A GNN-RNN Approach for Harnessing Geospatial and Temporal Information: Application to Crop Yield Prediction

Joshua Fan*, Junwen Bai*, Zhiyun Li*, Ariel Ortiz-Bobea, Carla Gomes
Cornell University
NeurIPS 2021 Workshop, Tackling Climate Change with Machine Learning

**equal contribution*

Climate change effects on crop yield

- Crop production is extremely sensitive to fluctuations in climatic factors such as temperature, precipitation, soil moisture (Ortiz-Bobea, 2018)
    - For example, extreme heat or drought severely hurts crop growth
    - Climate change has already reduced agricultural productivity growth by 21% (Ortiz-Bobea, 2021)
- To adapt to these effects, we would like to **forecast** crop yields in advance
    - Important for food security, supply stability, seed breeding, economic planning

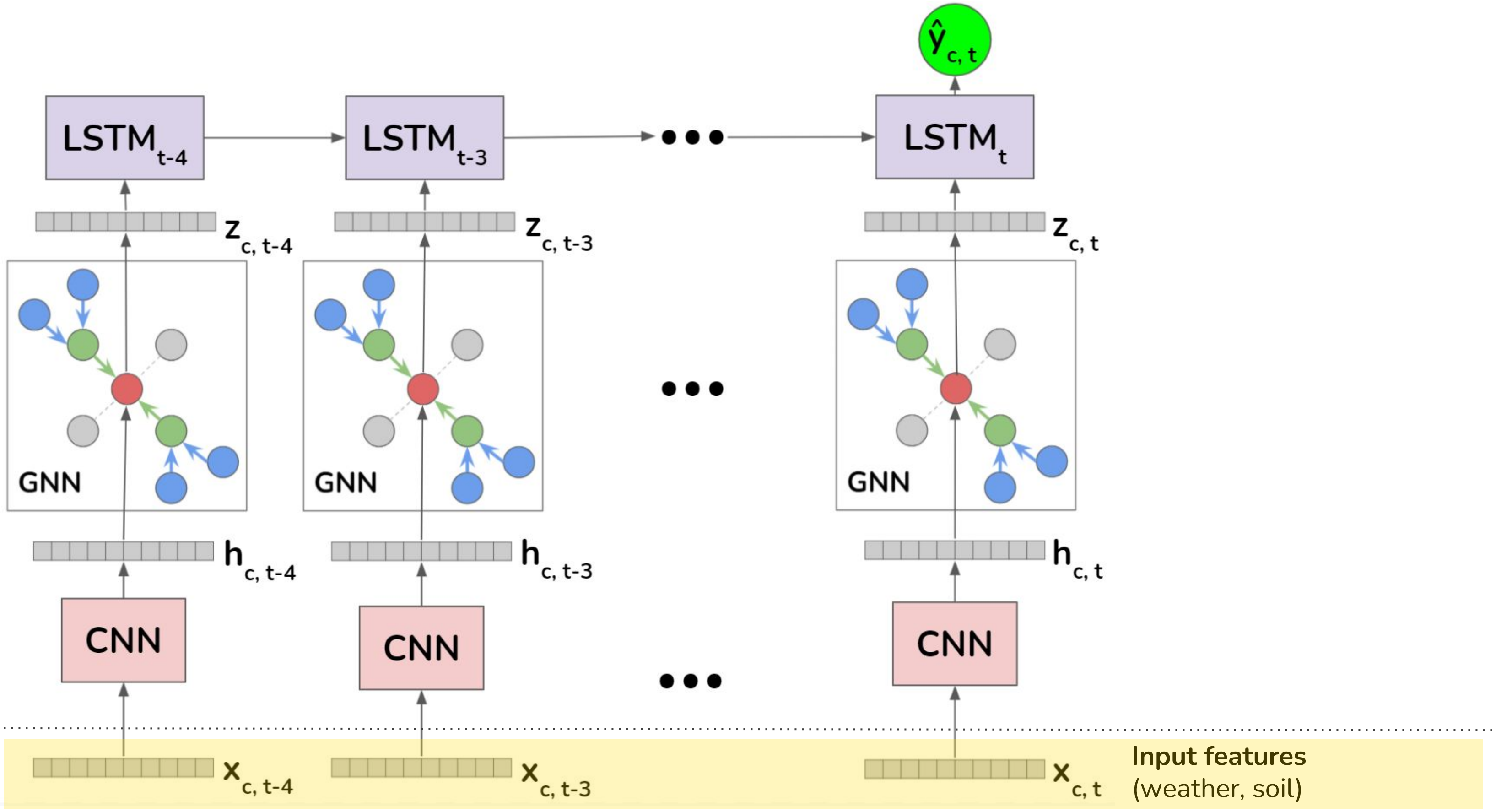
Goal: forecast crop yields in advance

- Use weather and soil data to forecast crop yields, preferably in the middle of the growing season (before harvest)
- For this study, we focus on corn and soybean yields for US counties
- Example features for each county/year: temperature, precipitation, humidity, soil moisture, soil temperature, soil quality features (e.g. available water capacity, bulk density and electrical conductivity, pH, and organic matter)
  - Weather features have a value for each week
  - Aggregated to county-level

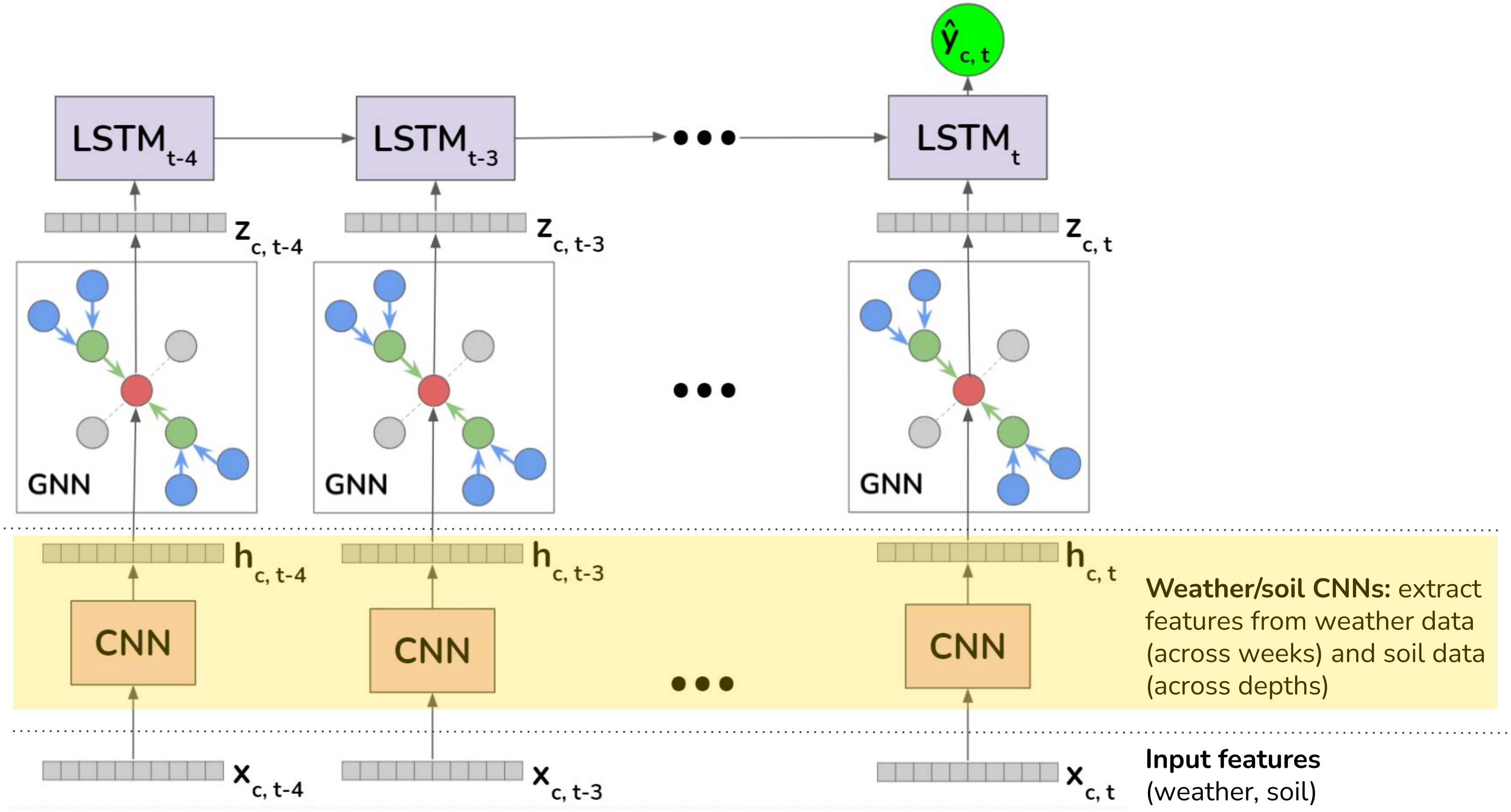
Existing work on crop yield prediction

- Most existing work uses standard feature-based supervised learning methods
  - Neural networks, tree-based methods (e.g. decision tree, random forest), linear regression
- A recent paper (Saeed & Khaki, 2021) introduced a **CNN-RNN** framework to extract features across different temporal scales (weekly/yearly) and across soil depth layers
- However, they do not make use of **geospatial** information
  - Example: nearby counties are correlated and have more in common
  - Should make use of this geospatial and temporal structure!

Our GNN-RNN framework



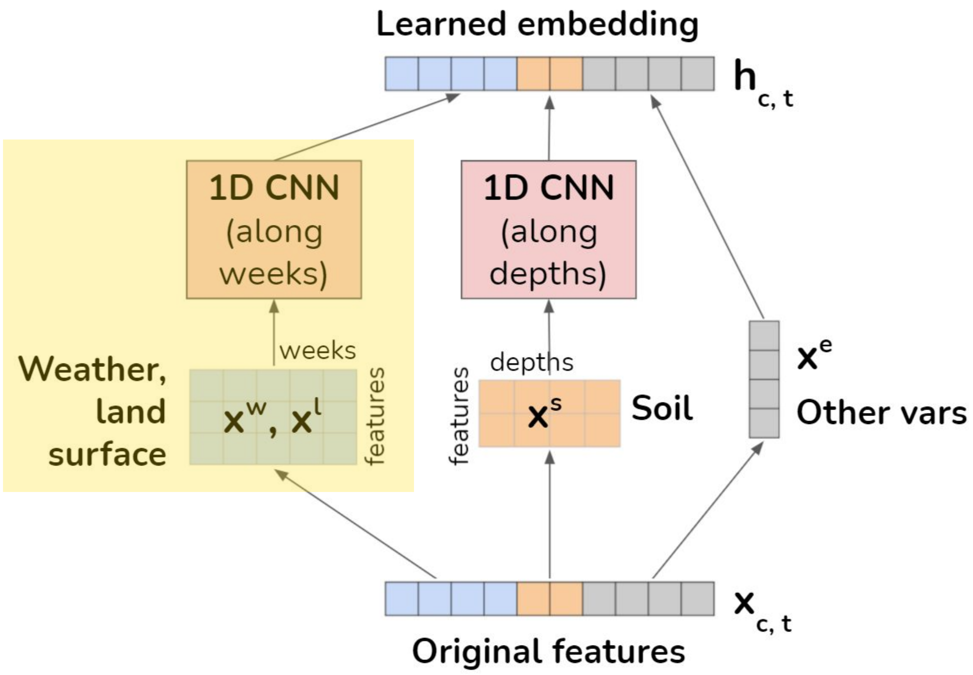
Our GNN-RNN framework



Extracting features within a year

For a **single datapoint (one year and county):**

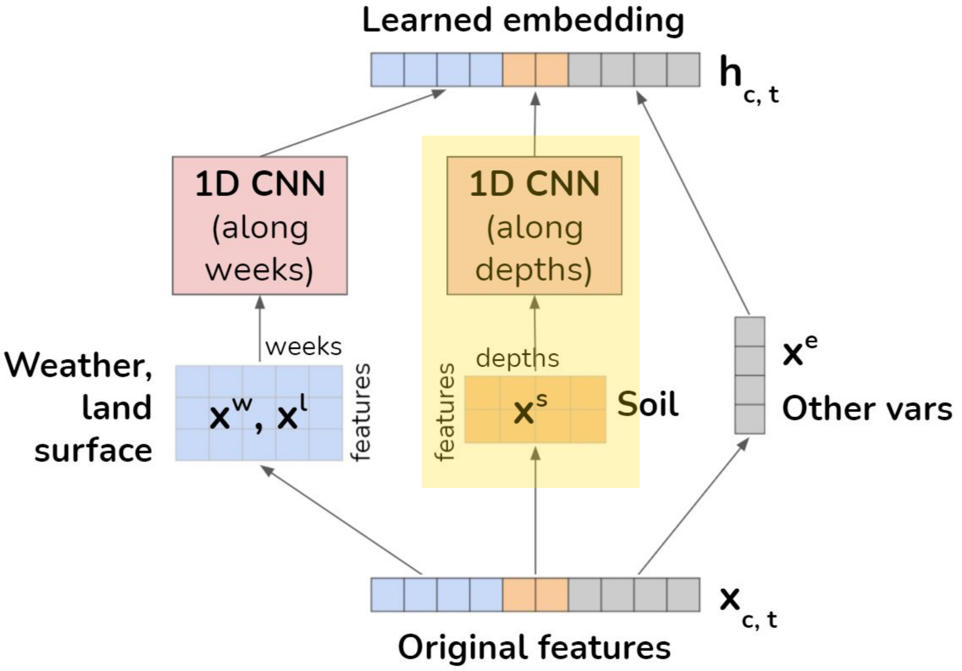
- Use 1D CNN to extract features from **weekly time-series of weather and land surface features**
  - Temperature, precipitation, soil moisture, etc



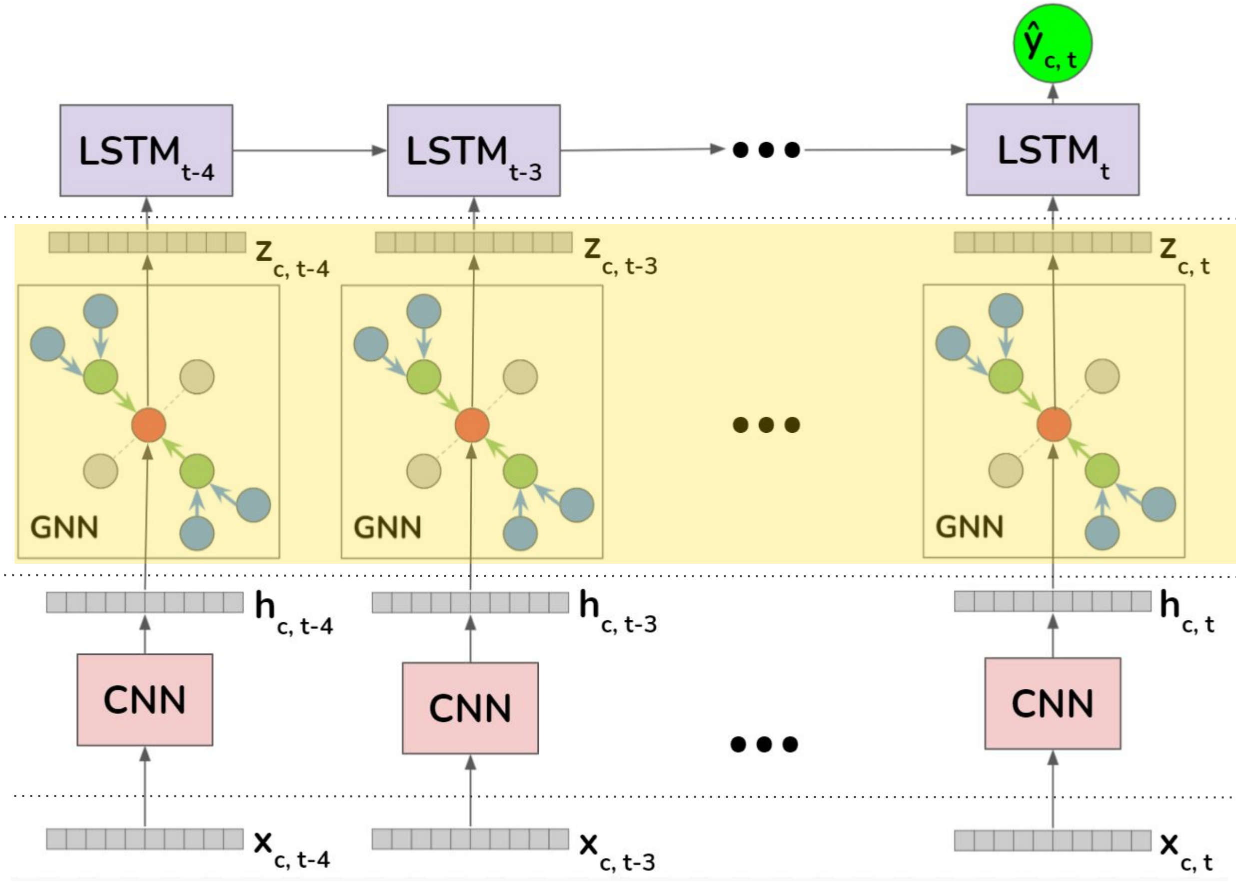
Extracting features within a year

For a **single datapoint (one year and county):**

- Use 1D CNN to extract features from **weekly time-series of weather and land surface features**
    - Temperature, precipitation, soil moisture, etc
- Use 1D CNN to extract features from **soil data** (varying across depths)



Our GNN-RNN framework



Graph neural network: aggregate info from nearby counties

Weather/soil CNNs: extract features from weather data (across weeks) and soil data (across depths)

Input features (weather, soil)

Using graph neural networks

We use a graph neural network (GNN) method, GraphSAGE (Hamilton et al, 2018), to allow each county's feature representation to be informed by **geospatial context**

- Each county is a node, and edges connect neighboring counties

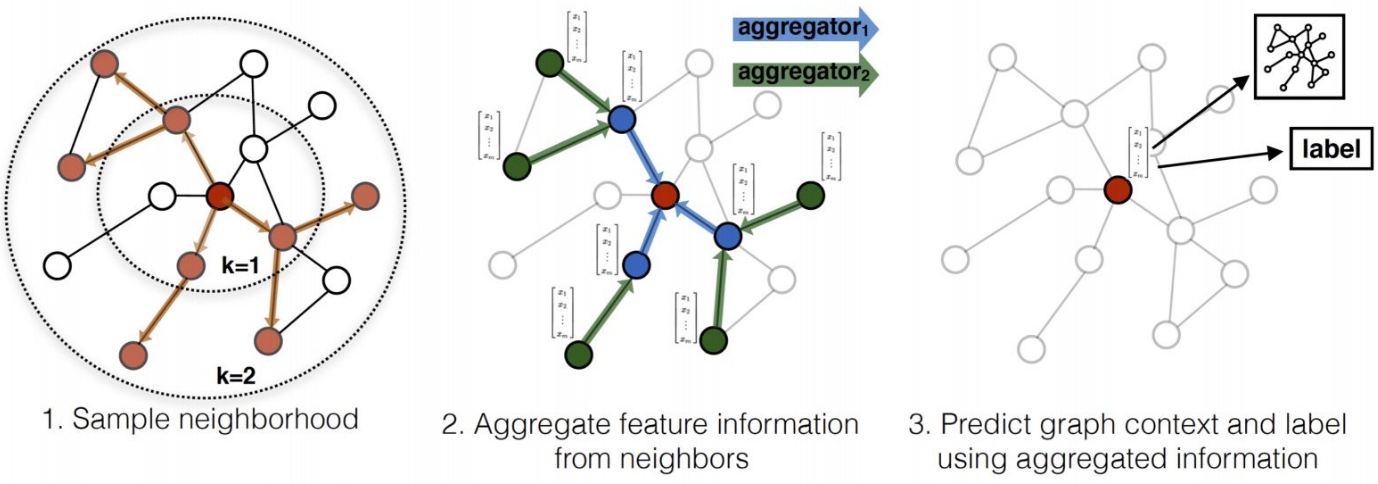
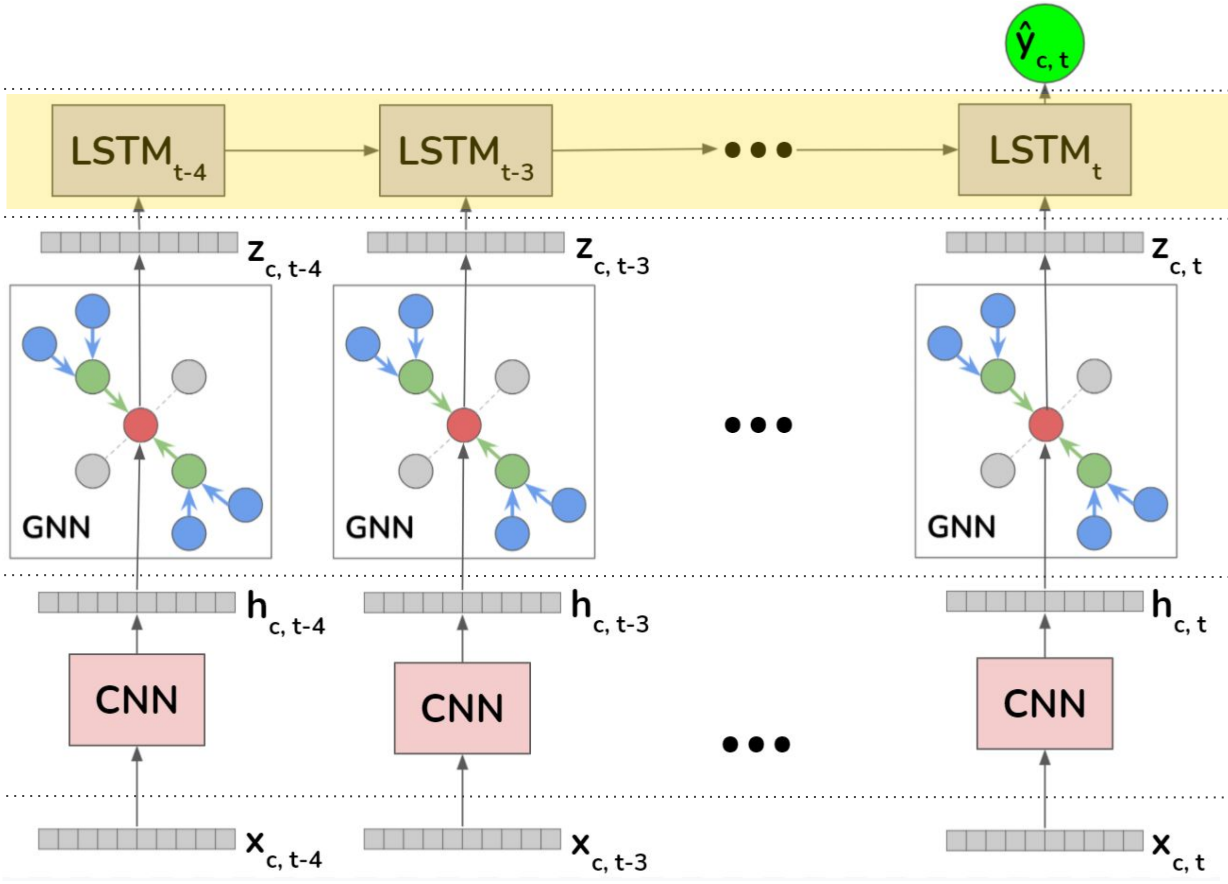


Figure from Will Hamilton et al. Inductive Representation Learning on Large Graphs. *Advances in Neural Information Processing Systems*, pages 1024–1034, 2017.

Our GNN-RNN framework



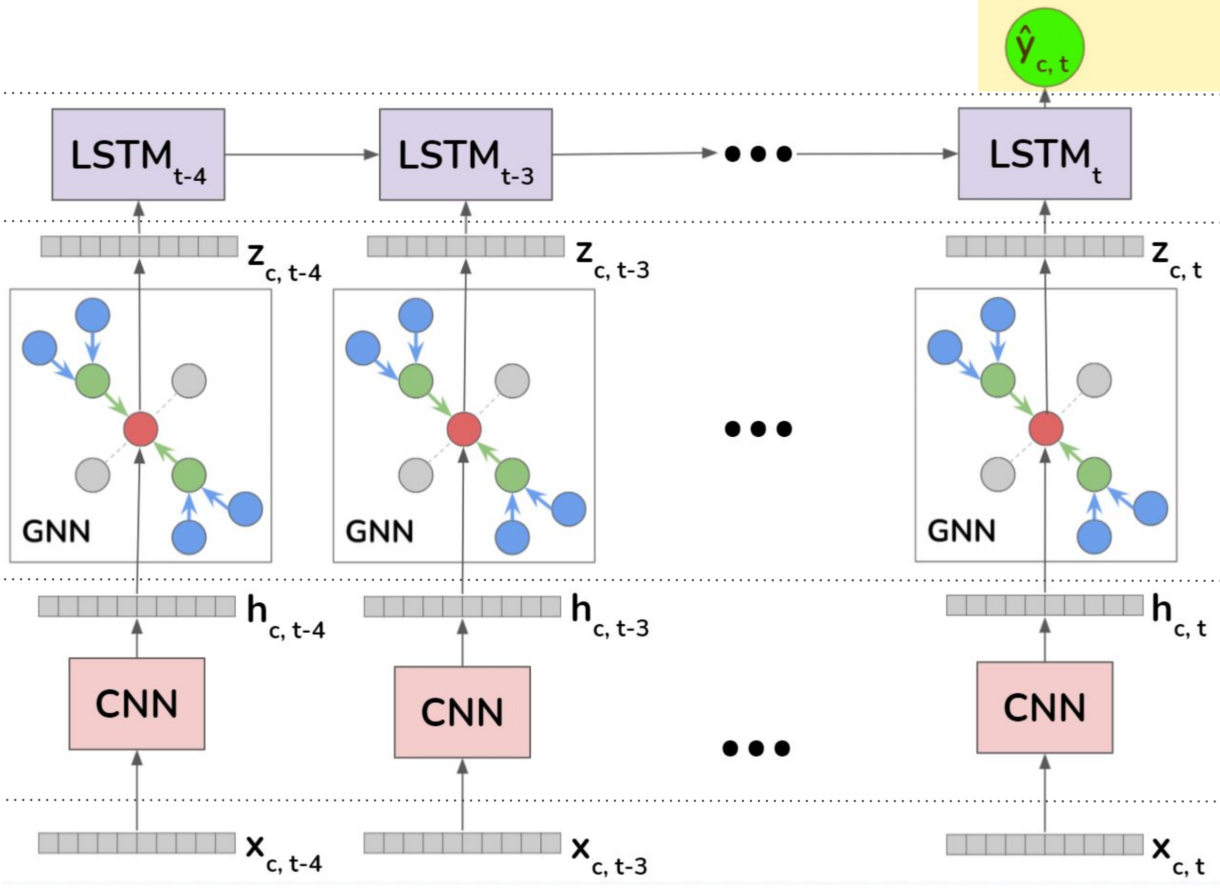
Recurrent neural network: process features from series of years

Graph neural network: aggregate info from nearby counties

Weather/soil CNNs: extract features from weather data (across weeks) and soil data (across depths)

Input features (weather, soil)

Our GNN-RNN framework



Predicted crop yield
(for given county, year)

Recurrent neural network: process features from series of years

Graph neural network: aggregate info from nearby counties

Weather/soil CNNs: extract features from weather data (across weeks) and soil data (across depths)

Input features
(weather, soil)

Results

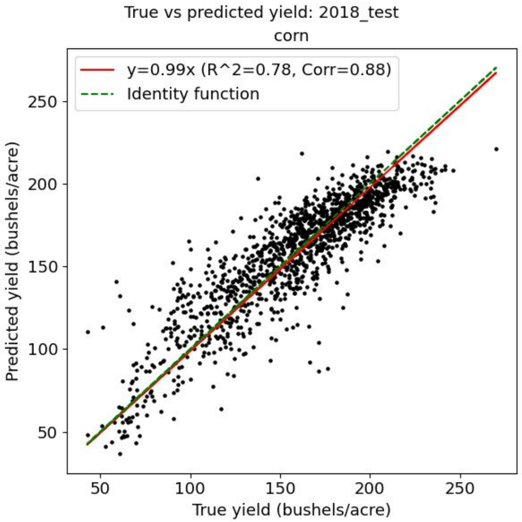
- We tested our model on (corn, soybean) for years (2018, 2019).*
  - For test year *t*, we train on years from 1981 to *(t-2)* inclusive, validate on year *(t-1)*, and test on year *t*
- Example result on 2018 corn:

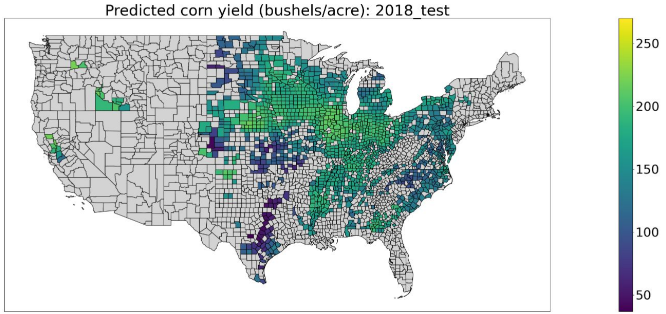
Method	RMSE	R^2	Corr
lasso 1y	0.7846	0.3839	0.7778
ridge 1y	0.9255	0.1428	0.7626
gradient-boosting 1y	0.7402	0.4516	0.7794
gru 1y	0.5938	0.6472	0.8158
lstm 1y	0.6146	0.6220	0.8303
cnn 1y	0.5824	0.6606	0.8235
gnn 1y **(ours)**	**0.4846**	**0.7517**	**0.8759**
gru 5y	0.6765	0.5419	0.8194
lstm 5y	0.6542	0.5716	0.8060
cnn-rnn 5y	0.5511	0.6936	0.8425
gnn-rnn 5y **(ours)**	**0.4900**	**0.7595**	**0.8731**
(std)	*(0.0191)*	*(0.0186)*	*(0.0092)*

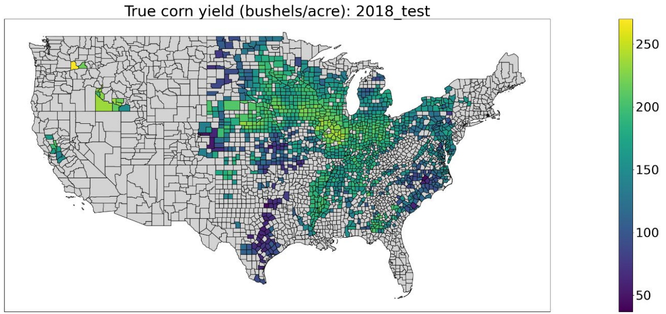
(a) 2018 corn results

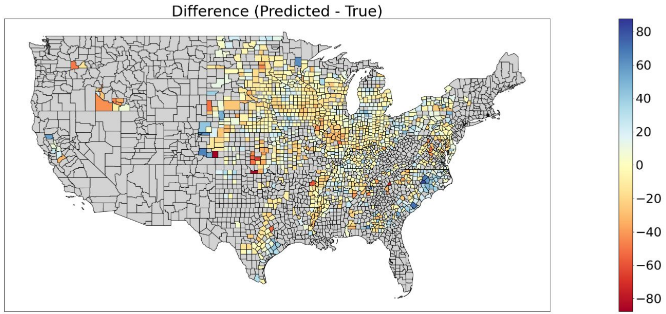
- For methods that only use a single year: GNN outperforms other single-year methods
- For methods that use a 5-year sequence: GNN-RNN outperforms all other methods

Example result









Early prediction results

- We are also interested in predicting in the middle of the year, before harvest.
- To simulate this, we choose a date (e.g. June 1). At test time, up to that date we use the actual features for the year, and after that date we replace the features with **historical averages**
- GNN-RNN framework still does best in this setting

Method	**RMSE**	R^2	**Corr**
lstm 1y	0.6347	0.5968	0.8148
cnn 1y	0.7253	0.4736	0.7004
gnn 1y **(ours)**	**0.5877**	**0.6543**	**0.8124**
lstm 5y	0.7004	0.5091	0.7708
cnn-rnn 5y	0.6532	0.5730	0.7732
gnn-rnn 5y **(ours)**	**0.5836**	**0.6591**	**0.8259**

Table 2: Early prediction results (2018 corn, after June 1).

Conclusion

- We developed a GNN-RNN framework to harness **geospatial** and **temporal** structure for crop yield prediction
  - Nearby counties are correlated and share information
- Outperforms existing methods across variety of crops and years
- Model's predictions could help us forecast the effects of climate change, allow for better adaptation, and facilitate humanitarian/economic planning to alleviate food security challenges

Acknowledgements

Funding sources:

- USDA cooperative agreement 58-6000-9-0041
- USDA NIFA Hatch project 1017421

Rich Bernstein - technical support, proofreading

Samuel Porter - help with processing gSSURGO soil quality data

References

Ariel Ortiz-Bobea, Toby R Ault, Carlos M Carrillo, Robert G Chambers, and David B Lobell. Anthropogenic climate change has slowed global agricultural productivity growth. *Nature Climate Change*, 11(4):306–312, 2021.

Ariel Ortiz-Bobea, Erwin Knippenberg, and Robert G Chambers. Growing climatic sensitivity of us agriculture linked to technological change and regional specialization. *Science Advances*, 1884(12):eaat4343, 2018.

Saeed Khaki et al. A cnn-rnn framework for crop yield prediction. *Frontiers in Plant Science*, 21710:1750, 2020