Using Neural Networks in Agricultural Economics – An Application to Modeling Costs of the Federal Crop Insurance Program

Andrew Crane-Droesch and Joseph Cooper

Selected Paper prepared for presentation at the International Agricultural Trade Research Consortium's (IATRC's) 2019 Annual Meeting: Recent Advances in Applied General Equilibrium Modeling: Relevance and Application to Agricultural Trade Analysis, December 8-10, 2019, Washington, DC.

Using Neural Networks in Agricultural Economics -- An Application to Modeling Costs of the Federal Crop Insurance program

Andrew Crane-Droesch, Perelman School of Medicine, Univ. of Pennsylvania

Joseph Cooper, Office of the Chief Economist, USDA

2019 IATRC Annual Meeting
Washington DC, 2019

The findings and conclusions in this presentation are those of the authors and should not be construed to represent any official USDA or U.S. Government determination or policy. This presentation was supported by the U.S. Department of Agriculture, Office of the Chief Economist.

Overview

- Causal inference and machine learning
- Semiparametric neural nets
  - What, why, and how
- Application 1: crop yield prediction and climate change impact assessment in agriculture
- Application 2: the impact of climate change on the cost of the Federal Crop Insurance Program

Causal Inference vs Machine Learning

Given outcomes y and data $\boldsymbol{X}$,

- **Causal inference:** Estimate $\frac{\partial y}{\partial x}$
  - Goal is to understand marginal effects – *what will happen if I exogenously change x?*
  - *Any* bias is generally unacceptable
- **Machine learning:** Predict y^*
  - Goal is not to understand how x affects y, but to predict new values of the outcome y
  - Usually willing to trade bias for variance

Hybrids

Heterogeneous treatment effects:

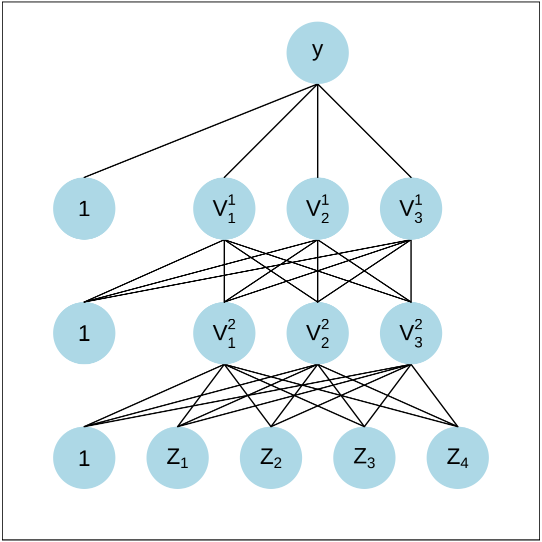
- Given outcomes y, a "treatment" D, and covariates $\boldsymbol{X}$
- Estimate $\tau(\boldsymbol{X}) = \mathbb{E}\left[y^{D=1} - y^{D=0} | \boldsymbol{X} = \boldsymbol{x}\right]$
- Promises to estimate *Individualized treatment effects*
- Papers:
  - Causal Trees: Athey & Imbens (2015)
  - Causal Forests: Wager & Athey (2016)

High-Dimensional Regression Adjustment:

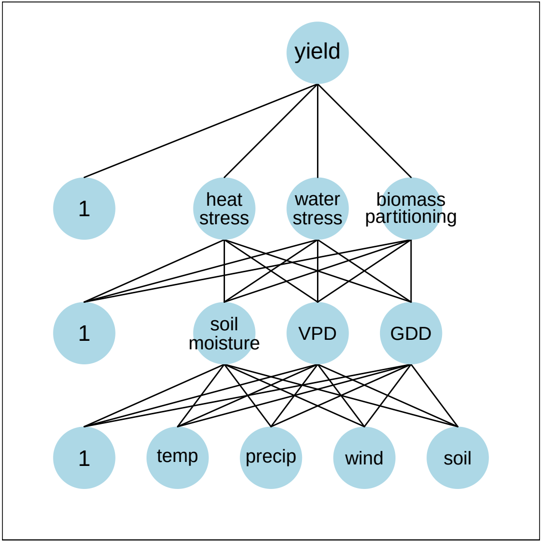
- Given a model $y = \boldsymbol{\theta} D + g(\boldsymbol{Z}) + \epsilon$
  - Where $\boldsymbol{\theta}$ is of interest and $\boldsymbol{Z}$ is confounding via $\mathbb{E}[\epsilon|\boldsymbol{Z}] \neq 0$
- Use ML to estimate an orthagonalizing transformation V such that $VD \perp Z$
- Papers: Chernozhukov et al. 2016 "Double Machine Learning"

Deep instrumental variables (Hartford et al. 2017)

Neural networks



Representation Learning



Neural networks

$$
\begin{aligned}
y &= \gamma^1 + \boldsymbol{V}^1\Gamma^1 + \epsilon \\
\boldsymbol{V}^1 &= a\left(\gamma^2 + \boldsymbol{V}^2\boldsymbol{\Gamma}^2\right) \\
\boldsymbol{V}^2 &= a\left(\gamma^3 + \boldsymbol{V}^3\boldsymbol{\Gamma}^3\right) \\
&\vdots \\
\boldsymbol{V}^L &= a\left(\gamma^L + \boldsymbol{Z}\boldsymbol{\Gamma}^L\right)
\end{aligned}
$$

- y – a (continuous) outcome
- ϵ – additive error
- $\boldsymbol{Z}$ – data
- $\boldsymbol{V}^l$ – "nodes": derived variables
- $a()$ – the "activation function". Maps the real line to some subset of it. Modern nets use variants of the ReLU: $a(x) = max(0, x)$
- Dimension of $\boldsymbol{\Gamma}^{1:L}$ controls number of nodes per layer

Semiparametric neural nets

- The top layer of a neural net is an OLS regression in derived variables $\boldsymbol{V}$

$$y = \gamma + \boldsymbol{V}\Gamma + \epsilon$$

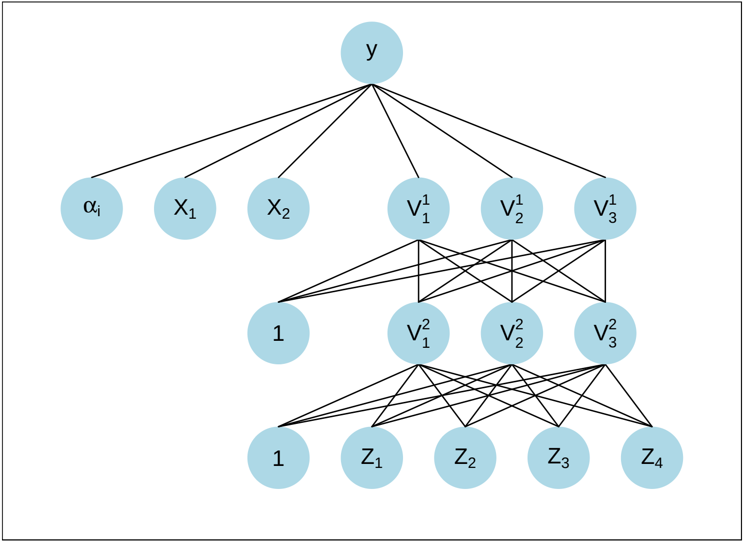
- It is simple to add linear terms to the model, where a linear-in-parameters relationship is known to be appropriate:

$$y = \gamma + \boldsymbol{X}\beta + \boldsymbol{V}\Gamma + \epsilon$$

- Likewise, panel structure can be accounted-for by adding unit-specific intercepts at the top level:

$$y_{it} = \alpha_i + \boldsymbol{X}_{it}\beta + \boldsymbol{V}_{it}\Gamma + \epsilon_{it}$$

Semiparametric and panel neural nets



Why might you want to do this?

- **Statistical efficiency**
- When the goal is to do prediction, but you're working in a topic area in which people have been doing inference and understand the data-generating process somewhat
- When a purely nonparametric model has a hard time representing specific kinds of structure:
  - Longitudinal structure
  - Secular trends
  - Response heterogeneity
- (With caveats) for certain sorts of causal inference tasks
  - Unbiased estimates of estimates of β from a model $y = \boldsymbol{X}\beta + f(\boldsymbol{Z}) + \epsilon$ where $f()$ unknown, $\boldsymbol{Z}$ high-dimensional, and $\boldsymbol{X} \not\perp f(\boldsymbol{Z})$
  - Examples: Instrumental variables models, high-dimensional regression adjustment

Training: backpropagation

No closed form solution to the parameters in a neural net. Training done by gradient descent. If the loss is

$$R = N^{-1} \sum (y - \hat{y})^2 + \lambda \sum \boldsymbol{\theta}^{\mathbf{2}}$$

then the gradients are:

$$\begin{aligned}
\frac{\partial R}{\partial \Gamma_1} &= -2\boldsymbol{V}_1^T \hat{\epsilon} + 2\lambda\Gamma_1 \\
\frac{\partial R}{\partial \Gamma_2} &= \boldsymbol{V}_2^T \underbrace{a'(\boldsymbol{V}_2\boldsymbol{\Gamma}_2) \odot -2\hat{\epsilon}\boldsymbol{\Gamma}_1^T}_{\text{stub}_1} + 2\lambda\Gamma_2 \\
\frac{\partial R}{\partial \Gamma_3} &= \boldsymbol{V}_3^T \underbrace{a'(\boldsymbol{V}_3\boldsymbol{\Gamma}_3) \odot \text{stub}_1 \boldsymbol{\Gamma}_2^T}_{\text{stub}_2} + 2\lambda\Gamma_3
\end{aligned}$$

et cetera...

updates are thus

$$\Gamma_{new} \leftarrow \Gamma_{old} + \text{learning rate} \times \frac{\partial R}{\partial \Gamma}$$

For a semiparametric net, the derivative WRT the linear slope coefs is just $-2\mathbf{X}^T\hat{\epsilon}$

The "OLS trick"

Gradient descent is inexact, and it's rare to train a neural net all the way to convergence

- More common to stop early based on test set performance

But if the parametric terms really are important, gradient descent might take a long time to figure this out.

- The nonparametric part of the net might start overfitting before the parametric slope coefs reach their true values

BUT we can exploit the fact that the top of a neural net is a linear regression, and use the closed-form solution:

$$\mathcal{B} = (\boldsymbol{W^T} \boldsymbol{W} + \lambda I)^{-1} \boldsymbol{W}^T y$$

where $\mathcal{B}$ are the weights at the top level, and $\boldsymbol{W}$ is the concatenaton of the parametric design matrix and the top-level derived variables.

Specifying some entries of λ to be zero allows for that some terms to be left unpenalized – useful if inference is desired.

Fixed and random effects

The model $y_{it} = \alpha_i + \boldsymbol{X}_{it}\beta + f(\boldsymbol{Z}_{it}) + \epsilon_{it}$ has an intercept for every cross-sectional unit.

- Estimating these by gradient descent is problematic for the reasons described above
- Estimating them by OLS involves inverting a huge matrix

But because of the neural net's structure, we can apply classical linear regression machinery to remove/estimate α.

- For fixed effects:

$$(y_{it} - \bar{y}_i) = \alpha_i - \bar{\alpha}_i + \left(\boldsymbol{X}_{it} - \bar{\boldsymbol{X}}_i\right)\beta + \left(\boldsymbol{V}_{it} - \bar{\boldsymbol{V}}_i\right)\Gamma + \epsilon_{it} - \bar{\epsilon}_{it}$$

- For random effects:

$$\left(y_{it} - \tilde{\lambda}\bar{y}_i\right) = \alpha_i - \tilde{\lambda}\alpha_i + \left(\boldsymbol{X}_{it} - \tilde{\lambda}\bar{\boldsymbol{X}}_i\right)\beta + \left(\boldsymbol{V}_{it} - \tilde{\lambda}\bar{\boldsymbol{V}}_i\right)\Gamma + \epsilon_{it} - \tilde{\lambda}\bar{\epsilon}_{it}$$

where $\tilde{\lambda} \in [0, 1]$ controls the variance of the random effects and corresponds to a L2 penalty between $[0, \infty]$

Gradient descent can proceed on transformed terms, and $\hat{\alpha}$ recoverable with algebra

Application 1

The method

- It has been implemented in the R package `panelNNET`, and a paper on the method has been published in *Environmental Research Letters*

IOPscience Journals Books Publishing Support Login Search IOPs

Environmental Research Letters

ACCEPTED MANUSCRIPT • OPEN ACCESS

Machine learning methods for crop yield prediction and climate change impact assessment in agriculture

Andrew Crane-Droesch[1]

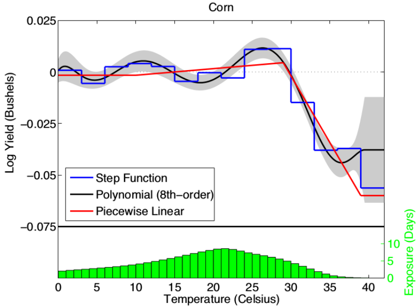
Accepted Manuscript online 14 September 2018 • Not subject to copyright in the USA. Contribution of United States Department of Agriculture

- It achieves state-of the art predictive skill in its domain, outperforming fully-nonparametric neural nets as well as parametric statistical models

Baseline parametric yield model

$$y_{it} = \alpha_i + \sum_r \text{GDD}_{rit}\beta_r + \boldsymbol{X}_{it}\boldsymbol{\beta} + \epsilon_{it}$$

- GDD = "growing degree days" – proportion of the year spent between specific temperature bands, e.g. 27-28°C
- Pioneered by Schlenker & Roberts (2009)



→ small shifts in heat have severe impacts on yields

Semiparametric Specification

$$
\begin{aligned}
y_{it} = \alpha_i + \sum_r \mathrm{GDD}_{rit}\beta_r + \boldsymbol{X}_{it}\boldsymbol{\beta} + \boldsymbol{V}_{it}\boldsymbol{\Gamma} + \epsilon_{it} \quad & \boldsymbol{\Gamma}\text{: } 100 \times 1 \\
\boldsymbol{V}_{it}^1 = a\left(\gamma^2 + \boldsymbol{V}_{it}^2\boldsymbol{\Gamma}^2\right) \quad & \boldsymbol{\Gamma}^2\text{: } 100 \times 100 \\
\vdots \quad & \\
\boldsymbol{V}_{it}^{10} = a\left(\gamma^{10} + \boldsymbol{Z}_{it}\boldsymbol{\Gamma}^{10}\right) \quad & \boldsymbol{\Gamma}^{10}\text{: } 1800 \times 100
\end{aligned}
$$

- *Identical* to parametric model, with addition of 100-node neural network layer
- 10 layers, 100 nodes each. **270042 parameters (!!!)** Regularization is extremely important.
- Activation is the "leaky ReLU" $a(x) = x$ if $x > 0$ else $x/100$

Hyperparameter optimization

Many hyperparameters:

- λ, dropout probability, minibatch size, gradient step size, etc.
- Grid-search cross-validation is expensive, and doesn't avail itself of benefits of model averaging

Instead, for each of B bootstrap samples of unique years, define bootstrap sample data D_{is} and out-of-bag D_{oob}. If hyperparameters are η, fit

$$\underset{\eta}{argmin} \|y_{oob} - y^*_{oob}\|_2$$

where

$$y^*_{oob} = \mathcal{M}_\eta \left(D_{oob}, \hat{\theta}\right)$$

and model $\mathcal{M}$ is fit by

$$\underset{\theta}{argmin} \|y_{is} - \mathcal{M}_\eta \left(D_{is}, \theta\right)\|_2$$

Final prediction is average of each of B models – e.g.: bootstrap aggregation or "bagging."

Data

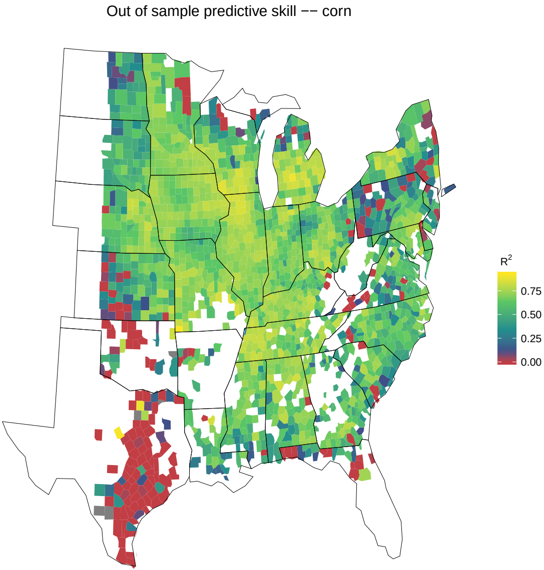
- Maize yield data from states comprising the US corn belt, from NASS, 1979 - 2016
- Historical weather data from METDATA (Abatzoglou 2013)
  - Daily observations of min/max temperature, min/max relative humidity, precipitation, wind speed, insolation
  - 4km resolution
  - Aggregated to counties and weighted by agricultural area
- Climate scenario data from Multivariate Adaptive Climate Analogs (MACA) dataset (Abatzoglou 2012)
  - Statistical downscaling of CMIP5 climate model runs, for RCP4.5 and 8.5
  - Same variables, resolution, processing
- Other variables: soil, time, lat/lon, proportion irrigated

Results: Predictive Skill

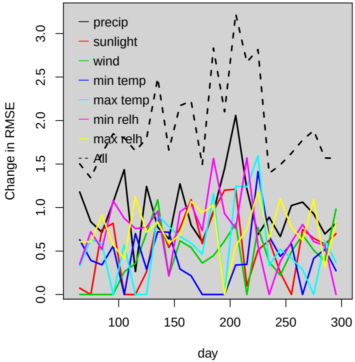
Model	Bagged	$\widehat{\text{MSE}}_{oob}$
Parametric	no	367.9
Semiparametric neural net	no	292.8
Parametric	yes	334.4
Fully-nonparametric neural net	yes	638.6
Semiparametric neural net	yes	251.5

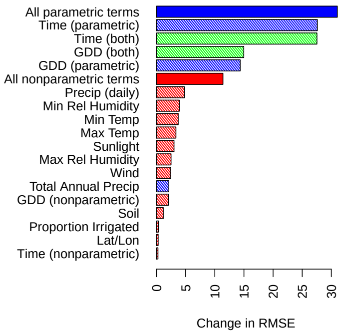
- Semiparametric neural nets outperform parametric model and regular neural net
- Improved skill derives both from semiparametric structure and from bagging

Model fit – Corn



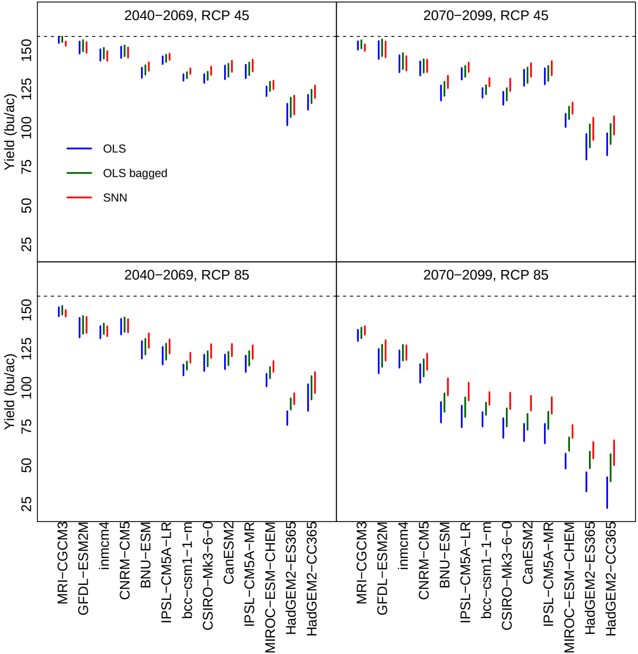
Variable Importance





Permutation importance: computed by randomly permuting input variables, predicting, and comparing resultant MSE against baseline

Projections by GCM



Application 2

United States Department of Agriculture

Economic
Research
Service

Economic
Research
Report
Number 266

July 2019

Climate Change and Agricultural Risk Management Into the 21st Century

Andrew Crane-Droesch, Elizabeth Marshall,
Stephanie Rosch, Anne Riddle, Joseph Cooper, and
Steven Wallander

Research Question

- How might climate change affect the cost of the Federal Crop Insurance Program?
  - And US Agricultural risk management programs more broadly (not just FCIP)
- What aspects of climate change impacts on agriculture are most important for the cost of the FCIP?
  - Price levels? Yield risk? Price risk?
  - Important to focus on drivers, given that the structure of the program could change over time.
- How might farmer adaptation change the government's fiscal exposure?

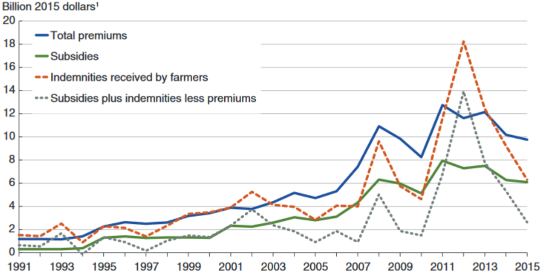
Cost of FCIP increasing over time

- FCIP premiums are actuarially-fair, but subsidized to increase participation
- Costs to government consist of subsidies plus indemnities, minus premiums
- Costs have averaged ~$7B/yr over the past 10 years
- Costs increase with acreage insured and spike with adverse weather
  - 2011-2013
  - 2008
  - Possibly this year (TBD)

Figure 3

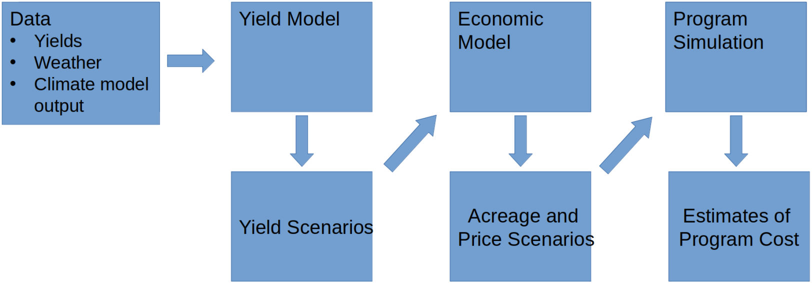
Federal crop insurance subsidies, indemnities, and premiums for crop years 1991-2015

Subsidies and premiums increased slowly in the 1990s and then rapidly in the mid-2000s before declining between 2013 and 2015

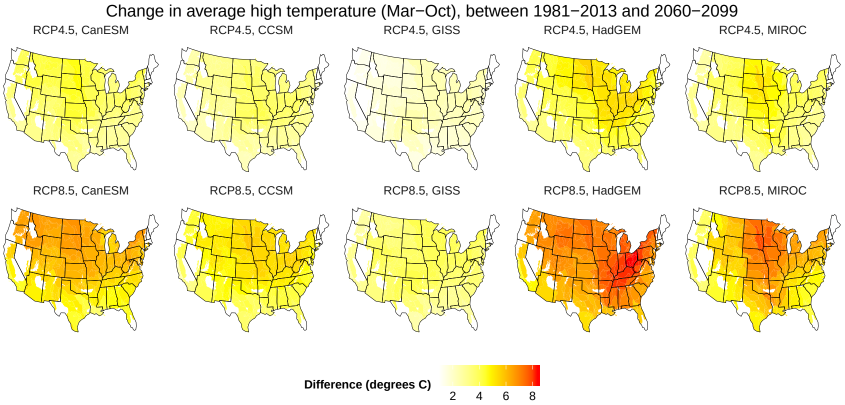


[1]Total premiums, subsidies, and indemnity payments are expressed in 2015 dollars using the gross domestic product (GDP) chain-type price index to adjust for price changes.
Source: USDA, Risk Management Agency, Federal Crop Insurance Corporation, Summary of Business Reports and Data, 1989-93, 1994-2003, 2004-13, and 2014-17 (crop years to date).

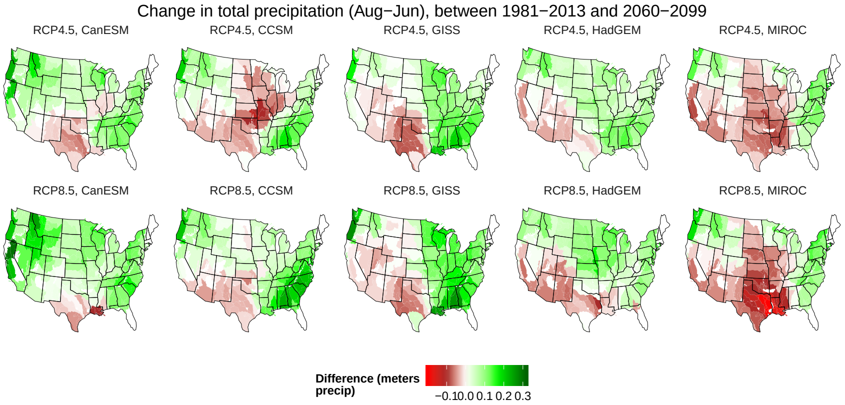
Model approach



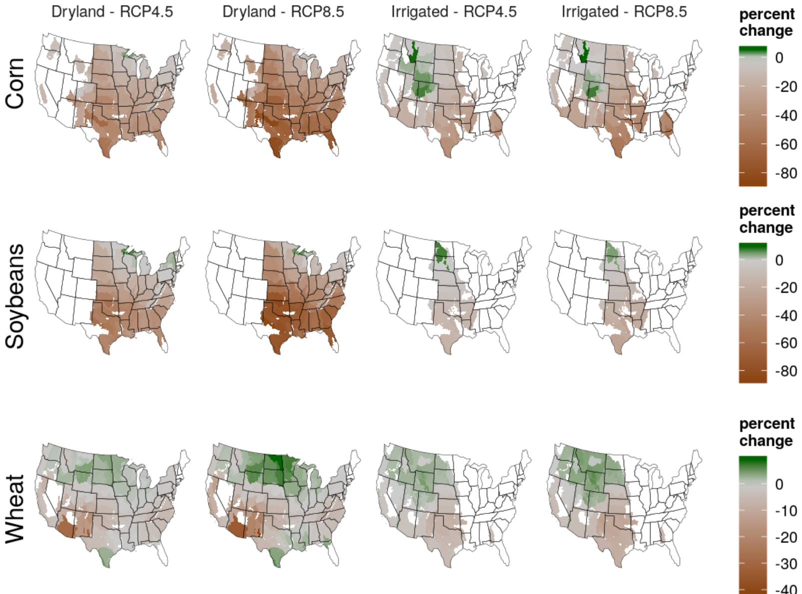
Inter-model variability in severity and pattern of warming



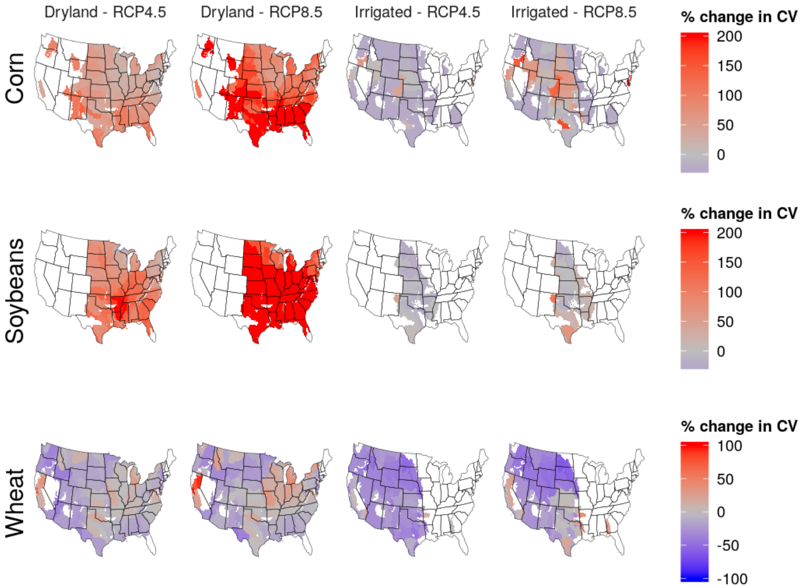
Inter-model variability in changes to precip



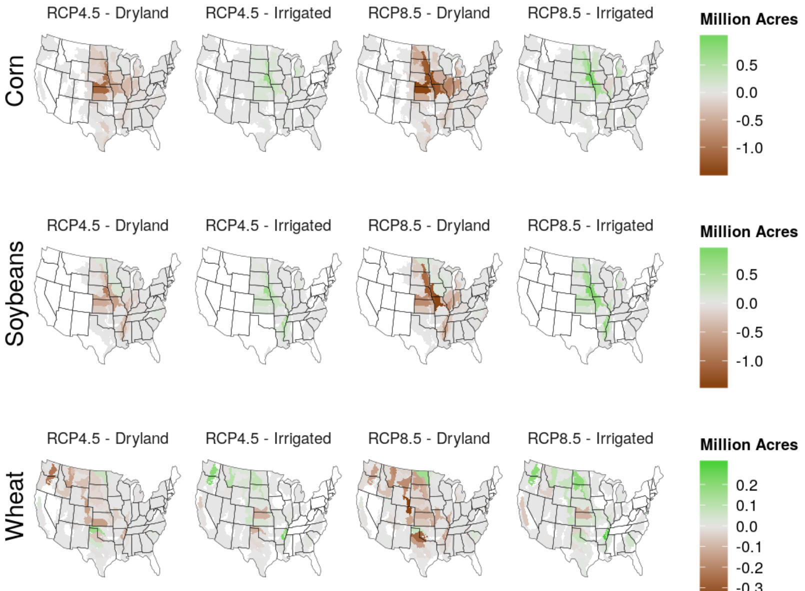
Scenarios: Large negative yield response, especially in dryland production, compared to today's climate



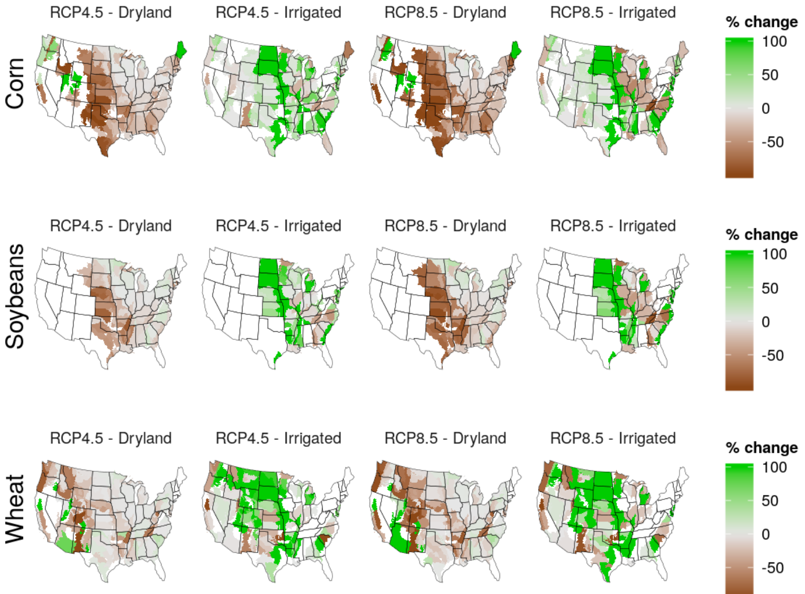
Changes in yield risk (proxied by CV) also large and substantial



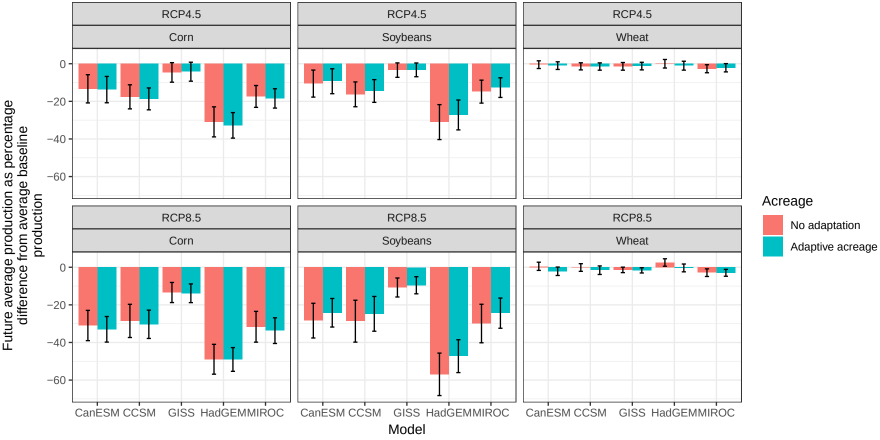
In the scenarios, effects on yields drives acreage out of dryland and into irrigated production



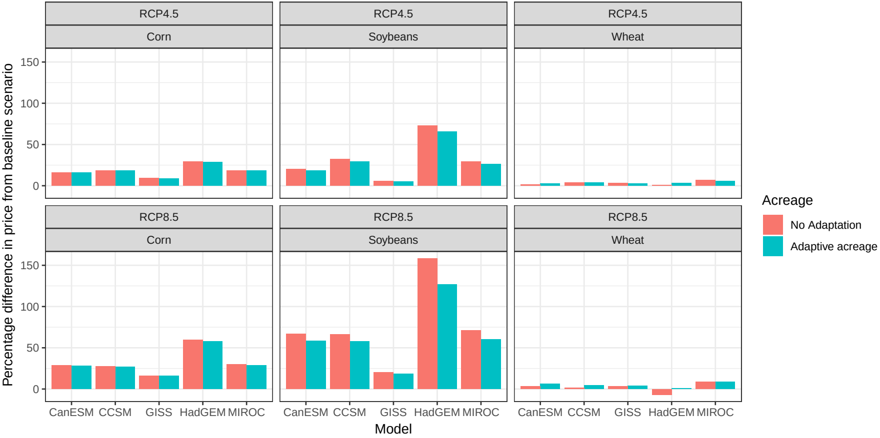
Scenarios: Climate change's effect on yield concentrates production into core of corn belt, and into irrigable areas



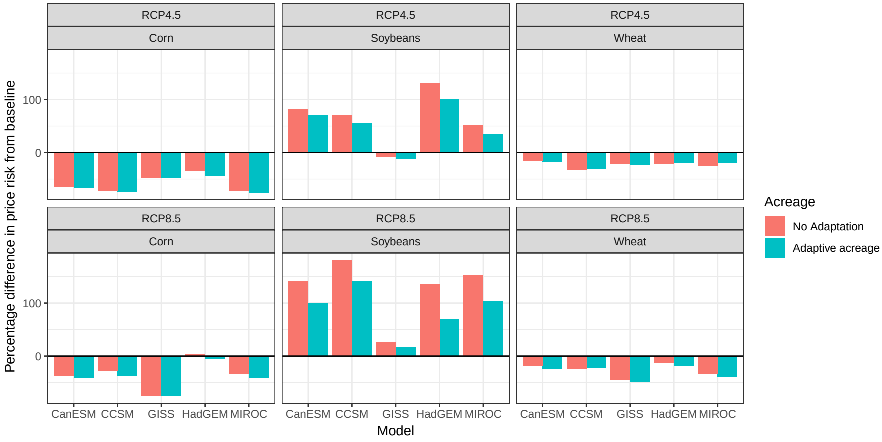
In the scenarios, climate change will reduce total production compared to a future with today's climate



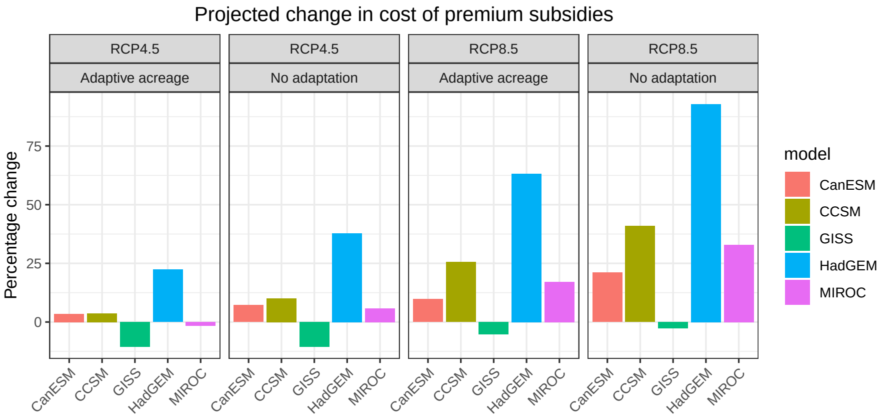
Scenarios: Impact of climate change on production will increase prices



Scenarios: Climate change impact on price risk varies by crop



Higher prices, more yield risk, and changed price risk increase fiscal exposure under most scenarios



Drivers of Change

- Scenarios are imperfect – Not possible to model everything
- Important to understand the mechanisms underlying our result

		1% increase in:		
		Yield risk	Price risk	Price levels
		Would lead to a ___% increase in average premiums		
Average premiums	Corn	0.7%	1.4%	1.5%
	Soybeans	0.5%	0.5%	0.9%
	Winter Wheat	0.7%	0.5%	1.1%

- Corn and soybean prices projected to increase
- Corn and soybean yield risk projected to increase
- Mixed signals on price risk
- Less impact on wheat

Conclusions of the Machine Learning Application

- Many factors that are **not** in our models could change with the climate
  - Examples include pests, disease, seeds
- Cost of agricultural risk management likely to increase with severity of climate change in the scenarios
- The principle driver likely to be average prices
- Yield volatility also important
- Price risk is an important driver of cost, but its direction of change is uncertain
- Adaptation to climate change through changes to planted acreage will reduce cost increases, but not completely