Robustly modeling the nonlinear impact of climate change on agriculture by combining econometrics and machine learning

Climate Change AI Workshop ICLR 2023: Tackling climate change with machine learning

Benedetta Francesconi

VU Amsterdam

Amsterdam, NL

benedetta.francesconi.1993@gmail.com

Ying-Jung Chen

Descartes Lab and Georgia Institute of Technology

San Francisco, US

yingjungcd@gmail.com

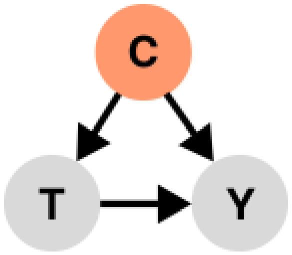
Problem and motivation

Temperature and other climatic variables have been changing during the last decades, as a result of climate change

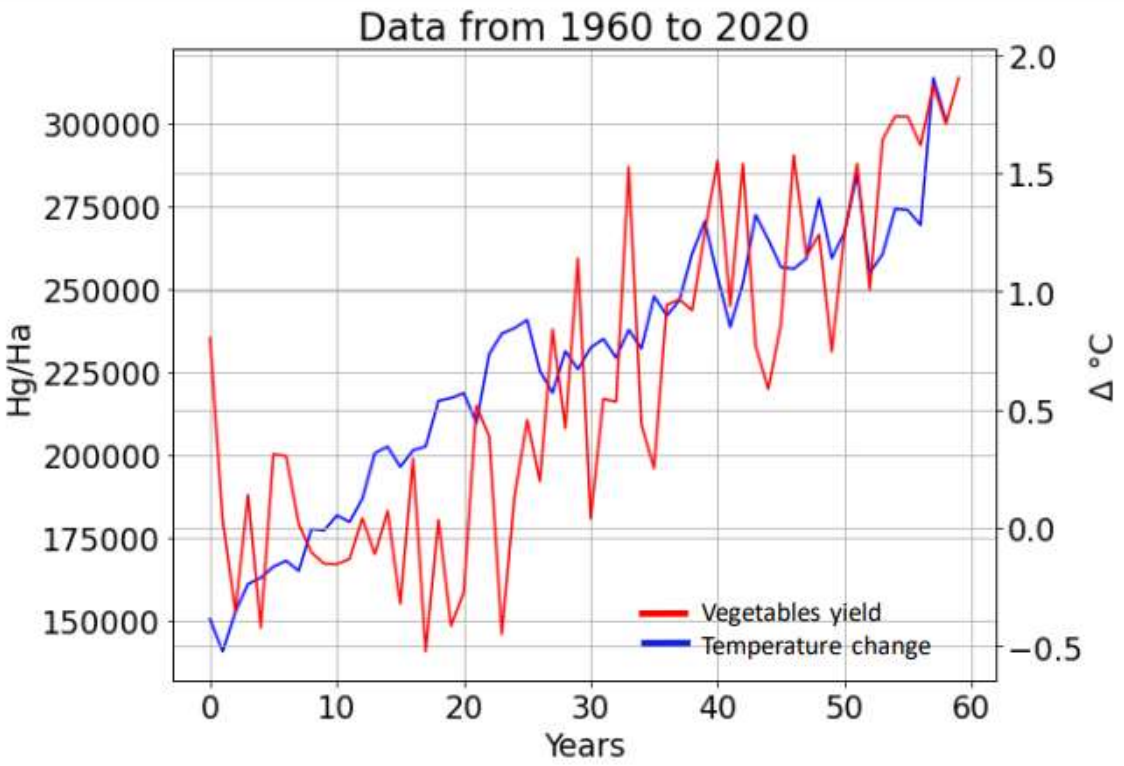
All sectors of society are likely to be affected by such drastic changes. The production sectors that are still highly tied to weather patterns, like agricultural production, will be those most affects

In order to design effective mitigation and adaptation policies, it is necessary to understand which exact variables have an influence on final crop yield

Problem and motivation



Both machine learning and econometrics have been attempting to perform causal estimation. Both approaches currently lack sufficient methodologies to be able to assess causality



Is temperature really affecting vegetables yield or is this a spurious relation? Causal analysis can help answer this question

Current approaches

Econometrics

Econometric approaches mainly rely on:

- Cointegration
- Granger causality

Which mainly use GARCH or VECM models. These models are not flexible enough to handle nonlinear phenomena and they generally lack an integration of causality concepts by default

Machine learning

Machine learning methods are in large part focused on obtaining good predictions, not on estimating causality

Our approach: data

Climate data

NOAA dataset:

- Average temperature data at 5x5 lat/lon scale from 1985 to 2022
- Average precipitation data at 5x5 lat/lon scale from 1985 to 2022
- Wind data at 1.9x1.9 scale data from 1979 to 2022
- Soil moisture at 0.5x0.5 resolution level from 1948 to 2022
- Shortwave and longwave radiation flux at 0.3 degrees resolution level from 1979 to 2022

Our approach: data

Agricultural data

Monthly crops data are rare. We propose to apply the approach in Wisser et alii. (2022) to obtain monthly gridded crop data up to 2022 by leveraging:

- The annual country level FAOSTAT database on crop harvested and crop production
- GAEZ* v4 gridded global annual harvested area, yield and production by crop
- GAUL** 2012 dataset, which reports the fraction of each global 5-minute grid cell that falls within a given country or disputed territory

Annual production

- Extract and calculate crop changes by country from 2010 to 2022
- Aggregate FAOSTAT-based ratios to match GAEZ+15 crop categories
- Apply country level ratios to grid cells
- Compute 2022 crop yields

↓

Monthly production

- Harmonize GAEZ+15 and MIRCA2000 crop lists
- Apply MIRCA2000 monthly crop dataset to GAEZ+15 annual data

* *Global Egro-Ecological Zones*
** *Global Administrative Unit Layer*

Our approach: methodology

Cointegration analysis

- Johansen cointegration
- Threshold cointegration
- Time-varying cointegration

Granger causality

Perform the test using a DeepAR model

Counterfactual analysis

- Use of knockoffs
- Generate a toy dataset and use it to compare results

Limitations

As agricultural production is not affected by climate only, in the future it will be necessary to include other variables as well, which also affect production. In particular:

- Economic factors
- Social factors
- Technology development
- Use of chemical fertilizers