate agricultural outcomes. Machine learning (ML) has been used to predict outcomes such as crop yields, soil health or water use from observational data. However, ML often cannot explain how these outcomes are affected by human or natural interventions, nor can it reliably generalize across regions and management contexts. Causal machine learning (causal ML) combines causal inference with ML to address these limitations. It enables estimation of causal effects for agricultural questions and improves prediction by focusing on stable and generalizable features. In this perspective, we introduce methods for causal inference from observational data and approaches that embed causal knowledge into predictive models. We then outline applications in research, policy and practice and conclude with key challenges and future directions for sustainable agriculture.

## 1 Introduction

Agriculture is a complex system where biological, physical, social and economic factors interact. Farms are not isolated production units but part of broader systems that include climate, land use, ecosystems and human activities [8]. Current practices often fail to balance these elements, contributing to environmental pressures such as biodiversity loss, soil degradation and greenhouse gas emissions [30, 13]. Building resilient food systems while reducing these impacts requires tools that can untangle these complex interactions and allow us to intervene effectively. Machine learning (ML) is excellent at identifying patterns and making predictions in agricultural systems [21]. For example, ML can forecast crop yields from historical observations and environmental measurements. However, predictive ML alone cannot explain why outcomes occur and may fail when applied to new regions or management scenarios [41]. Without causal reasoning, ML can produce statistically accurate but misleading insights for sustainability-oriented decision-making. Causal machine learning (causal ML) addresses these limitations (see Sec. A.1: predictive vs causal ML) in two complementary ways (see Fig. 1): (i) by answering causal questions, such as “Does El Niño Southern Oscillation (ENSO) cause soil moisture anomalies in Southern and Eastern Africa?” or “How does crop rotation affect productivity across different local conditions?” and (ii) by improving predictive models’ robustness and generalizability through stable causal features [15]. Despite its promise, adoption in agriculture

Accepted to the NeurIPS 2025 Workshop on CauScien: Uncovering Causality in Science.

remains limited compared to ecology [37] and climate sciences [39]. This perspective paper aims to make causal ML concepts accessible for stakeholders and scientists in agriculture, in an attempt to promote their wider adoption in the domain.

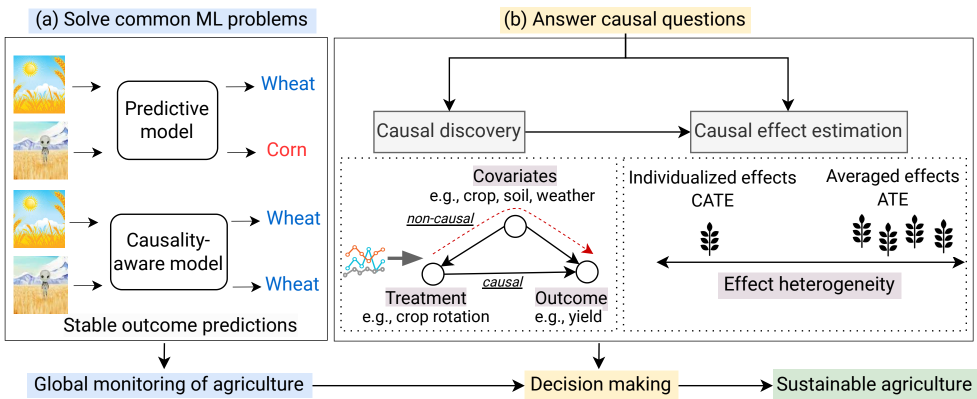


Figure 1: **Causality in agriculture:** (a) Improve predictive models by enhancing robustness and generalization across interventions and geographies. (b) Answer causal questions: causal discovery to construct causal graphs and causal effect estimation to guide effective interventions.

## 2 Causal Machine Learning in Agriculture

### 2.1 ML For Answering Causal Questions

Randomized experiments remain the gold standard for causal inference in agriculture, providing unbiased effect estimates by manipulating treatments and controlling variables [29]. Yet, they have notable limitations: small sample sizes often fail to capture environmental heterogeneity, results may not generalize to new regions or scales [34] and many agricultural processes involve complex interactions that a single intervention cannot isolate [16]. Observational causal inference offers a practical alternative. Large datasets from farm records, surveys or remote sensing allow researchers to study interventions at broader scales and under realistic conditions. Advances in data and machine learning for representation learning help extract informative features from high-dimensional inputs like satellite imagery [19]. However, confounding, where external factors influence both treatment and outcome, remains a key challenge, and ML-learned features may not correspond to variables with clear physical or agronomic meaning. Embedding these analyses within rigorous causal frameworks is therefore essential for producing valid, actionable insights. To exploit ML's potential while addressing these issues, we present a causal ML workflow for agriculture:

**Defining the causal question.** A causal question can be either qualitative or quantitative. For instance, a qualitative question might explore whether there is a causal relationship between soil microbiome diversity and crop yield. Such questions are addressed using *causal discovery* methods. On the other hand, a quantitative question might seek to determine the extent of the impact, such as estimating how much the use of a particular pesticide increased the crop yield. These types of questions are tackled using *causal effect estimation* methods.

**Collecting data.** To answer causal questions in agricultural systems, data on relevant variables must be collected. For causal discovery, this means measuring the variables of interest and potential confounders, factors that influence multiple variables and may create spurious associations. For causal effect estimation, information is needed on the treatment (e.g., fertilizer use), the outcome (e.g., farm profitability) for specific units (e.g., fields) and confounders such as temperature, soil or crop characteristics [27]. Confounders can be identified through causal discovery algorithms, statistical analyses (e.g., correlations suggesting hidden links) and expert agronomic knowledge. Relevant data sources include remote sensing, reanalysis products, process-based model outputs, farm management systems, crop calendars, accounting records, field experiments and plant phenotyping platforms.

**Making assumptions.** To answer causal questions without randomized experiments, certain assumptions are required, structural, process or statistical. *Structural assumptions* concern the existence of unmeasured confounding or the time lag of cause-effect relationships. *Process assumptions* address whether chosen models (e.g., linear) are adequate for capturing causal dynamics. *Statistical assumptions* involve whether the available data are sufficient and representative. For causal discovery, the core requirement is *causal sufficiency*, meaning that all relevant variables are measured. For causal effect estimation, the key requirement is a valid *causal graph*, which specifies cause-effect relationships and can be informed by expert knowledge or data-driven methods. Rubin's potential outcomes framework [35] complements Pearl's structural causal models [31] by detailing the assumptions needed for valid causal inference.

**Selecting causal ML methods.** The choice of method depends on the causal question, available data and assumptions. For *causal discovery*, the goal is to uncover cause-effect relationships. Constraint-based approaches (e.g., PC, FCI) [43] and score-based approaches (e.g., GES) [6] require relatively weak assumptions but may not recover the full graph. Variants exist for both cross-sectional [12] and time-series data [3, 40]. Asymmetry methods [48] assume specific functional forms of relationships to improve recovery. For *causal effect estimation*, the focus is on quantifying the impact of interventions. The Average Treatment Effect (ATE) captures the overall impact of a treatment (e.g., *'What is the average effect of organic farming across Europe?'*), typically estimated via matching [36] or propensity scores [32]. The Conditional Average Treatment Effect (CATE) refines this for subgroups (e.g., *'How does the impact of organic farming vary across countries in Europe?'*), enabling tailored insights, with methods such as the X-learner [18] and Double Machine Learning (DML) [5].

**Checking robustness.** When relying on observational rather than experimental data to answer causal questions, assumptions must be made about the data and the process that generated them. Robustness checks help stress-test these assumptions. For *causal discovery*, this involves comparing inferred relationships with domain knowledge or testing methods on synthetic datasets generated by process-based models with known causal graphs. If the method correctly recovers the model's structure, it increases confidence in its applicability to the real system. For *causal effect estimation*, refutation tests [42] assess whether assumptions yield coherent estimates. A common approach is to add noise variables as spurious confounders and verify that estimated effects remain stable.

### 2.2 Using Causality to Improve ML Predictions

Predictive ML models in agriculture often fail to generalize across space, time or under interventions, as they rely on correlations that may not hold outside the training data. Causality-aware ML improves predictive performance by emphasizing features with stable, direct causal relationships to the outcome [28]. Models trained on these causal features are more robust, interpretable and generalizable. Invariant Causal Prediction (ICP) [28] identifies features whose influence remains consistent across environments, while Anchor Regression [33] balances predictive accuracy and stability by weighting non-causal features without discarding them entirely. Other approaches, including Invariant Risk Minimization (IRM) [2] and Risk Extrapolation (REx) [17], similarly leverage causal principles to enhance robustness. Traditional domain adaptation, domain generalization and transfer learning methods [47] focus on improving predictive performance in new domains. Causality-aware ML complements these by explicitly identifying which relationships are invariant under interventions and which are susceptible to change, providing a principled framework for reasoning under distribution shifts and designing models that remain reliable in diverse agricultural contexts.

## 3 Applications of Causal ML in Agriculture

Causal ML is about answering causal queries, avoiding being right for the wrong reasons. Next, we present several applications to inspire how causal ML can drive sustainability in agriculture.

**Advancing Science** *(i) Understanding complex and dynamic systems:* Systems like food security and land use change are shaped by complex socio-environmental drivers that evolve across space and time. For example, in the case of food security, environmental stressors (e.g., droughts) interact with local socio-economic factors (e.g., poverty, market access) and global market dynamics to drive outcomes [22, 9] (Fig. 3, app 1). Domain knowledge alone often cannot keep up with these shifting dynamics. Data-driven causal discovery provides a method for constructing graphs that reveal hidden drivers and quantify changing relationships over time and space. For example, Mwebaze et al. [24] applied an ensemble of causal discovery algorithms to identify drivers of famine risk in Uganda, supporting improved prediction and intervention strategies. *(ii) Comparing crop growth models:* Process-based crop models are widely used to simulate crop development and yield under different environmental and management conditions. These models often differ substantially in their sensitivity to factors such as temperature, precipitation, soil characteristics, etc., which introduces uncertainty in yield projections [25, 14]. Causal discovery can be applied to the outputs of multiple models to identify the implied causal relationships between inputs (e.g., weather, soil, management) and outputs (e.g., yield, biomass). By comparing these causal structures against each other and against observational data, researchers can evaluate which models best capture the true underlying mechanisms. This evaluation goes beyond traditional sensitivity analysis, allowing identification of systematic biases or missing interactions and providing a principled way to improve model design and parameterization [26]. Such an approach can increase confidence in model predictions and guide both scientific understanding and practical decision-making in agricultural planning (Fig. 3, app 2).

**Assisting Policymakers and Farmers** *(i) Estimating impacts of human and natural interventions:* Causal ML supports evidence-based policy by quantifying the effects of interventions. For example, cash transfers in drought-affected regions of the Horn of Africa were evaluated using ATE to measure their impact on food security, providing actionable guidance for humanitarian organizations [4]. Similarly, understanding how extreme weather events affect crop yields helps design risk management strategies for farmers and policymakers (Fig. 3, app 3,4). *(ii) Personalizing sustainable practices:* Agricultural interventions often have heterogeneous impacts due to local environmental and management conditions. CATE methods allow geospatial tailoring of practices, prioritizing the most beneficial strategies for each land unit [10] (Fig. 3, app 5). For instance, [11] demonstrated how crop rotation effects on productivity vary with temperature and water deficit, informing localized recommendations. *iii) Evaluating digital agriculture recommendations:* Causal ML can be used to assess the effectiveness of specific decision support systems (Fig. 3, app 6). Using ATE estimation, [45] found that following optimal day-of-sowing recommendations increased yield by 12–17%. Such evidence builds trust, guides cost-benefit analyses and supports fair pricing of digital services.

**Improving Predictive Modeling** Causality-aware ML can enhance predictive modeling in agriculture by improving model stability [7] (Fig. 3, app 7, 8). For example, national-level crop yield forecasting is vital for food security, yet training data are often scarce or unevenly distributed, especially in food-insecure regions. By focusing on stable, causal variables, causality-aware methods improve geographic generalization [20]. They also increase robustness over time, such as in pest prediction models that must adapt to new control strategies or unusual environmental events [46].

## 4 Discussion and Outlook

Causal ML can promote agricultural sustainability through two complementary approaches: enhancing predictive model stability (causality for ML) and answering causal questions (ML for causality), providing actionable insights for stakeholders from farmers to policymakers. A causal ML workflow ensures effective application and includes five key elements, with their challenges and opportunities:

**Causal question.** Explicit causal framing is essential for addressing management decisions. Expert knowledge can be encoded in causal graphs to guide variable selection and method choice [27]. Questions must be relevant, precise and feasible. **Data.** Agroecosystems are complex, nonlinear and non-stationary [3, 39, 38]. Effective analysis requires curated, multi-modal data across environmental, economic and social variables. Temporal and spatial scales must align with the causal question and proxies should closely reflect conceptual variables. **Assumptions.** When randomization is impossible, structural and statistical assumptions are needed to infer causality [42]. Assumptions should be explicit and plausible, with spatial dependencies considered to avoid biases (e.g., pesticides affecting neighboring fields) [1]. **Methods.** Method choice depends on data, assumptions and question type. Open challenges in methods include scalable algorithms for high-dimensional, mixed and non-stationary data [38], spatial interactions [1] and causal representation learning [41]. **Validation.** Expert-derived causal graphs can benchmark discovery methods. Effect estimation can be evaluated using process-based models or synthetic data, and combined with randomized experiments when available [16]. Causality-aware ML models should be assessed for geographic generalization, transferability and robustness [44, 23]. Initiatives like AgML/AgMIP [44] and CauseMe [23] support standardization, collaboration and continuous evaluation.

Integrating causal thinking into agroecosystem sustainability enables transparent assessment of interventions and supports informed decision-making while navigating trade-offs between productivity, environmental health and social outcomes.

## References


[1] Akbari, K., Winter, S., and Tomko, M. (2023). Spatial causality: A systematic review on spatial causal inference. *Geographical Analysis*, 55(1):56–89.

[2] Arjovsky, M., Bottou, L., Gulrajani, I., and Lopez-Paz, D. (2019). Invariant risk minimization. *arXiv preprint arXiv:1907.02893*.

[3] Camps-Valls, G., Gerhardus, A., Ninad, U., Varando, G., Martius, G., Balaguer-Ballester, E., Vinuesa, R., Diaz, E., Zanna, L., and Runge, J. (2023). Discovering causal relations and equations from data. *Physics Reports*, 1044(1):1–68.

[4] Cerda-Bautista, J., Tarraga, J. M., Sitokonstantinou, V., and Camps-Valls, G. (2024). Assessing the causal impact of humanitarian aid on food security. In *IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium*, pages 1546–1552.

[5] Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters.

[6] Chickering, D. M. and Heckerman, D. (2002). Optimal structure identification with greedy search. *Journal of machine learning research*, 3(Dec):507–554.

[7] Cui, P. and Athey, S. (2022). Stable learning establishes some common ground between causal inference and machine learning. *Nature Machine Intelligence*, 4(2):110–115.

[8] Francis, C., Lieblein, G., Gliessman, S., Breland, T., Creamer, N., Harwood, R., Salomonsson, L., Helenius, J., Rickerl, D., Salvador, R., Wiedenhoeft, M., Simmons, S., Allen, P., Altieri, M., Flora, C., and Poincelot, R. (2003). Agroecology: The ecology of food systems. *Journal of Sustainable Agriculture*, 22:99–118.

[9] Fyles, H. and Madramootoo, C. (2016). Key Drivers of Food Insecurity. In Madramootoo, C., editor, *Emerging Technologies for Promoting Food Security*, Woodhead Publishing Series in Food Science, Technology and Nutrition, pages 1–19. Woodhead Publishing, Oxford.

[10] Giannarakis, G., Sitokonstantinou, V., Lorilla, R. S., and Kontoes, C. (2022a). Personalizing sustainable agriculture with causal machine learning. In *NeurIPS 2022 Workshop on Tackling Climate Change with Machine Learning*.

[11] Giannarakis, G., Sitokonstantinou, V., Lorilla, R. S., and Kontoes, C. (2022b). Towards assessing agricultural land suitability with causal machine learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1442–1452.

[12] Glymour, C., Zhang, K., and Spirtes, P. (2019). Review of causal discovery methods based on graphical models. *Frontiers in genetics*, 10:524.

[13] Godfray, H. C. J., Beddington, J. R., and Crute, I. R. (2010). Food security: the challenge of feeding 9 billion people. *Science*, 327(5967):812–818.

[14] Jägermeyr, J., Müller, C., Ruane, A. C., Elliott, J., Balkovic, J., Castillo, O., Faye, B., Foster, I., Folberth, C., Franke, J. A., Fuchs, K., Guarin, J. R., Heinke, J., Hoogenboom, G., Iizumi, T., Jain, A. K., Kelly, D., Khabarov, N., Lange, S., Lin, T.-S., Liu, W., Mialyk, O., Minoli, S., Moyer, E. J., Okada, M., Phillips, M., Porter, C., Rabin, S. S., Scheer, C., Schneider, J. M., Schyns, J. F., Skalsky, R., Smerald, A., Stella, T., Stephens, H., Webber, H., Zabel, F., and Rosenzweig, C. (2021). Climate impacts on global agriculture emerge earlier in new generation of climate and crop models. *Nature Food*, 2(11):873–885.

[15] Kaddour, J. and Preux, P. (2022). Causal machine learning: A survey and open problems. *arXiv preprint arXiv:2206.15475*.

[16] Kluger, D. M., Owen, A. B., and Lobell, D. B. (2022). Combining randomized field experiments with observational satellite data to assess the benefits of crop rotations on yields. *Environmental Research Letters*, 17(4):044066.

[17] Krueger, D., Caballero, E., Jacobsen, J.-H., Zhang, A., Binas, J., Zhang, D., Le Priol, R., and Courville, A. (2021). Out-of-distribution generalization via risk extrapolation (rex). In *International Conference on Machine Learning*, pages 5815–5826. PMLR.

[18] Künzel, S., Sekhon, J. S., Bickel, J., Yu, B., Bennett, C., Xie, M., and Tibshirani, R. (2019). Metalearners for estimating heterogeneous treatment effects using machine learning. *Proceedings of the National Academy of Sciences*, 116(10):4156–4165.

[19] Li, J., Hong, D., Gao, L., Yao, J., Zheng, K., Zhang, B., and Chanussot, J. (2022). Deep learning in multimodal remote sensing data fusion: A comprehensive review. *International Journal of Applied Earth Observation and Geoinformation*, 112:102926.

[20] Liu, J., Wu, J., Pi, R., Xu, R., Zhang, X., Li, B., and Cui, P. (2022). Measure the predictive heterogeneity. In *The Eleventh International Conference on Learning Representations*.

[21] Meshram, V., Patil, K., Meshram, V., Hanchate, D., and Ramkteke, S. (2021). Machine learning in agriculture domain: A state-of-art survey. *Artificial Intelligence in the Life Sciences*, 1:100010.

[22] Misselhorn, A. A. (2005). What drives food insecurity in southern africa? a meta-analysis of household economy studies. *Global Environmental Change*, 15(1):33–43.

[23] Munoz-Marí, J., Mateo, G., Runge, J., and Camps-Valls, G. (2020). Causeme: An online system for benchmarking causal discovery methods. *In Preparation*.

[24] Mwebaze, E., Okori, W., and Quinn, J. A. (2010). Causal structure learning for famine prediction. In *2010 AAAI Spring Symposium Series*.

[25] Müller, C., Jägermeyr, J., Franke, J. A., Ruane, A. C., Balkovic, J., Ciais, P., Dury, M., Falloon, P., Folberth, C., Hank, T., Hoffmann, M., Izaurralde, R. C., Jacquemin, I., Khabarov, N., Liu, W., Olin, S., Pugh, T. A. M., Wang, X., Williams, K., Zabel, F., and Elliott, J. W. (2024). Substantial Differences in Crop Yield Sensitivities Between Models Call for Functionality-Based Model Evaluation. *Earth's Future*, 12(3):e2023EF003773.

[26] Nowack, P., Kretschmer, M., and Marotzke, J. (2020). Causal networks for climate model evaluation and constrained projections. *Nature Communications*, 11(1):1415.

[27] Pearl, J. (2009). *Causality*. Cambridge university press, Cambridge, UK and New York, NY, USA.

[28] Peters, J., Bühlmann, P., and Meinshausen, N. (2016). Causal inference by using invariant prediction: identification and confidence intervals. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 78(5):947–1012.

[29] Petersen, R. G. (1994). *Agricultural field experiments: design and analysis*. CRC Press, FL, USA.

[30] Pretty, J. (2013). Agricultural sustainability: concepts, principles and evidence. *Philosophical Transactions of the Royal Society of London, Series B*, 363:447–465.

[31] Richardson, T. S. (2013). Single World Intervention Graphs (SWIGs): A Unification of the Counterfactual and Graphical Approaches to Causality.

[32] Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.

[33] Rothenhäusler, D., Meinshausen, N., Bühlmann, P., and Peters, J. (2021). Anchor regression: Heterogeneous data meet causality. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 83(2):215–246.

[34] Rothwell, P. M. (2005). External validity of randomised controlled trials:"to whom do the results of this trial apply?". *The Lancet*, 365(9453):82–93.

[35] Rubin, D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association*, 100(469):322–331.

[36] Rubin, D. B. (2006). Matched sampling for causal effects. *Journal of Computerized Tomography*, 30(3):16–19.

[37] Runge, J. (2023). Modern causal inference approaches to investigate biodiversity-ecosystem functioning relationships. *Nature Communications*, 14(1):1917.

[38] Runge, J., Bathiany, S., Bollt, E., Camps-Valls, G., Coumou, D., Deyle, E., Glymour, C., Kretschmer, M., Mahecha, M. D., Muñoz-Marí, J., et al. (2019a). Inferring causation from time series in earth system sciences. *Nature communications*, 10(1):2553.

[39] Runge, J., Gerhardus, A., Varando, G., Eyring, V., and Camps-Valls, G. (2023). Causal inference for time series. *Nature Reviews Earth & Environment*, 4(7):487–505.

[40] Runge, J., Nowack, P., Kretschmer, M., Flaxman, S., and Sejdinovic, D. (2019b). Detecting and quantifying causal associations in large nonlinear time series datasets. *Science advances*, 5(11):eaau4996.

[41] Schölkopf, B. (2022). Causality for machine learning. In *Probabilistic and causal inference: The works of Judea Pearl*, pages 765–804.

[42] Sharma, A., Syrgkanis, V., Zhang, C., and Kıcıman, E. (2021). Dowhy: Addressing challenges in expressing and validating causal assumptions. *arXiv preprint arXiv:2108.13518*.

[43] Spirtes, P., Glymour, C., and Scheines, R. (2000). Causation, prediction, and search.

[44] Sweet, L., Athanasiadis, I. N., van Bree, R., Castellano, A., Martre, P., Paudel, D., Ruane, A. C., and Zscheischler, J. (2024). Transdisciplinary coordination is essential for advancing agricultural modeling with machine learning. *Submitted manuscript*.

[45] Tsoumas, I., Giannarakis, G., Sitokonstantinou, V., Koukos, A., Loka, D., Bartsotas, N., and Athanasiadis, I. (2023a). Evaluating digital agriculture recommendations with causal inference. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 14514–14522.

[46] Tsoumas, I., Sitokonstantinou, V., Giannarakis, G., Lampiri, E., Athanassiou, C., Camps-Valls, G., and Athanasiadis, I. (2023b). Causality and explainability for trustworthy integrated pest management. In *NeurIPS 2023 Workshop on Tackling Climate Change with Machine Learning*.

[47] Wang, J., Lan, C., Liu, C., Ouyang, Y., Qin, T., Lu, W., Chen, Y., Zeng, W., and Yu, P. S. (2023). Generalizing to unseen domains: A survey on domain generalization. *IEEE Trans. on Knowl. and Data Eng.*, 35(8):8052–8072.

[48] Zhang, K., Wang, Z., Zhang, J., and Schölkopf, B. (2015). On estimation of functional causal models: general results and application to the post-nonlinear causal model. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 7(2):1–22.

# A Appendix

## A.1 Predictive vs Causal ML

**Predictive ML** focuses on finding statistical associations to make predictions.

**Causal ML** is the synergistic use of causal inference and machine learning to improve causality with ML (ML for causality) or improve ML with causality (causality for ML). It seeks to distinguish between types of relationships that result in statistical associations: direct causal, indirect or mediated causal, and non-causal associations. Pearl's Structural Causal Model (SCM) framework [27] provides the language to do this.

**Causal tasks:** Causal questions like *what happens if...* can be expressed using the SCM framework. ML for causality involves two main sub-tasks: causal discovery and causal effect estimation.

- *Task.* What is the effect of a crop rotation on the field's soil health?
- *Predictive ML* can detect correlations but may mistakenly attribute changes to crop rotation due to confounding variables like climate.
- *Causal ML* allows us to quantify the extent to which soil health changes can be attributed to crop rotations by separating out confounding effects.

**Predictive tasks:** Use patterns in a training dataset to model associative relationships.

- *Task.* Predict crop yield in Africa with a model trained in Europe.
- *Predictive ML* relies on training and test data being very similar and cannot generalize well to different data distributions. The differences between Africa and Europe can make the model obsolete.
- *Causal ML* focuses on causal features that have a higher potential for generalization and robustness. By filtering out non-causal relationships, the model relies on more stable and fundamental patterns.

### A.2 Supplementary Figures

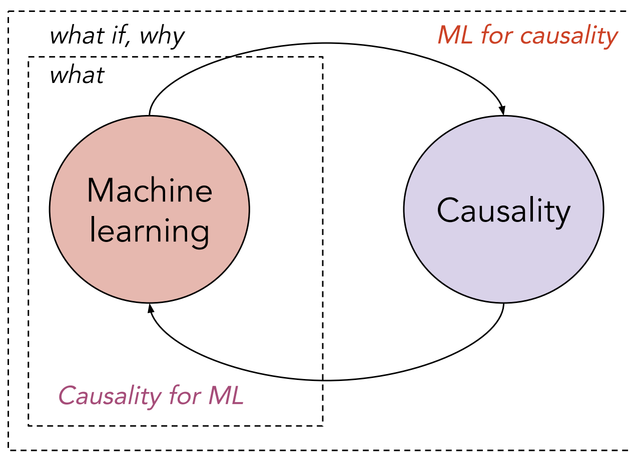


Figure 2: Visual representation of the relationship between predictive ML, causal inference, and causal ML. Causal ML builds upon the strengths of both.

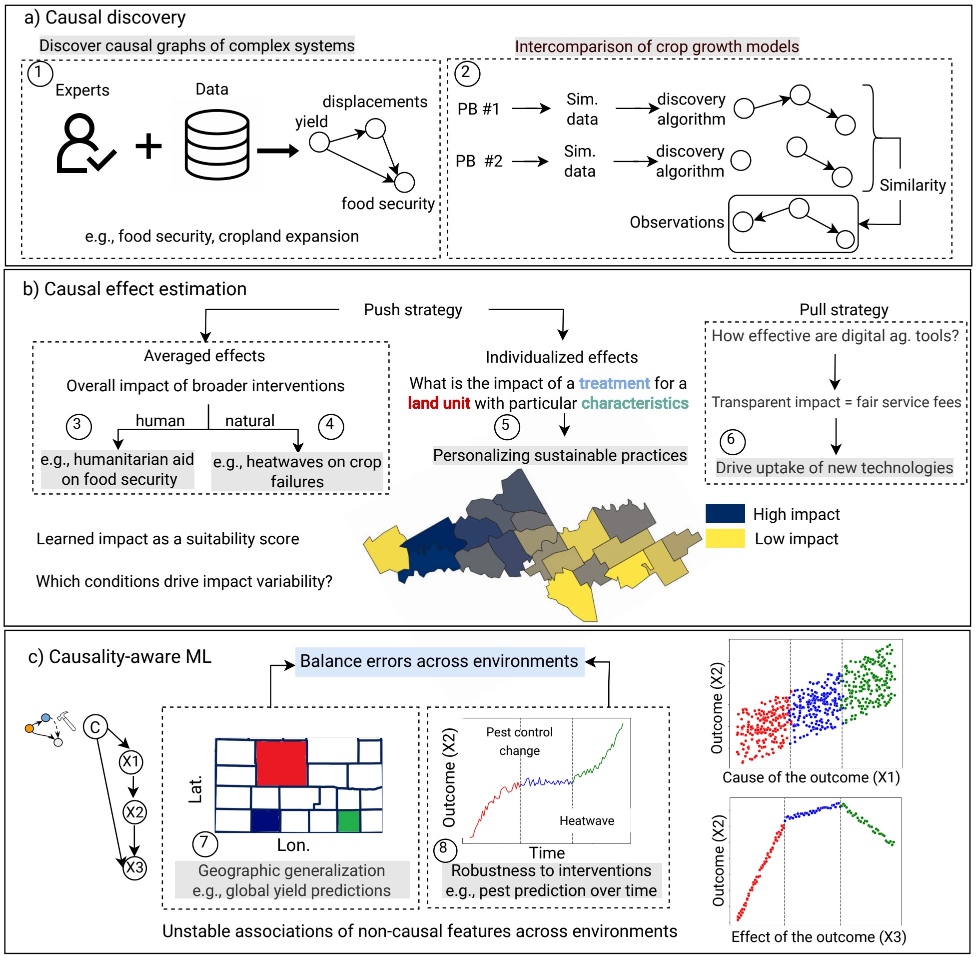


Figure 3: **Applications of causal ML for agriculture.** Panel a) Causal discovery applications: Data-driven causal discovery (1) unveils causal mechanisms and (2) evaluates process-based (PBs) models. Panel b) Causal effect estimation applications: Support evidence-based decisions by evaluating the impact of (3) human actions and (4) climate/weather events. (5) Sustainable practices can be spatially tailored by estimating each land unit's individualized impact. Panel c) Causality-aware ML applications: To achieve geographic generalization (7) and robustness to interventions (8) in predictive models, it is key to balance errors across various environments, which helps identify causal features that maintain a stable relationship with the outcome.