Causal Machine Learning for Sustainable Agroecosystems

Vasileios Sitokonstantinou[1*], Emiliano Díaz Salas-Porras[1], Jordi Cerdà-Bautista[1], Maria Piles[1], Ioannis Athanasiadis[2], Hannah Kerner[4], Giulia Martini[3], Lily-belle Sweet[5], Ilias Tsoumas[2,6], Jakob Zscheischler[5] and Gustau Camps-Valls[1]

[1*]Image Processing Laboratory, Universitat de València, Spain.
[2]Artificial Intelligence, Wageningen University and Research, The Netherlands.
[3]World Food Program, UN, Rome, Italy.
[4]Arizona State University, USA.
[5]Department of Compound Environmental Risks, Helmholtz Centre for Environmental Research – UFZ, Leipzig, Germany.
[6]National Observatory of Athens, Greece.

*Corresponding author(s). E-mail(s): Vasileios.Sitokonstantinou@uv.es;

Abstract

Sustainable agriculture is essential for food security and environmental health in a changing climate. However, it is challenging to understand the complex interactions among its biophysical, social, and economic components. Predictive machine learning (ML), with its capacity to learn from data, is leveraged in sustainable agriculture for applications like yield prediction and weather forecasting. Nevertheless, it cannot explain causal mechanisms and remains descriptive rather than prescriptive. To address this gap, we propose causal ML, which merges ML's data processing with causality's ability to reason about change. This facilitates quantifying intervention impacts for evidence-based decision-making and enhances predictive model robustness. We showcase causal ML through eight diverse applications that benefit stakeholders across the agri-food chain, including farmers, policymakers, and researchers.

1 Introduction

The perception of agriculture is evolving. Nowadays, policymakers, researchers, farmers, and consumers recognize farms as integral components of larger, interconnected agroecosystems that include biological, physical, social, and

economic dimensions [1]. However, current agricultural practices often fail to balance these elements, relying heavily on agricultural land expansion, external inputs like synthetic fertilizers and pesticides, and resource extraction. These practices contribute to environmental impacts like biodiversity loss and soil degradation that pose significant threats to future generations [2].

Sustainably achieving food security targets is a complex task. Production must be intensified to meet the growing food demand [3, 4], even on lands already degraded [5], while simultaneously reducing greenhouse gases emissions [6]. This challenge is compounded by slow technology adoption [7], inefficient policies [8], and unforeseen crises like pest outbreaks, pandemics, and financial disruptions [9]. Beyond achieving intensification without harming the environment and adding to climate change, agriculture has the potential to mitigate these problems by sequestering carbon in soils and enhancing biodiversity. Thus, agriculture can be both a contributor to and a solution for climate change [10]. To harness agriculture's potential for positive impact, we need a comprehensive framework that evaluates strategies based on their contributions to sustainability across the entire agroecosystem. Improved understanding of sustainability across all dimensions of agriculture can help prioritize the most impactful actions [11].

In this context, where agricultural decisions have far-reaching impacts, traditional approaches to modeling and decision-making may struggle to capture the complexity of agroecosystems. Machine learning (ML) methods have emerged as powerful tools for finding patterns within large datasets and making predictions based on historical data [12]. However, while ML excels at predicting outcomes, it cannot explain the underlying causality, which limits its effectiveness in performing robustly in new, changing environments [13] and evaluating the impact of interventions (see Box 1).

To address the limitations of predictive ML and enhance our understanding of causal mechanisms in agricultural systems, we introduce causal ML (see Fig. 1). Causal ML includes methods designed to infer causality from data, but also leverage causal knowledge to enhance predictive ML models [14]. Despite its potential, the adoption of causal ML in agriculture has been slower compared to related disciplines like ecology [15], public policy [16], and Earth and climate sciences [17]. This paper aims to bridge this gap by expressing complex ideas in a language that stakeholders in the food and agriculture sectors can understand and use. Integrating causal ML into decision-making processes can facilitate a transition towards evidence-based decisions that improve the sustainability of our food systems.

Box 1: Key differences between predictive ML and causal ML

Predictive ML focuses on finding statistical associations or dependencies among input variables to make predictions of an outcome.

Causal ML is the synergistic use of causal inference and machine learning to either, improve causality with ML (ML for causality) or improve ML with causality (causality for ML).
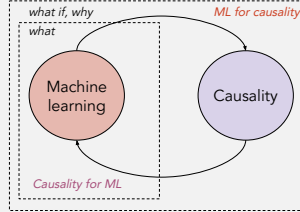
Causal questions: Causal questions such as *what happens if...* or *what would have happened if...* can be expressed using Pearl's Structural Causal Model (SCM) framework [18]. In this framework, a system is described with interdependent fundamental processes and random quantities. Each process is represented by an equation that includes qualitative elements, namely a list of causative variables, and quantitative elements, namely functions that describe relationships between the process variable and its causes. ML for causality involves two main sub-tasks: causal discovery, which determines SCM's qualitative elements, and causal effect estimation, which addresses SCM's quantitative elements.

- *Question.* What is the effect of a crop rotation on the field's soil health?
- *Predictive ML* can detect correlations based on historical data and estimate soil health from crop rotation data. However, it may mistakenly attribute soil health changes to crop rotation.
- *ML for causality* allows us to express the decisions as causal questions and to use ML guided by causal reasoning to answer them, i.e. to quantify to what extent the observed changes in soil health can be attributed to crop rotations.

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

- *Task.* Predict yield in Africa by exploiting the relationship between yield and biomass observed in Europe.
- *Predictive ML.* It exploits all patterns produced by associative relationships to obtain accurate models. However, it relies on training and test data being very similar. It cannot discriminate between types of association (e.g., cause-effect) to generalize to new data distributions.
- *Causality for ML.* Focuses on causal features with higher potential for generalization and robustness.
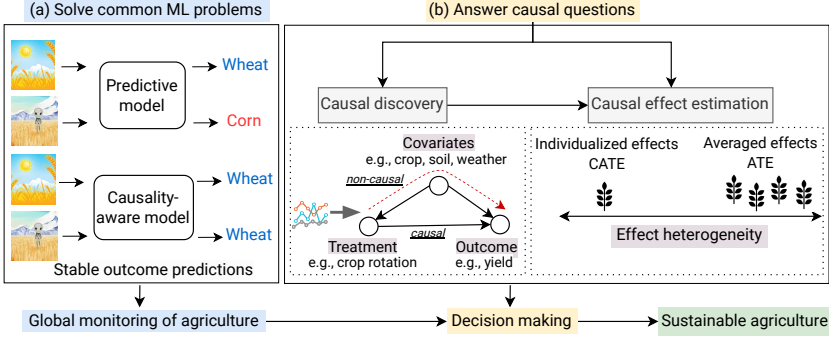
Fig. 1 How causality can drive sustainability in agriculture: a) Solve common ML problems: Address robustness to interventions over time (e.g., new policies) and geographic generalization to develop predictive models that can continuously and globally predict agricultural activity and ecological conditions. For instance, a standard model might fail to recognize wheat in snowy conditions if it was trained only on sunny conditions. A causality-aware model, however, learns the stable underlying features of wheat, allowing accurate identification in any weather. b) Answer causal questions: Use large-scale, robust predictions from (a) and/or other agricultural observations with causal discovery methods to produce causal graphs that map cause-effect relationships. Based on these graphs, estimate the effects of treatments (averaged or individualized) on outcomes of interest, as the relevant covariates for isolating non-causal associations (red line) are now known. With this knowledge, the most effective solutions can be identified and prioritized.

2 Causal ML for agriculture

2.1 ML for answering causal questions

Researchers have traditionally used randomized experiments to answer causal questions, namely establishing causal links between variables or estimating the effects of various practices or policies (treatments) [19]. Examples of such questions are provided in Box 2. While these experiments are known for providing unbiased estimates, they come with important challenges, including small sample sizes that lead to less precise results [20] and findings that might not apply universally to (even slightly) different conditions [21]. Advancements in big data and satellite-based remote sensing now enable the investigation of the impacts of agricultural practices or policies on sustainability outcomes using extensive datasets that cover diverse conditions. Deep learning techniques, for example, allow for the efficient processing of large amounts of data from multiple sources at different spatiotemporal scales [22]. These techniques enable the transformation and reduction of the raw data into useful representations. However, these learned representations are not necessarily aligned with the physical variables that best describe the system. Another issue is that these datasets are observational, meaning they lack randomized treatments and rely on the assumption that no external factors are influencing both the treatment and the outcome. Violating these assumptions can lead to biased estimates of causal effects.

To exploit ML's potential while addressing these issues we introduce the following 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 (see apps in Sec. 3.1). 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 (see apps in Sec. 3.2 and 3.3).

- **Collecting data.** To answer causal questions, we need to collect data on the agricultural system under study. To determine if cause-effect relationships exist between a group of variables (causal discovery), we need information on the variables of interest and any potential confounding variables: variables that may cause at least two of the variables of interest. To estimate the effectiveness of practices, policies, or technologies (causal effect estimation), we require data on the treatment of interest (e.g., fertilizer application), the observed outcome (e.g., farm profitability) for a specific unit (e.g., a field), and potential confounders like environmental factors (e.g., temperature), and crop and soil characteristics [18]. This information can be sourced from remote sensing observations, reanalysis data, process-based model simulations, farm management systems, crop calendars, farm accounting, agricultural field experiments, and plant phenotypical platforms.

- **Making assumptions.** To answer causal questions without randomized experiments, we must make some assumptions, either structural, process, or statistical. *Structural assumptions* concern the nature of potential causal relationships: Is unmeasured confounding present? What is the maximum delay for cause-effect relationships? *Process assumptions*, for example, might question whether linear models are sufficient for capturing causal relationships. *Statistical considerations* focus on evaluating whether data is adequate and representative of key variables. The core assumption for causal discovery is causal sufficiency (structural), ensuring all relevant variables are measured. For causal effect estimation, the key assumption is the causal graph (structural), which outlines all cause-effect relationships and can be derived from various sources including expert knowledge and data-driven methods. Rubin's potential outcomes framework [23] complements Pearl's SCM [24] (cf. Box 1) by providing a detailed list of necessary assumptions for a correct causal graph.

- **Selecting causal ML method.** The choice of causal ML method depends on the causal question, the available data, and the assumptions. For causal discovery questions, constraint methods (e.g., the PC and FCI algorithms) [25] and score methods (e.g., GES) [26] make minimal data assumptions but do not guarantee the discovery of the full causal graph. There exist variations for both cross-sectional [27] and time-series [28, 29] data. Asymmetry methods [30] make additional (process) assumptions regarding the types of

functions that describe the cause-effect relationships to guarantee that they are all discovered.

Answering causal effect questions involves estimating two key quantities: the Average Treatment Effect (ATE) and the Conditional Average Treatment Effect (CATE). The ATE summarizes the treatment's effectiveness by comparing average outcomes between treated and untreated groups, e.g., answering *'What is the average impact of organic farming (treatment) in Europe (entire population)?'* To estimate the ATE from observational data, researchers employ techniques such as matching [31], where similar treated and control groups are paired for comparison or propensity score methods [32] which account for the probability of an individual receiving treatment. The CATE adopts a more granular approach by focusing on specific subgroups defined by observed covariates, e.g., answering *'How does the impact of organic farming (treatment) vary across countries in Europe (population subgroups)'* By estimating CATE, researchers can identify which subgroups benefit most from the treatment, allowing for the individualization of treatment recommendations. Techniques like the X-learner [33] and Double Machine Learning (DML) [34] are commonly employed.

- **Checking robustness.** When we use observational data instead of data from randomized experiments to answer causal questions, we need to make assumptions about the data and the process that generated them. Robustness checks must be made to stress-test the plausibility of these assumptions. For causal discovery, this involves contrasting discovered relationships with domain knowledge. It may also involve testing the methods with artificially generated data, such as that generated with process-based models, and for which ground-truth causal graphs are available. Refutation tests [35] are used for causal effect estimation to check if the assumptions made result in coherent estimates. Typical refutation methods involve manipulating the data so that the impact of this manipulation on results is predictable, provided the analysis is correct. For example, noise variables can be added as additional confounder variables to check that the estimated effect remains stable.

2.2 Causality for improved ML predictions

Predictive ML models do not consider the causal structure of agricultural systems, thus they cannot adapt predictions accordingly when changes across space or time occur (cf. Box 1). Causality-aware ML leverages the idea that associations coming from direct cause-effect relationships are the most stable (see apps in Sec. 3.4). By selecting these direct causes, models can rely on relationships that persist across different environments or contexts [36]. This enhances the models' robustness to interventions, spatio-temporal generalizability, and explainability. This selection can be guided by prior knowledge of the causal graph or through data-driven techniques such as Invariant Causal Prediction (ICP) [37], which aims to discover features that exhibit consistent predictive capacity under various experimental settings, environments,

regimes, or interventions. However, solely relying on direct causes may be too conservative for some problems. Anchor Regression [38] introduces a relaxation by incorporating weighted features that may extend beyond direct causes. This method allows the model to leverage additional information while prioritizing causal relationships, thus balancing accuracy and stability. Invariant Risk Minimization (IRM) [39] and Risk extrapolation (REx) [40] are other methods that leverage causal principles to enhance model stability.

3 Applications of causal ML in agriculture

Causal ML is about answering causal queries (like the ones in Box 2), avoiding being right for the wrong reasons. We next showcase eight indicative applications of causal ML for sustainable agriculture (see Fig. 2); from advancing science to assisting policymakers, empowering farmers, and improving models. We characterize them with estimated levels of leverage, adoption risk, and impact. By leverage, we mean that the application can be realized and provide major changes and improvements with relatively small or focused efforts. Adoption risk refers to the challenges in getting scientists, policymakers, or farmers to use the outputs of the applications. Impact signifies the application's potential to create significant positive outcomes in the targeted areas.

Box 2: Questions that causal ML can answer.

Scientists 🐝, policymakers 🦎, farmers 🧑‍🌾, and general public 👥

Causal discovery

- **What is the relationship between X and Y?** 🐝 Does El Niño Southern Oscillation (ENSO) cause soil moisture anomalies in Southern and Eastern Africa?
- **What are the drivers of complex systems?** 🐝, 🦎 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?

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?

3.1 Advancing science

Understanding complex systems

Food security exemplifies complex systems with causal mechanisms that vary
across scales, contexts, and time. Next to environmental stressors (e.g., per-
sistent droughts), socio-economic factors such as poverty, poor market access,
and lack of employment drive chronic food insecurity at the regional level [41].
Local shocks, like climate extremes, and food price spikes drive sudden drops
in regional food availability and exacerbate food security issues. Additionally,
global market dynamics and trade routes, influenced by regional disruptions,
impact food prices and availability elsewhere [42]. The complexity of these sys-
tems challenges domain experts. Data-driven causal discovery techniques can
complement expertise by analyzing large datasets to uncover hidden or emerg-
ing patterns (Fig. 2, app 1). For instance, researchers [43] used a committee
of causal discovery algorithms that collectively voted on the causal relation-
ships between socioeconomic factors and famine risk in Uganda. Capturing the
relationships in a causal graph enhances scientific understanding of underlying
mechanisms and how they vary, aids in improving predictions of famine across
different environments and contexts (cf. Sec. 3.4), and facilitates the estima-
tion of effects of various interventions on food security (cf. Sec. 3.2). We assess
this application as having low leverage, a medium level of adoption risk among
domain scientists, and a moderate impact.

Intercomparison of crop growth models

Crop growth models are essential for predicting yields, simulating environmen-
tal impacts, and optimizing resource use. However, current crop models exhibit
large differences in their sensitivities to drivers such as carbon dioxide, tem-
perature, and water [44], leading to uncertainty in projections of future global
yields [45]. The Agricultural Model Intercomparison and Improvement Project
(AgMIP) aims to improve crop growth models for assessing climate impacts
on agriculture by conducting extensive protocol-based model intercompar-
isons. Inspired by the work of [46], who applied causal discovery algorithms
to compare climate models, a similar approach could be implemented for the
inter-comparison of crop growth models (Fig. 2, app 2). Simulations generated
by various process-based crop growth models serve as inputs for causal dis-
covery algorithms. The resulting causal graphs can be then compared using
specific causal graph metrics with those derived from observational data to
evaluate alignment with the actual causal structure of crop growth. This
method offers an objective pathway for process-oriented model evaluation and
improvement that goes beyond sensitivity analyses and enhances the reliability
and accuracy of crop growth predictions. We assess this application as having
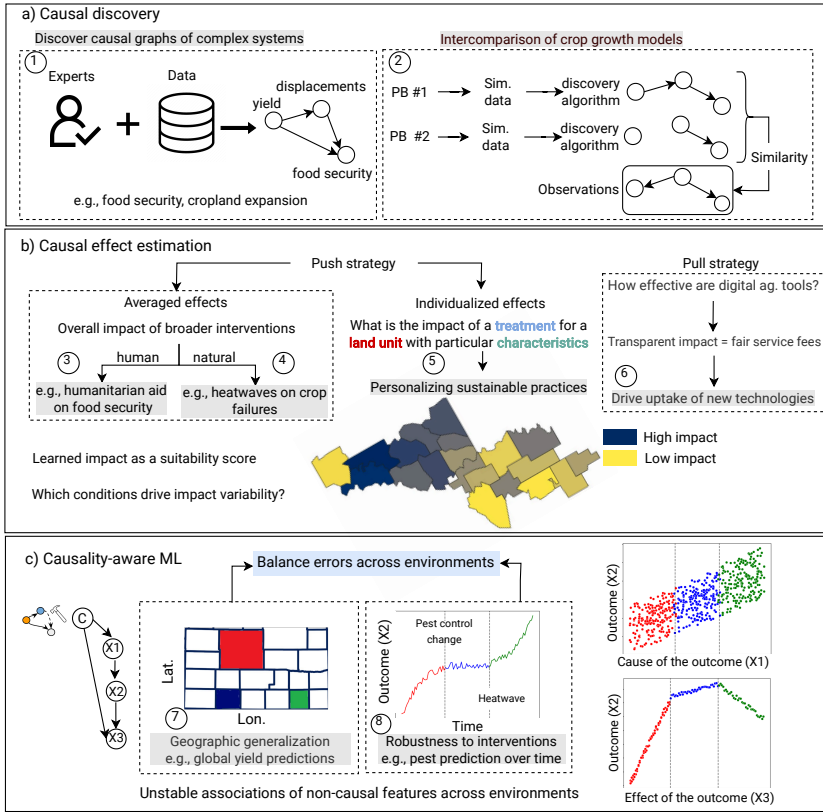
Fig. 2 Applications of causal ML for agriculture. Panel a) Causal discovery applications: Data-driven causal discovery (1) unveils causal mechanisms in complex systems like food security, enhancing domain expertise, and (2) evaluates process-based (PBs) models by comparing causal graphs from model simulations and observational data. Panel b) Causal effect estimation applications: Support evidence-based decisions by evaluating the impact of (3) human actions and (4) climate/weather events on sustainability outcomes. (5) Sustainable practices can be spatially tailored by estimating each land unit's individualized impact based on their characteristics. Which factors are responsible for the differences in 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 (e.g., geographic areas, anomalous weather events - variable C). This helps identify causal features, like X1, which maintain a stable relationship with the outcome, X2, under different conditions.

high leverage, since crop growth models are already in place, a medium level of adoption risk among domain scientists, and a moderate impact.

3.2 Assisting policymakers

Impact of human and natural interventions

Assessing the impact of policies, farmer decisions, or subsidies (treatments) on sustainability outcomes in agriculture is vital for evidence-based decision-making (Fig. 2, app 3). For example, we can assess the impact of humanitarian aid, where cash transfers provide financial assistance to individuals and households during extreme droughts in the Horn of Africa [47]. By estimating the ATE, we can measure the overall impact of these transfers on food security indicators, such as the Integrated Food Security Phase Classification (IPC), thus providing humanitarian organizations with evidence to design actionable responses. Shifting perspective, it is also useful to cast natural phenomena as the treatment of interest. For example, by understanding how extreme weather events, such as heatwaves or droughts, affect crop production, we can develop risk management strategies and support farmers in adapting (Fig. 2, app 4). Quantifying the overall impact of broader human and natural interventions is assessed as having high leverage, a medium level of adoption risk among policymakers, and a high impact.

Personalizing sustainable practices

Achieving sustainability goals in agriculture requires a shift from traditional one-size-fits-all approaches to geospatially tailored strategies [48]. The impact of agricultural practices varies across different regions due to local conditions. Leveraging CATE methods facilitates the assessment of these varying effects (Fig. 2, app 5) which in turn enables the prioritization of the most beneficial practices for each land unit. For example, [49] measured the CATE of crop rotation on net primary productivity (NPP) and proposed using it as a suitability score for applying crop rotation. They also explored which variables drove effect heterogeneity and showed that high min. temperatures combined with low water deficit increased the positive impact of crop rotation. We assess this application as having low leverage, as it requires large amounts of high-quality local data, a high level of adoption risk among policymakers and farmers, and a high impact.

3.3 Empowering farmers

Assessment of the effectiveness of digital agriculture

Digital agriculture tools, such as decision support systems and smart farming technologies, hold great promise for enhancing sustainability. However, low adoption rates remain a challenge. Evaluating their impact on farmers' profits can promote wider adoption (see Fig. 2, app 6). For instance, [50] estimated a 12-17% increase in crop yield for farmers who followed optimal day-of-sowing recommendations. Analyses like this can provide clear evidence, helping farmers make informed decisions and building trust in these technologies. Quantifying the effectiveness of digital agriculture tools also supports fair service fees and allows farmers to conduct accurate cost-benefit analyses.

We assess this application as having low leverage due to data access issues, a medium level of adoption risk among farmers, and a moderate impact.

3.4 Improving predictive modeling

Causality-aware ML can improve predictive modeling in agriculture by addressing challenges related to model stability [36] (Fig. 2, app 7 and 8). For instance, global information on national-level crop yield forecasting is key for food security crisis prevention and response planning. However, the uneven distribution of available training data, typically limited in time and quality for food-insecure countries, poses challenges for supervised models. Causality-aware ML methods (see Sec. 2.2), which focus on causal, stable variables, can offer a better geographic generalization for these models [51]. Additionally, these methods can offer robustness to interventions over time. For example, they can improve the resilience of pest prediction models when faced with changes in pest control strategies or anomalous environmental events [52]. We assess these applications as having high leverage, as they capitalize on existing predictive models and their datasets, a low level of adoption risk among data modelers, and a moderate impact.

4 Discussion and outlook

We highlighted the potential of causal ML in promoting agricultural sustainability and addressing food security problems through two key approaches: enhancing the stability of predictive models (causality for ML) and answering causal questions (ML for causality). The applications we discussed demonstrate how causal ML can generate actionable insights to empower stakeholders, from farmers to global agricultural managers. We introduced a causal ML workflow to ensure the effective and reliable application of causal ML for sustainable agroecosystems. Five elements should be worked out and integrated, all coming with their challenges and opportunities:

Causal question. In pursuit of effective agricultural decisions, addressing the causal questions inherent in agroecosystem management is key. However, explicit causal language necessary for articulating these questions is largely absent in agriculture. Domain experts possess valuable knowledge that can be translated into a causal graph, which, when faithfully constructed, can be used alongside ML methods to answer causal queries. Correctly defining the causal question is important; it must be relevant, precise, and feasible so that variables are measured accurately and appropriately [18].

Data. Agroecosystems are complex to model, predict, and understand, and involve nonlinearities and non-stationarities [17, 28, 53]. Effective analysis requires careful curation and harmonization of diverse data sources, including multi-modal Earth observation data and economic and social variables. High-resolution economic or social indicators like conflicts, GDP, and crop prices are often uncertain, less accessible, and inconsistently recorded compared to environmental data. The temporal and spatial scales of observation are important

considerations. For example, in shorter time frames, weather can cause fluctuations in crop yield, whereas over longer periods, the impact of agriculture on local climate can be measured. Data limitations are a significant concern, especially because in causal graphs, the variables at the nodes are conceptual and may not be directly observable. It is important to minimize the gap between proxies and the original concepts to maintain the integrity of the analysis.

Assumptions. When randomization of treatment allocation is not possible, making specific assumptions is necessary to move from correlations to causal conclusions (cf. Box 1 and Sec. 2.1). Structural, process, and statistical assumptions should be transparently laid out to ensure appropriate integration with data curation and method selection within the workflow. Assessing the plausibility of these assumptions is a necessary step [35]; it relies heavily on expert domain knowledge. Their adequacy may also be evaluated by checking the performance of the entire causal ML workflow during validation. An important type of assumption not discussed in Sec. 2.1 regards how causal structure leads to spatial dependence. If spatial assumptions are not adequate this may lead to spatial biases [54]. For example, chemical pesticides used by one farmer may impact the water quality for irrigation in neighboring fields as they seep into the groundwater (spillover).

Methods. Once the data curation and assumption elements have been established to answer the causal question of interest, selecting a method involves choosing one that aligns with these elements. For example, for causal discovery questions if we are unwilling to assume there is no unmeasured confounding then we must choose a method such as FCI [25] that relaxes this assumption. Gaps in the causal ML literature for agriculture include: scalable algorithms for mixed data types, high-dimensional data, and non-stationary causal relations [53]; methods that account for unit interactions [54]; and causal representation learning methods for raw unstructured data [13].

Validation. *Causal discovery:* Establishing a database of causal graphs for agriculture based on expert knowledge is recommended. These graphs can then serve as a benchmark for evaluating different causal discovery algorithms and guide the definition of downstream causal effect estimation tasks. *Effect estimation:* Ground truth data for effect estimates is typically not available. Instead, process-based models can be utilized: ground-truth causal effects can be extracted directly from their equations. Additionally, these equations can be used to generate artificial data. The causal effect estimation method is then evaluated on its ability to recover the ground-truth effects from the artificial data accurately. Another option is to combine observational effect estimates with randomized experiments [20]. *Causality-aware ML:* Model stability in causality-aware predictive models should be evaluated based on metrics like geographic generalization, transferability, or robustness to policy changes.

We recommend engagement in benchmark activities like those conducted by AgML/AgMIP [55] and making use of systematic evaluation platforms like causeme [56]. These initiatives foster collaboration, standardization, and continuous improvement in validating and evaluating (causal) ML methods in

geosciences and agriculture.

Integrating causal thinking into agroecosystem sustainability is not only sensible but necessary, as it provides a robust framework for making informed decisions with significant environmental, economic, and societal implications.

References

[1] 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., Poincelot, R.: Agroecology: The ecology of food systems. Journal of Sustainable Agriculture **22**, 99–118 (2003). https://doi.org/10.1300/J064v22n03_10

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

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

[4] Van Dijk, M., Morley, T., Rau, M.L., Saghai, Y.: A meta-analysis of projected global food demand and population at risk of hunger for the period 2010–2050. Nature Food **2**(7), 494–501 (2021)

[5] Eswaran, H., Lal, R., Reich, P.: Land degradation: an overview. Response to land degradation, 20–35 (2019)

[6] Ivanovich, C.C., Sun, T., Gordon, D.R., Ocko, I.B.: Future warming from global food consumption. Nature Climate Change **13**(3), 297–302 (2023)

[7] Takahashi, K., Muraoka, R., Otsuka, K.: Technology adoption, impact, and extension in developing countries agriculture: A review of the recent literature. Agricultural Economics **51** (2019). https://doi.org/10.1111/agec.12539

[8] Peer, G., van der Werf, W., Piro-Smith, E.: The eus common agricultural policy could be spent much more efficiently to address challenges (2020)

[9] Altieri, M.A., Nicholls, C.I.: Agroecology and the emergence of a post covid-19 agriculture. Agriculture and Human Values **37**, 525–526 (2020)

[10] Umesha, S., Manukumar, H.M., Chandrasekhar, B.: Sustainable agriculture and food security. In: Biotechnology for Sustainable Agriculture, pp. 67–92. Elsevier, NL (2018)

[11] Gliessman, S.R., *et al.*: Transforming food systems with agroecology.

Agroecology and Sustainable Food Systems **40**(3), 187–189 (2016)

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

[13] Schölkopf, B.: Causality for machine learning. In: Probabilistic and Causal Inference: The Works of Judea Pearl, pp. 765–804 (2022)

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

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

[16] Fougre, D., Jacquemet, N.: Causal inference and impact evaluation. Economie et Statistique/Economics and Statistics (2019)

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

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

[19] Petersen, R.G.: Agricultural Field Experiments: Design and Analysis. CRC Press, FL, USA (1994)

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

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

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

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

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

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

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

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

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

[29] Runge, J., Nowack, P., Kretschmer, M., Flaxman, S., Sejdinovic, D.: Detecting and quantifying causal associations in large nonlinear time series datasets. Science advances **5**(11), 4996 (2019)

[30] Zhang, K., Wang, Z., Zhang, J., Schölkopf, B.: 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 (2015)

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

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

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

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

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

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

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

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

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

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

[41] Misselhorn, A.A.: What drives food insecurity in southern africa? a meta-analysis of household economy studies. Global Environmental Change **15**(1), 33–43 (2005). https://doi.org/10.1016/j.gloenvcha.2004.11.003

[42] Fyles, H., Madramootoo, C.: Key Drivers of Food Insecurity. In: Madramootoo, C. (ed.) Emerging Technologies for Promoting Food Security. Woodhead Publishing Series in Food Science, Technology and Nutrition, pp. 1–19. Woodhead Publishing, Oxford (2016). https://doi.org/10.1016/B978-1-78242-335-5.00001-9. https://www.sciencedirect.com/science/article/pii/B9781782423355000019

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

[44] Mller, C., Jgermeyr, 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., Elliott, J.W.: Substantial Differences in Crop Yield Sensitivities Between Models Call for Functionality-Based Model Evaluation. Earth's Future **12**(3), 2023–003773 (2024). https://doi.org/10.1029/2023EF003773

[45] Jgermeyr, J., Mller, 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., Rosenzweig, C.: Climate impacts on global agriculture

emerge earlier in new generation of climate and crop models. Nature Food **2**(11), 873–885 (2021). https://doi.org/10.1038/s43016-021-00400-y

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

[47] Cerdà-Bautista, J., Tárraga, J.M., Sitokonstantinou, V., Camps-Valls, G.: Assessing the Causal Impact of Humanitarian Aid on Food Security (2024)

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

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

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

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

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

[53] 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.*: Inferring causation from time series in earth system sciences. Nature communications **10**(1), 2553 (2019)

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

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

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

Competing Interests

The authors declare no competing interests.