ML4CC: Lecture 5

Sit with your **new** discussion groups (1-10 left to right)!

Assignments reminder

Keep doing your weekly PMIRO+Q

Your second coding assignment is due Feb 29 before the start of class.

Recap of previous paper

P: Want to track glacier melting rate

M: Train a U-net to segment satellite images of glaciers and compare sizes over time

I: Fully automated method applied to the Alps, specific outlier filtering

R: Model has balanced, high performance in terms of precision and recall and can be used to measure glacier melt

O: Need to capture small glaciers, unclear if it will generalize to other glacier locations

Climate Change in the News

The Inflation Reduction Act, the 2021 U.S. climate law abbreviated IRA, primarily reduces emissions through financial incentives, rather than binding rules. But in addition to all its well-known carrots, lawmakers quietly included a smaller number of sticks — particularly when it comes to the potent greenhouse gas methane, which has proven to be a pesky source of increasing climate pollution with each passing year. New research suggests that those sticks could soon batter the oil and gas industry, which is responsible for a third of all methane emissions in the U.S.

An IRA provision directs the Environmental Protection Agency, or EPA, to charge $900 for every metric ton of methane above a certain threshold released into the atmosphere in 2024. The issue is particularly challenging to tackle in oil and gas fields because methane is the primary component in natural gas, and it leaks from hundreds of thousands of devices scattered across the country. In 2022, oil and gas facilities emitted more than 2.5 million metric tons of methane.

Grist

Biden's climate law fines oil companies for methane pollution. The bill is coming due.

Recent research suggests the IRA's methane fee could batter the oil and gas industry to the tune of more than $1 billion.

A new analysis by Geofinancial Analytics, a private data provider, found that some companies may be liable for tens of millions of dollars in fees — a possibility that could bankrupt some operators. The analysis, which relied on satellite data, found that the top 25 oil and gas producers in the country would together have been liable for as much as $1.1 billion if the methane fee was applied to emissions for a one-year period ending in March 2023.

Paper 4 Discussion

AtmoDist: Self-supervised Representation Learning for Atmospheric Dynamics

Sebastian Hoffmann
Dept. of Computer Science
Universität Magdeburg
sebastian1.hoffmann@ovgu.de

Christian Lessig
Dept. of Computer Science
Universität Magdeburg
christian.lessig@ovgu.de

Abstract

Representation learning has proven to be a powerful methodology in a wide variety of machine learning applications. For atmospheric dynamics, however, it has so far not been considered, arguably due to the lack of large-scale, labeled datasets that could be used for training. In this work, we show that the difficulty is benign and introduce a self-supervised learning task that defines a categorial loss for a wide variety of unlabeled atmospheric datasets. Specifically, we train a neural network on the simple yet intricate task of predicting the temporal distance between atmospheric fields from distinct but nearby times. We demonstrate that training with this task on ERA5 reanalysis leads to internal representations capturing intrinsic aspects of atmospheric dynamics. We do so by introducing a data-driven distance metric for atmospheric states. When employed as a loss function in other machine learning applications, this Atmodist distance leads to improved results compared to the classical ℓ_2-loss. For example, for downscaling one obtains higher resolution fields that match the true statistics more closely than previous approaches and for the interpolation of missing or occluded data the AtmoDist distance leads to results that contain more realistic fine scale features. Since it is derived from observational data, AtmoDist also provides a novel perspective on atmospheric predictability.

Attendance

Select one person from the group to go to this Google Doc and write down the names of all people present in the group (remember to mark who took attendance!)

https://docs.google.com/document/d/1PKhw9E2lJpAnFrFO88DOc2rscZFVlcIv47QNa5h1sGs/edit?usp=sharing (link is in Brightspace under Syllabus content)

Discussion Question 1

How did you feel about this paper?

Discussion Question 2

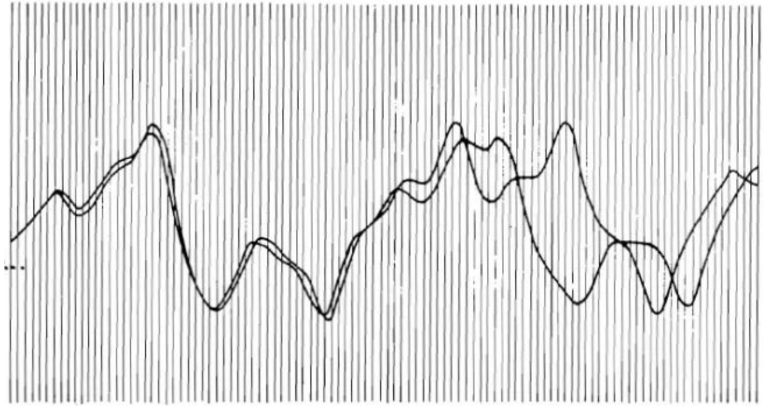
What does the highlighted sentence mean and how does it help motivate this work?

2.1 Geoscience

Distance measures for atmospheric states play an important role in classical weather and climate predictions. For example, ensemble methods require a well defined notion of nearby atmospheric states for their initialization. Various distance measures have therefore been proposed in the literature,

Ensemble climate models

To provide a distribution of possible future climate states, models need to be run multiple times with different but similar starting points. “Similar” is easy to define for a single scalar value like temperature, but not for complex high-dimensional states like atmospheric dynamics



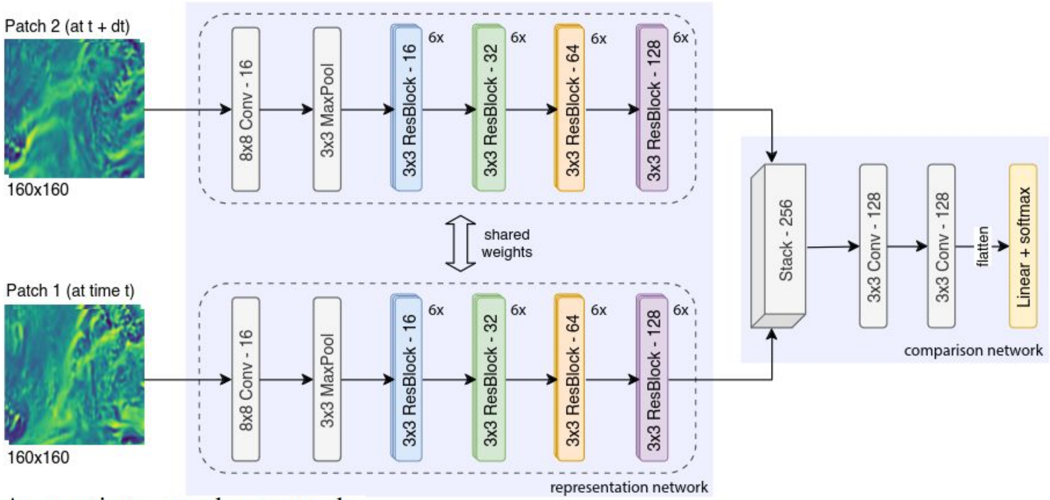
HOW TWO WEATHER PATTERNS DIVERGE. From nearly the same starting point, Edward Lorenz saw his computer weather produce patterns that grew farther and farther apart until all resemblance disappeared. (From Lorenz's 1961 printouts.)

Discussion Question 3

What self-supervised task is the model trained on and what specific loss function does it use? Give intuition why this task helps the model learn useful representations and why the loss function is appropriate

Pretext task: how many time points apart are these two images?

Intuition: if these two states happen close together in time they must be similar



two nearby states corresponds to an intrinsic distance between them. As a spatio-temporal pretext task for learning a distance measure for atmospheric dynamics, we thus use the prediction of the temporal separation between close-by states. More specifically, given two local patches of atmospheric states X_{t_1}, X_{t_2} centered at the same spatial location but at different, nearby times t_1 and t_2, the task for the neural network is to predict their temporal separation $\Delta t = t_2 - t_1 = n \cdot h_t$ given by a multiple of the time step h_t (3h in our case). The categorical label of a tuple (X_{t_1}, X_{t_2}) of input patches, each consisting of the vorticity and divergence field at the respective time $t_k = k \cdot h_t$ for the patch region, is thus defined as the number of time steps n in between them. Following standard methodology for classification problems, for each training item (X_{t_1}, X_{t_2}), our representation network predicts a probability distribution over the finite set of allowed values for n. Training can thus be performed with cross-entropy loss, which is known to be highly effective.

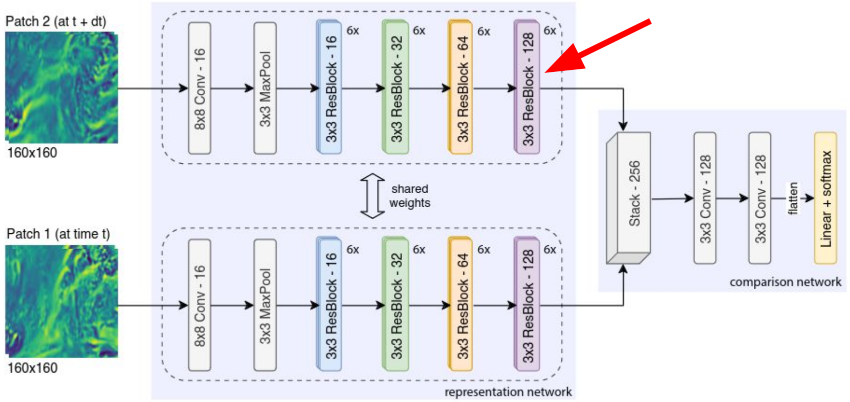
Is regular cross entropy the best loss function here? 🤔

Discussion Question 4

Which part of the model is used for the AtmoDist metric? What is the size of this representation and how does it compare to the size of the input data?

Learned representation

AtmoDist uses the final layer of the representation network

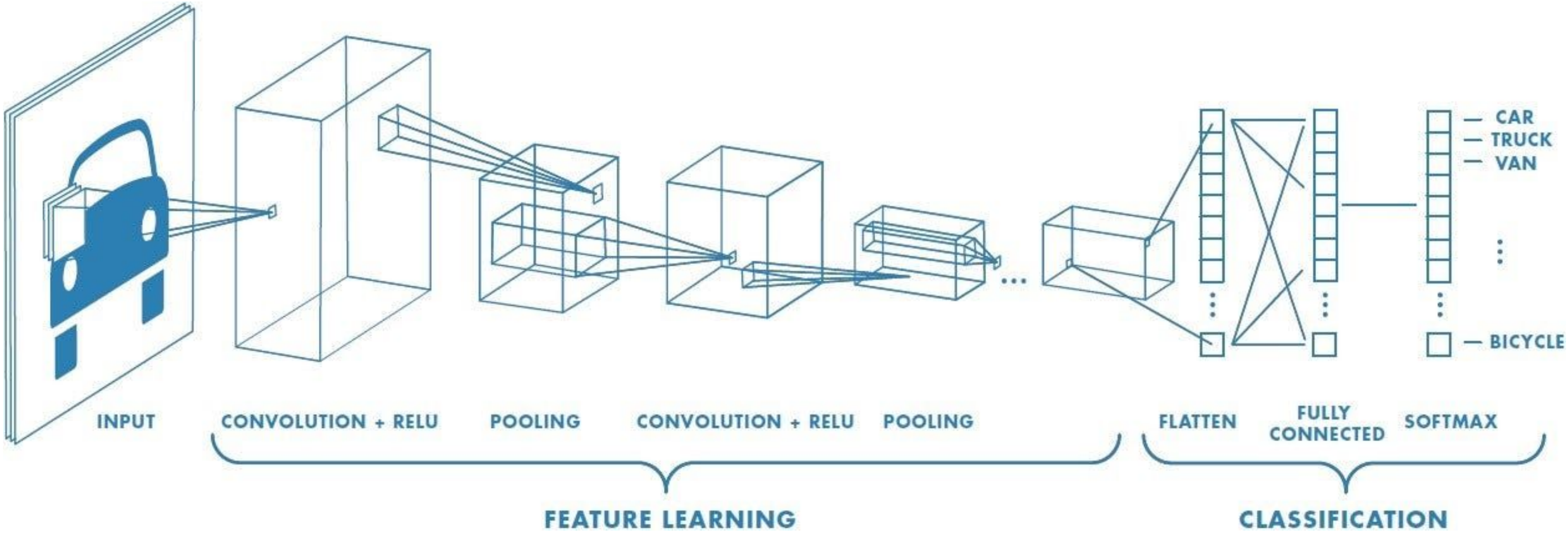


Divergence and vorticity plus spatial dimensions

128 features plus downsampled spatial dimensions

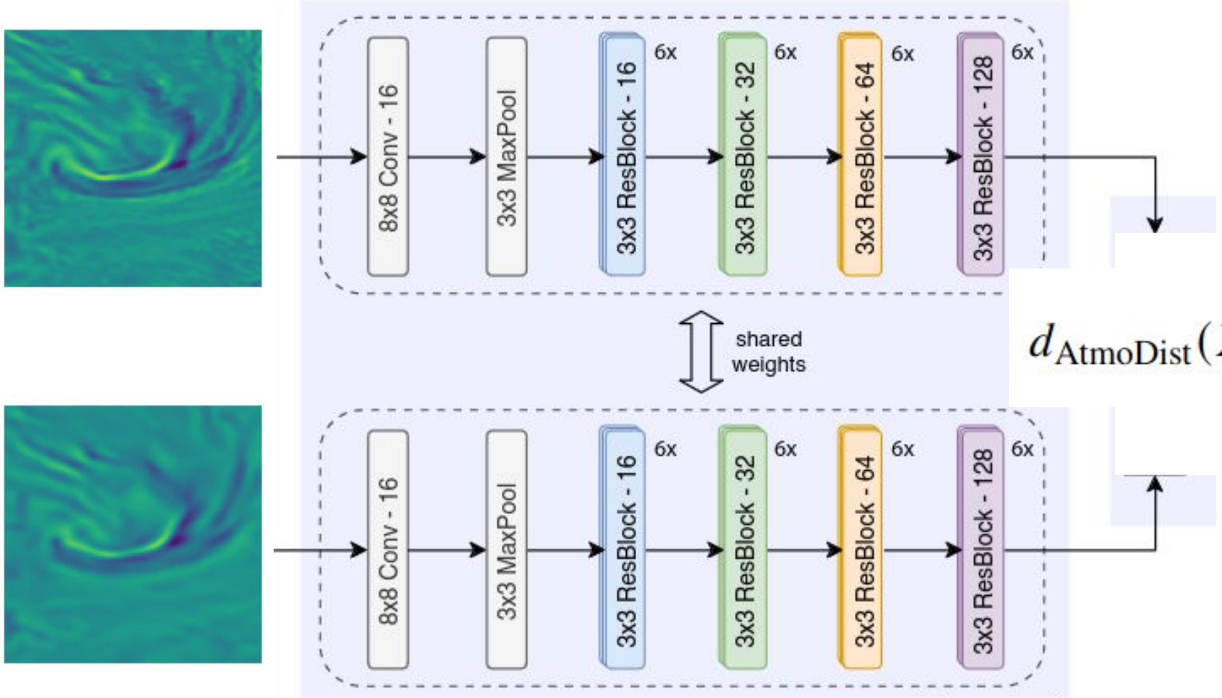
It maps an input X of size $2 \times 160 \times 160$ to a representation vector $\mathcal{F}(X)$ of size $128 \times 5 \times 5$. This corresponds to a compression rate of 16. The tail network is a simple convolutional network with a

Convolutional neural networks



Uses convolution and pooling operations

Distance Metric



$$d_{\text{AtmoDist}}(X_1, X_2) = \frac{1}{N}\left\| \mathcal{F}(X_1) - \mathcal{F}(X_2) \right\|^2$$

Discussion Question 5

Explain what figure 4 is showing in your own words.

Demonstrating superiority of AtmoDist representation

Intuitively, we would like the a measure of the difference between two inputs to grow as the time difference does, and be consistent.

Their measure (AtmoDist) is better than a basic pixel-wise loss function (l1 or l2) according to both overall shape and variability.

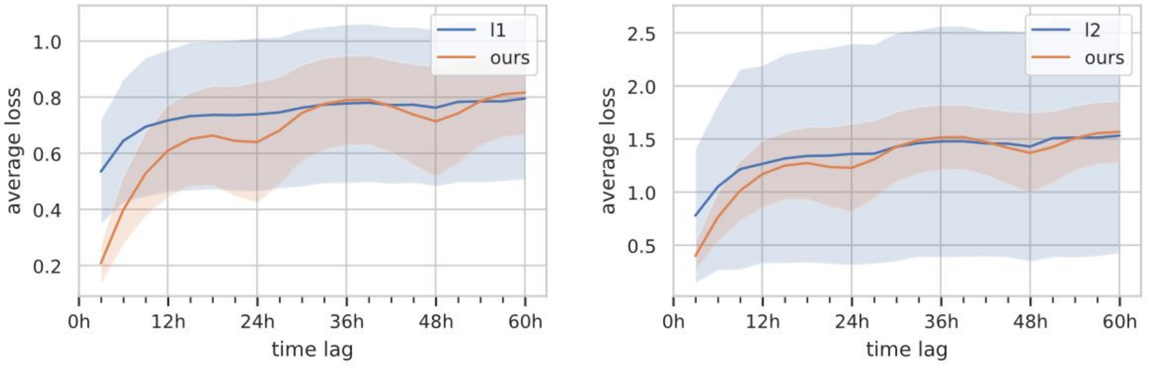
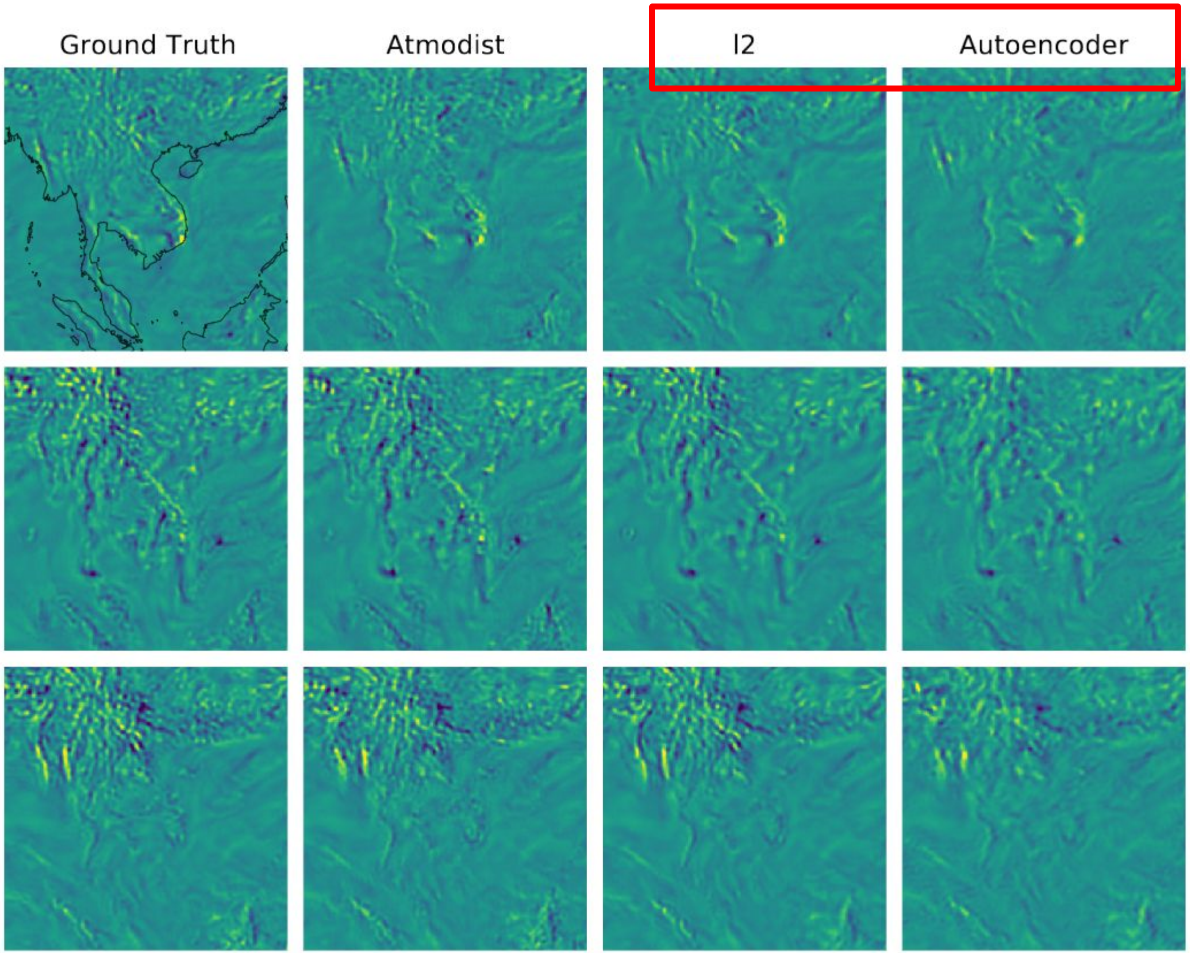


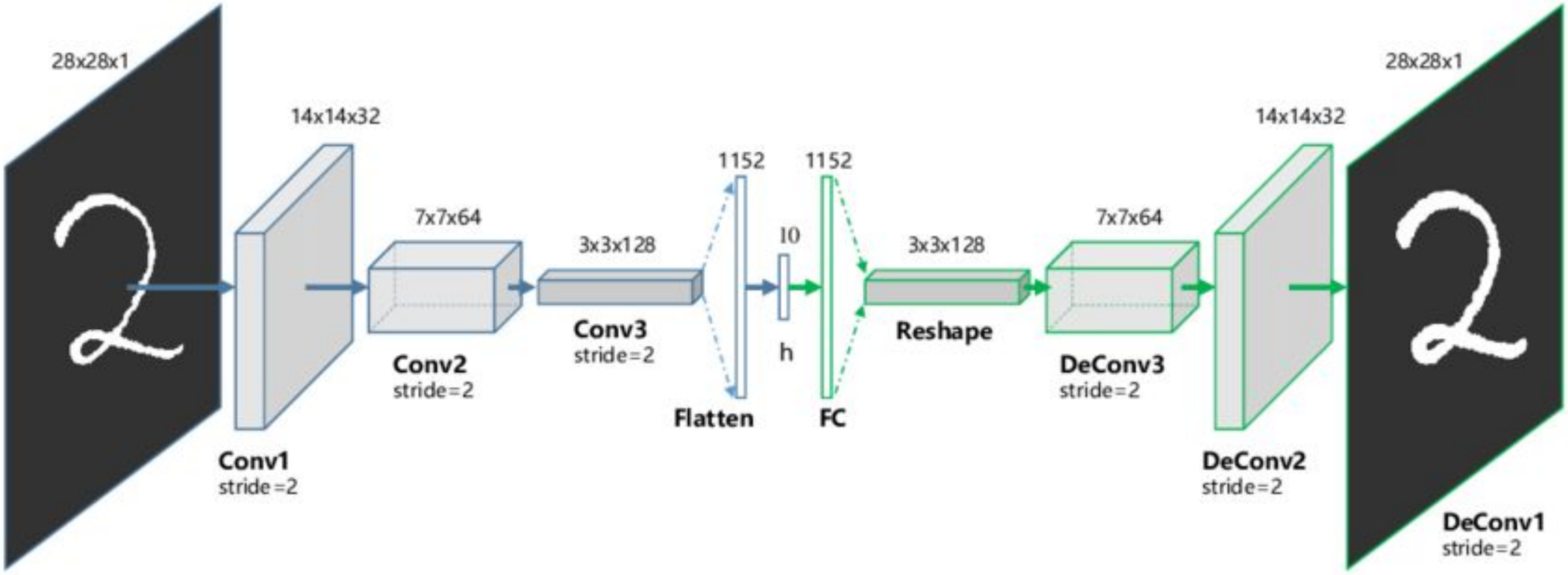
Figure 4: Mean ℓ_1-norm (left) and mean ℓ_2-norm (right) between samples that are a fixed time-interval apart, calculated on the training set. Shaded areas indicate standard deviation. For comparability, the AtmoDist distance has been normalized in each case with the method described in Appendix A.3. To give equal weight to divergence and vorticity, they have been normalized to zero mean and unit variance before calculating pixel-wise metrics.

Discussion Question 6

What other two content loss functions for the super resolution GAN was the AtmoDist loss compared to?


Ground Truth
Atmodist
l2
Autoencoder

Autoencoders are the original “self-supervised” models



Autoencoders are trained to reconstruct their input, but must pass through a “bottleneck” representation

Discussion Question 7

What did their “ablation study” show?

69h is the best maximum time difference to train on

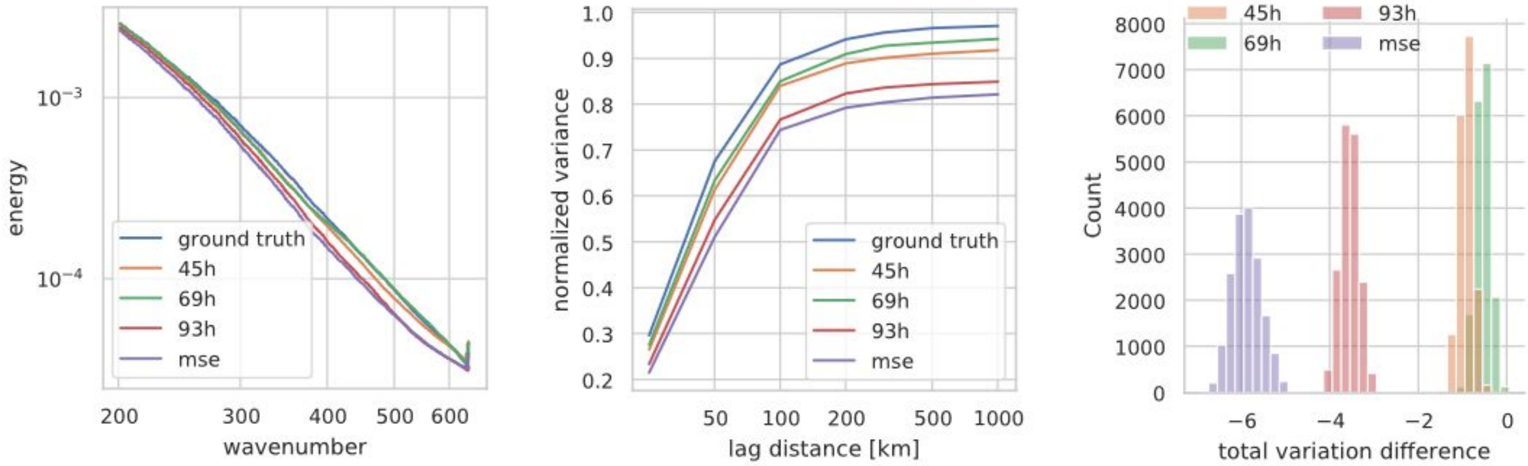


Figure 11: The energy spectrum (left), semivariogram (center), and distribution of total variation difference errors (right) for models trained with different maximum $\Delta t_{\max}$ for our ablation study. The semivariogram and error distributions are calculated on divergence, but qualitative similar results are obtained for vorticity.

Discussion Question 8

Share what questions you wrote in your PMIRO+Q and decide as a group what you’d like to ask.

Update your PMIRO+Q

Submit a second file to the Brightspace assignment (don’t overwrite the original):

It should:

Update your PMIRO as needed

Answer your own Q

You can be talking with your group during this!

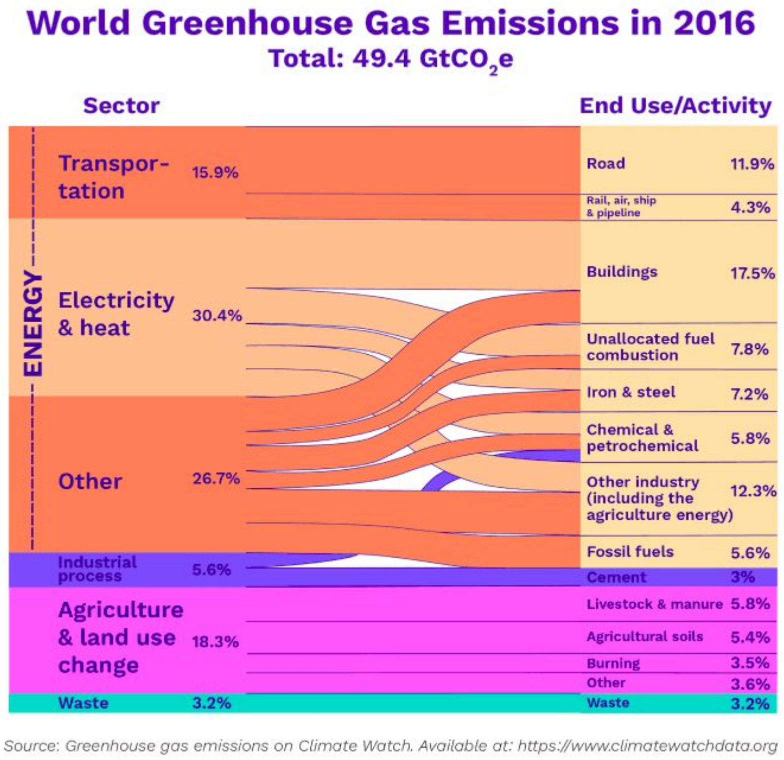
15 min break

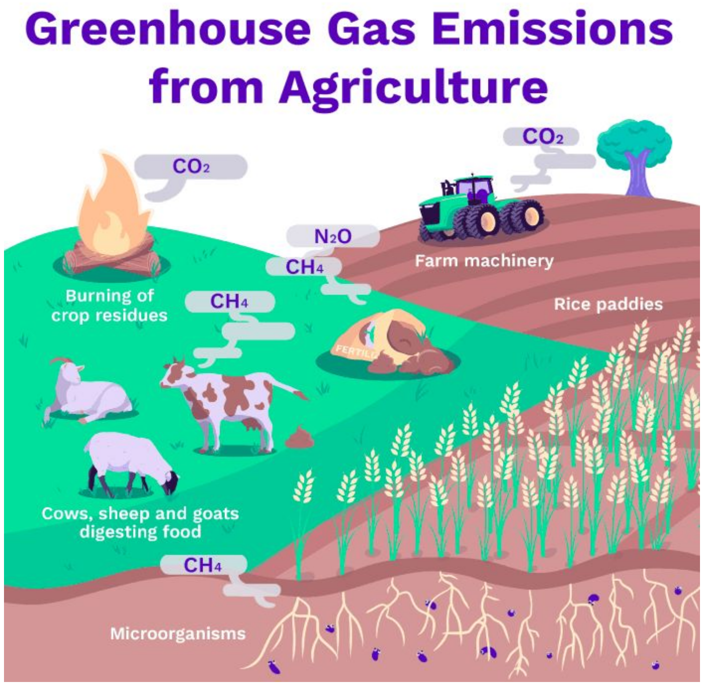
Lecture

Climate Change: food and agriculture

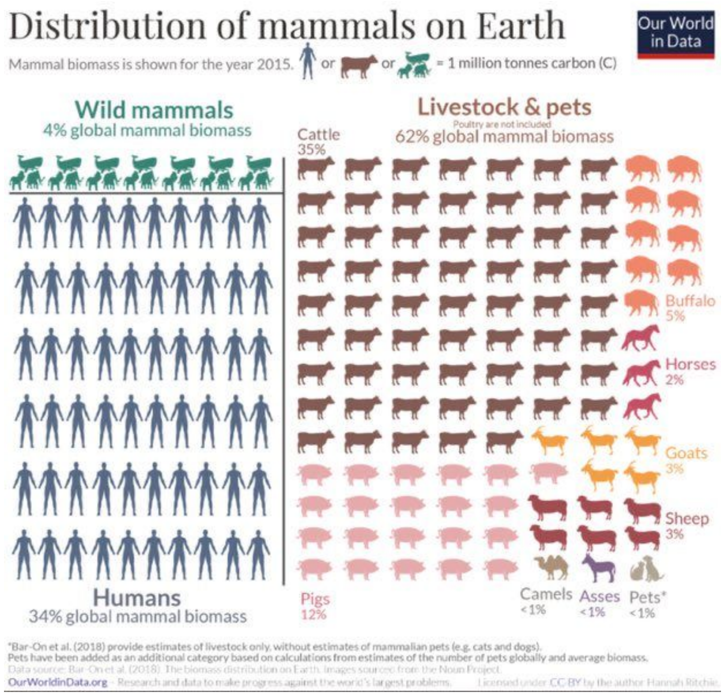
Machine Learning: time series, recurrent neural networks, and transfer learning

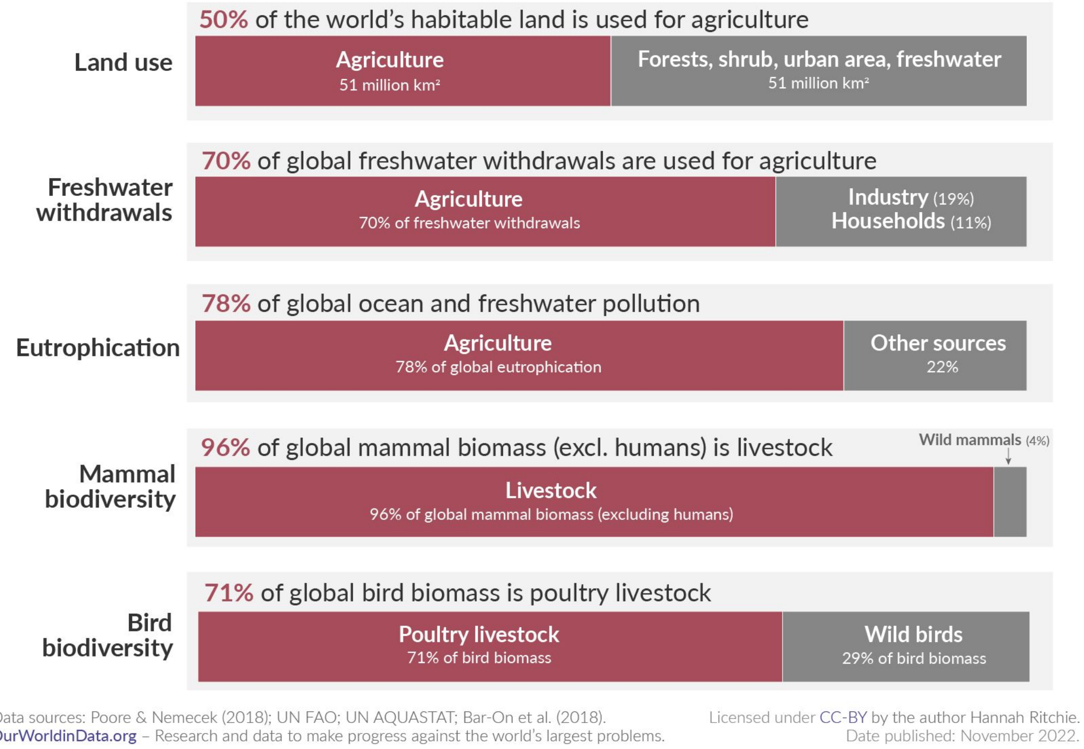
Which activities contribute to GHG emissions?





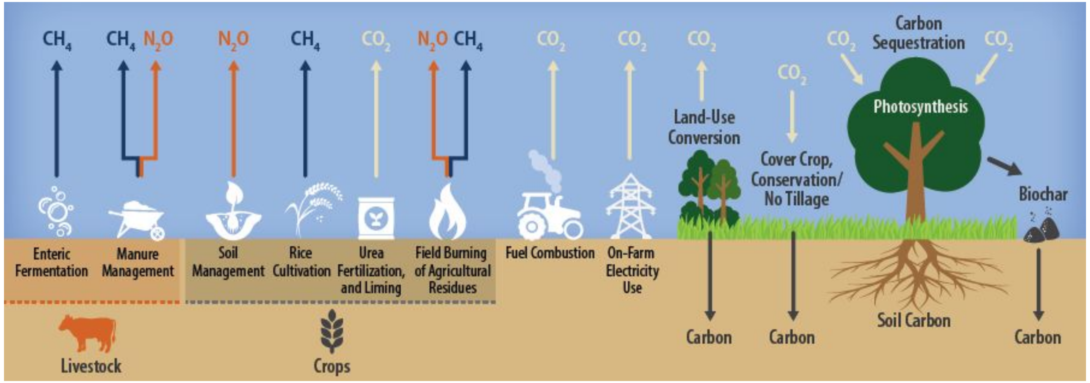
Agriculture, due to its scale, has a large impact on the planet



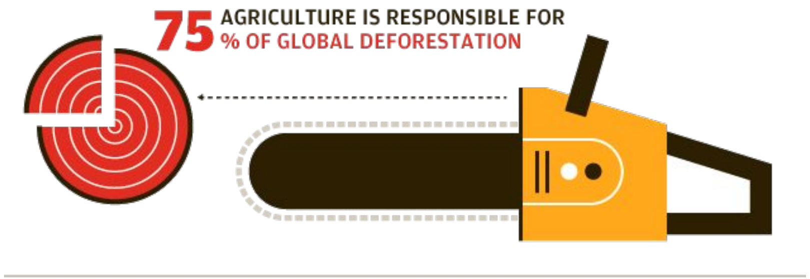


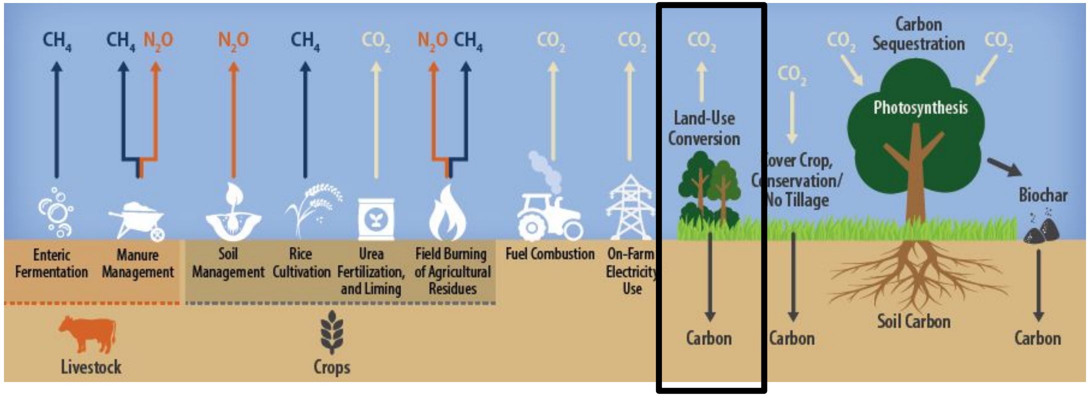
Agriculture has multiple impacts on GHGs

- Farming of plants and animals can capture CO2.
- However, modern farming practices lead to far more emissions than reductions in GHGs
- Furthermore, farming requires land that may have been used for other purposes
- GHGs other than CO2 are also released, which have higher warming potentials

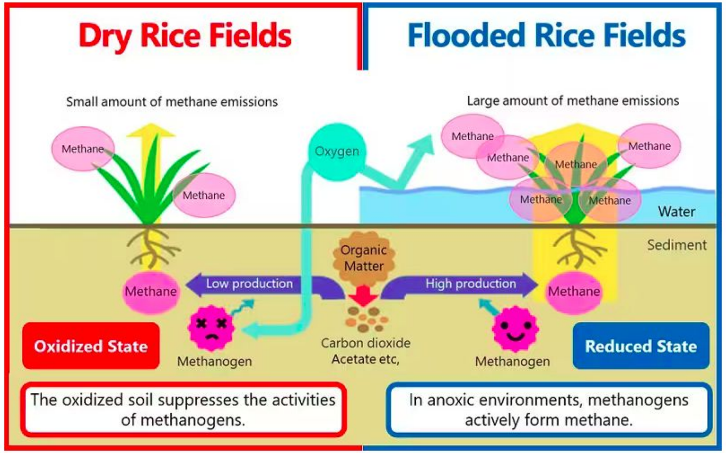


Forests are cut down for grazing land and farms

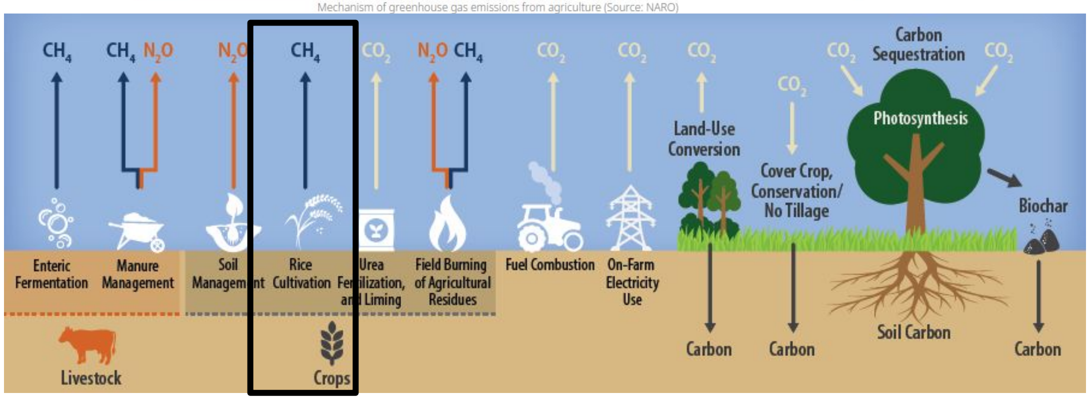




Flooding of rice paddies releases methane

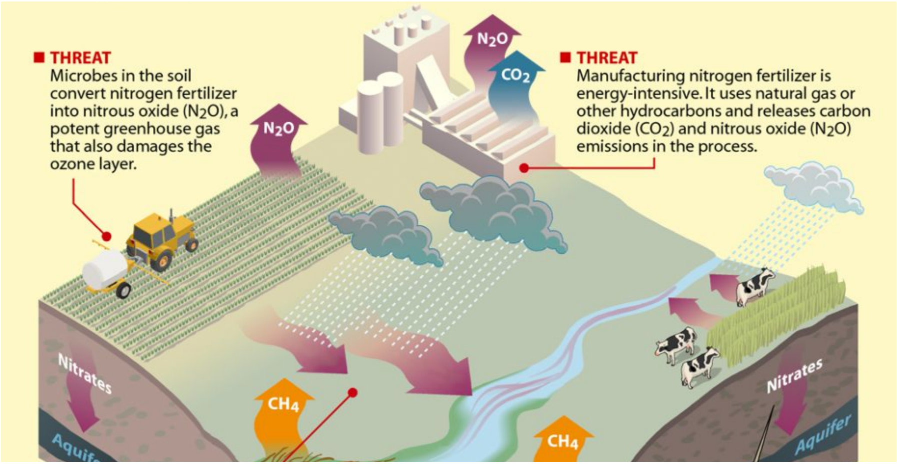


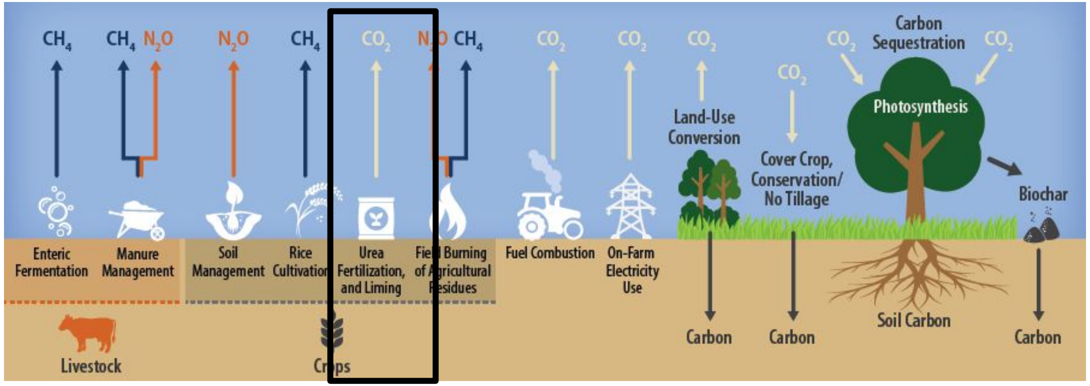
Mechanism of greenhouse gas emissions from agriculture (Source: NARO)



Rice paddies and their flooding is responsible for 10% of global methane release.

Manufacture and use of fertilizer releases multiple GHGs

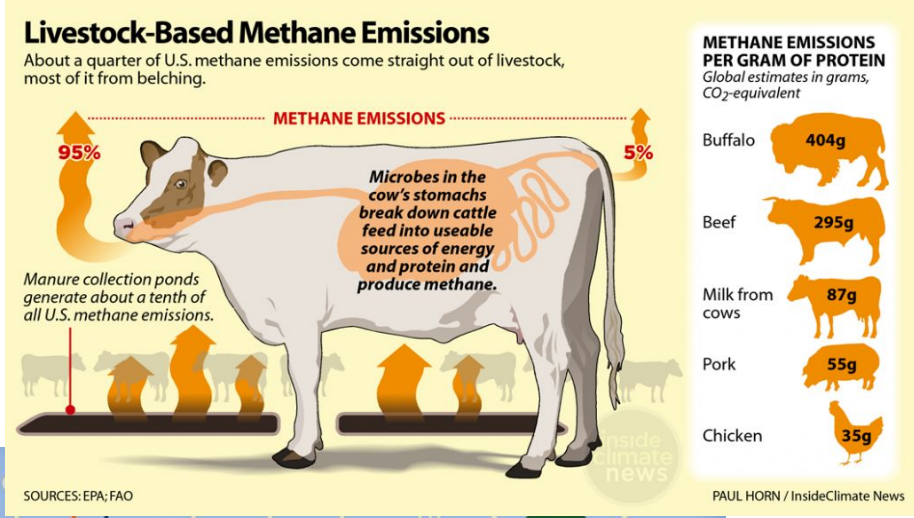


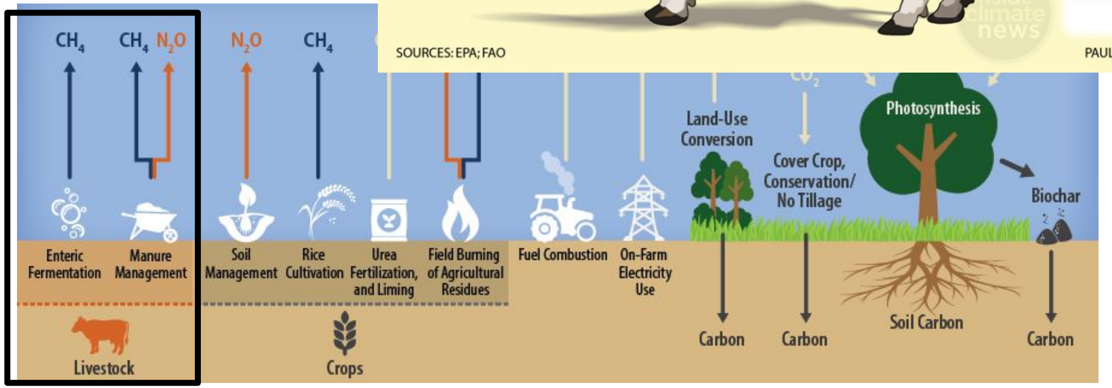


Livestock digestion is responsible for a large fraction of agriculture emissions

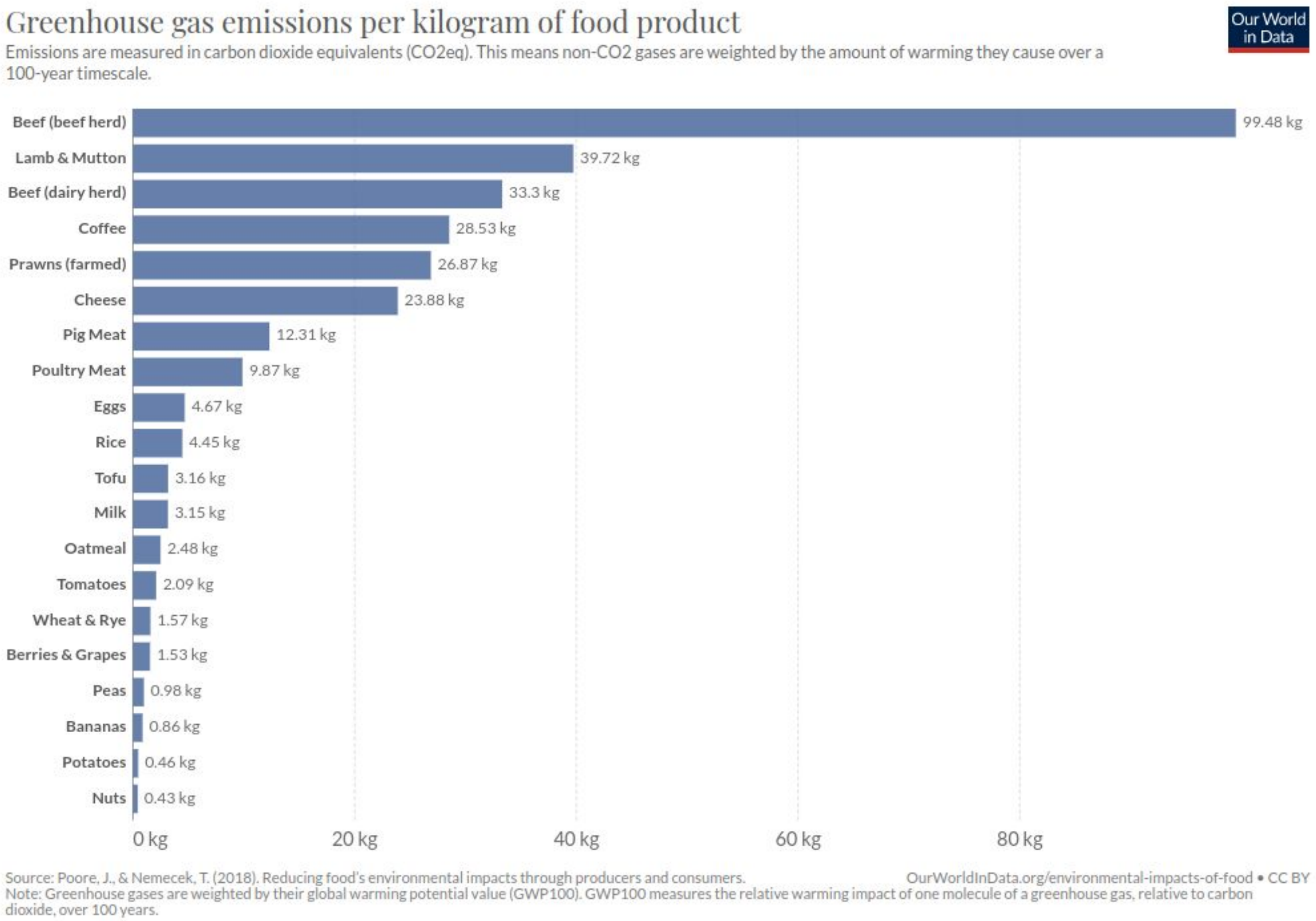
Gas from ruminant animals (cattle, sheep, goats, etc) causes methane release.

Manure stored in piles creates methane and nitrous oxide.



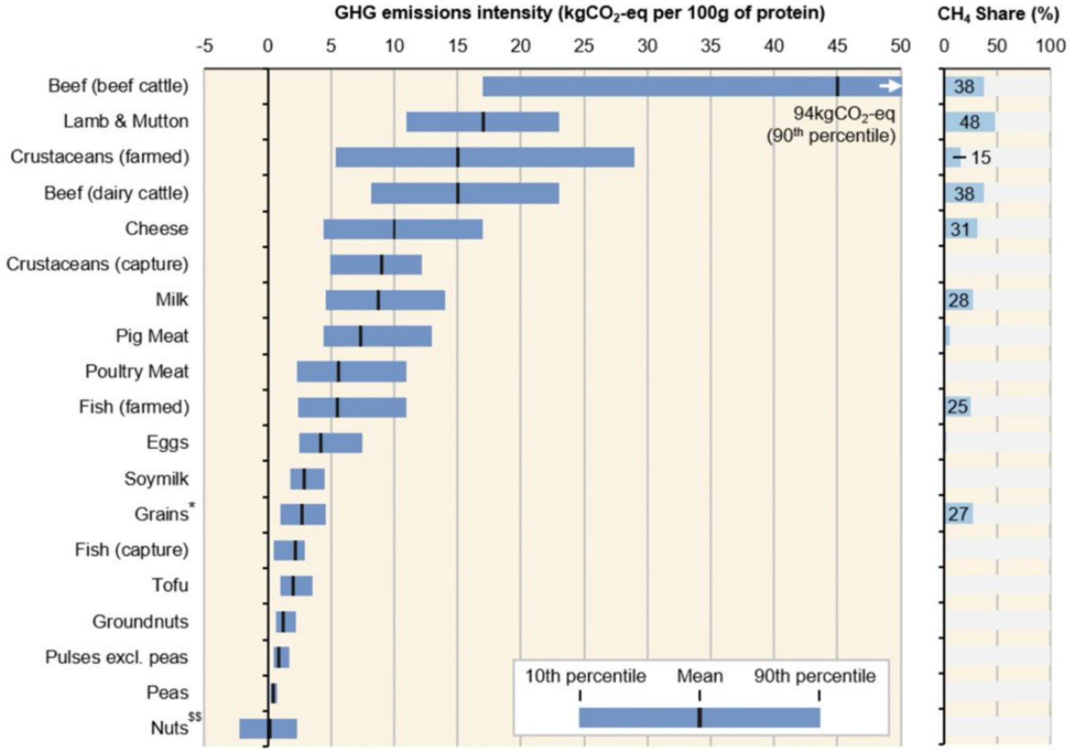


Emissions from different foods are *very* different

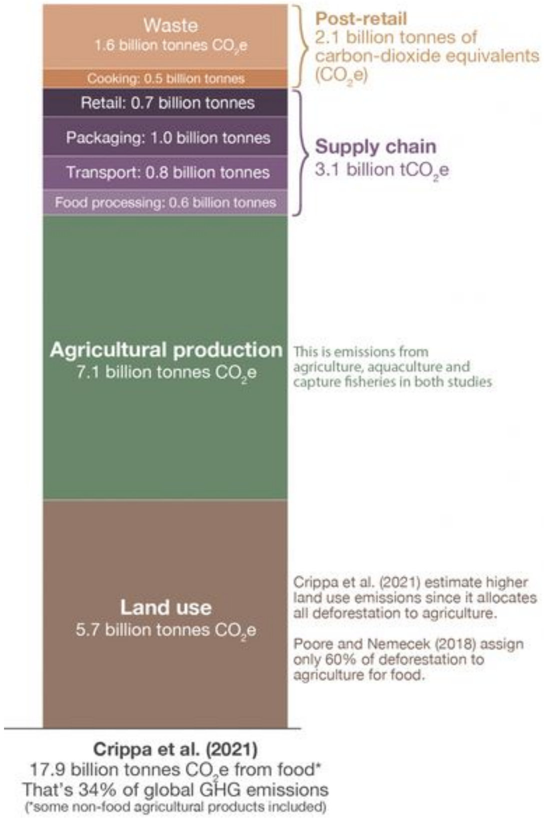


Emissions from different foods are *very* different

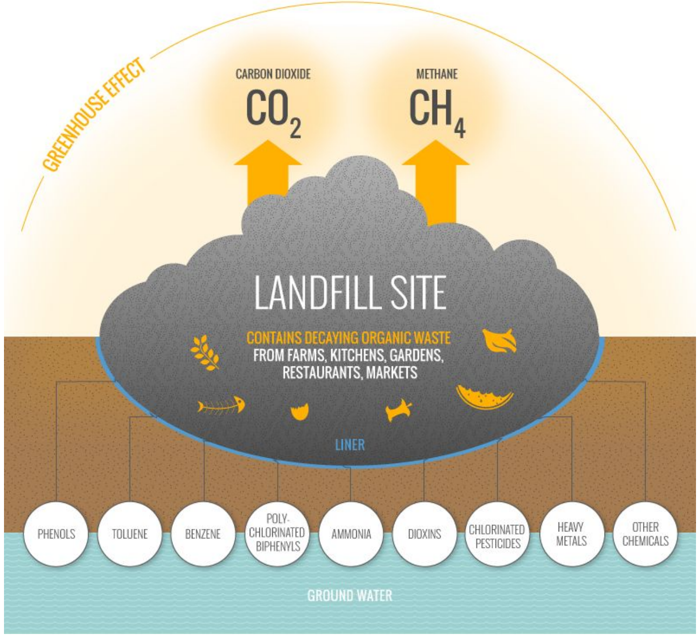
Emissions per 100g of protein



Food waste also contributes to emissions



74kg (163 lbs) of food waste per person, per year.



Too Good to Go

Too Good To Go

DISCOVER
delicious surplus food at a great discount around you

PICK-UP
delicious food from shops nearby

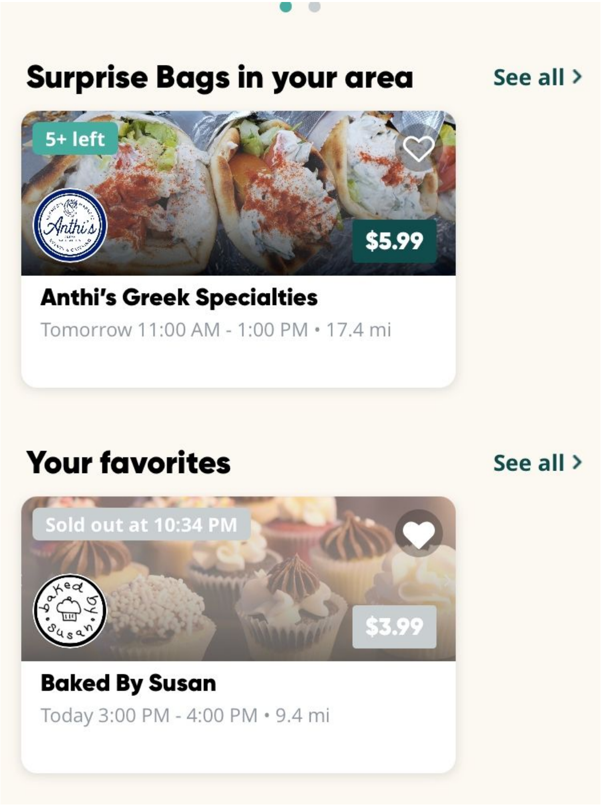
ENJOY
a tasty meal that helps the planet

WHY IT MATTERS

Globally, more than ⅓ of food is wasted - and that's bad news for our planet. Food waste is responsible for 10% of greenhouse gases, and we use a landmass the size of China to produce food we end up throwing away. It makes no sense, does it?

At Too Good To Go, we're determined to help fix the problem. Our app lets you rescue delicious, unsold food from businesses to save it from being thrown away.

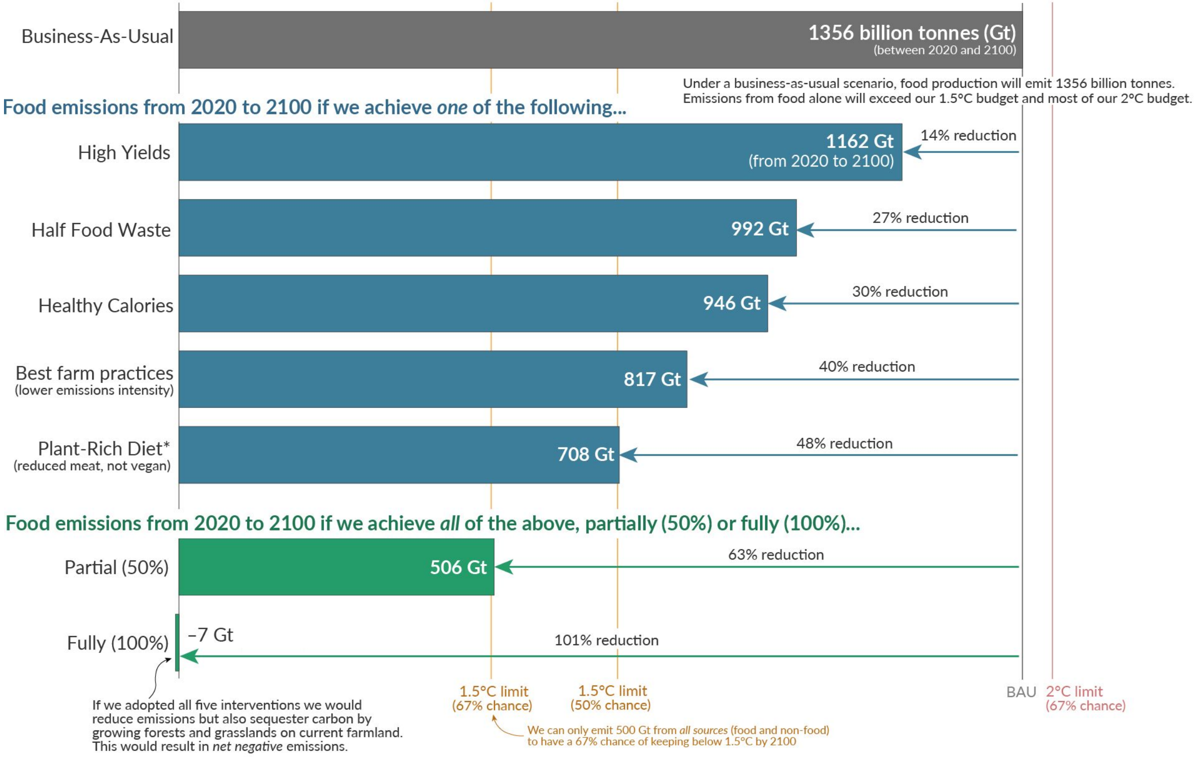
In turn, the app powers our efforts to build an anti-food waste movement. Globally, our dedicated team works within organizations like local governments and schools to shake up the food system, and change the way we think about food.



How can we reduce global greenhouse gas emissions from food?

Our World in Data

Shown are estimates of cumulative greenhouse gas emissions from food production from 2020 to 2100 under a business-as-usual scenario, and five interventions to reduce emissions.
This is measured in global warming potential (GWP*) CO_2 warming-equivalents (CO_2-we).

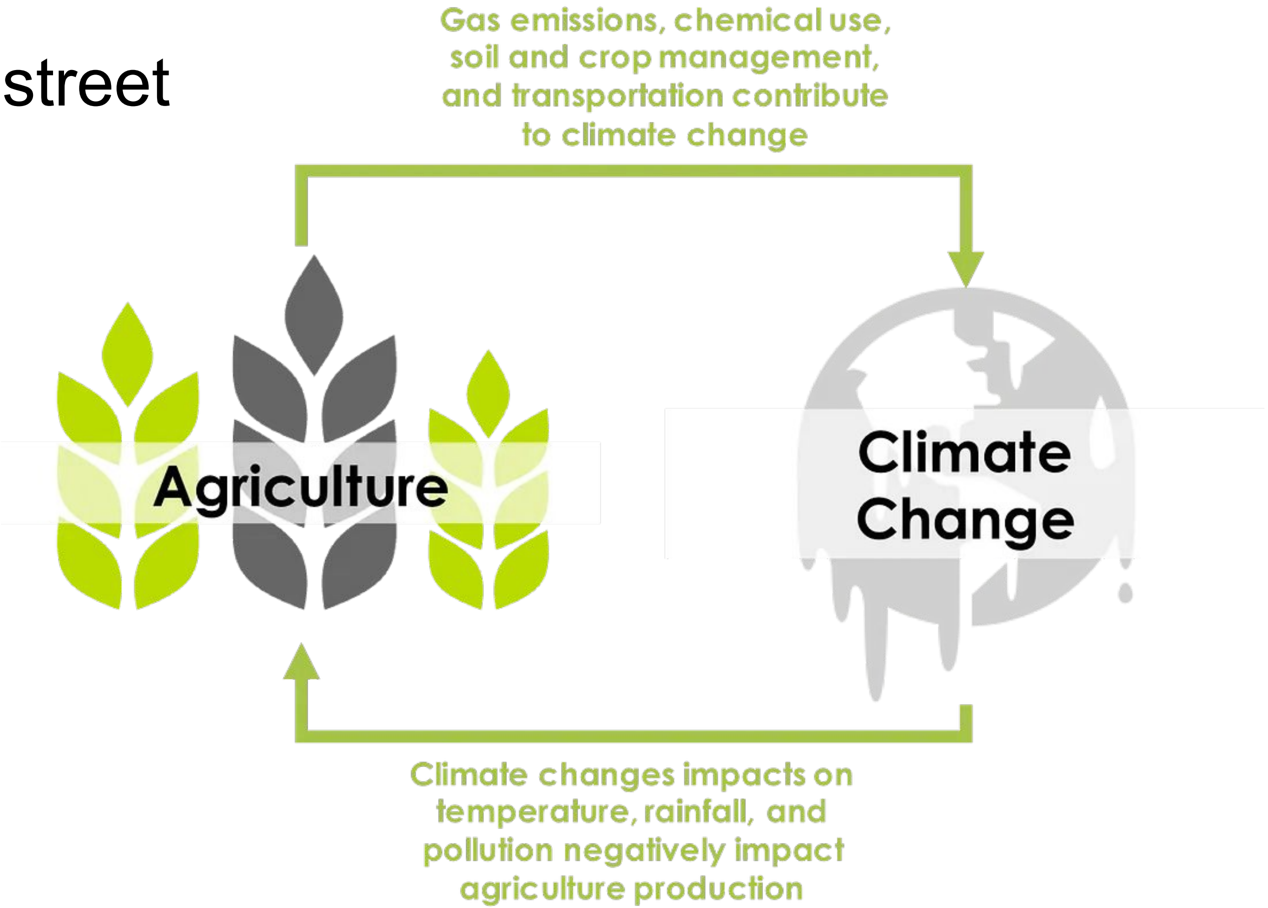


*Based on the EAT-Lancet Planetary Health diet which reduces but does not eliminate meat or dairy consumption.

Source: Michael Clark et al. (2020). Global food system emissions could preclude achieving the 1.5° and 2°C climate change targets. *Science*.

OurWorldinData.org – Research and data to make progress against the world's largest problems.

Two-way street



Foods at particular risk

Coffee - rising temperatures are estimated to reduce the suitable coffee-growing land by 50% by 2050 and support the growth of coffee rust fungus

Chocolate - yield will decline as soon as 2030. Growing regions will shift to South America

Shellfish - at risk from hotter and more acidic oceans

Corn - yield for corn grain will decrease within the U.S. Corn Belt by 20 to 40% from 1991-2000 levels by 2046-2055 with large economic impact

Livestock - impacted by more heat stress, disease, and water availability

Rice - increase and severity of hot weather can cause yields to decrease by 40% within this century

Honey - Shifts in the flowering plant cycles can cause nutritional stress on bee populations.

https://www.tastingtable.com/1016241/how-climate-change-is-impacting-these-foods/

Predicted impacts on food security

Shared Socioeconomic Pathway	Production		Storing/Processing/ Packaging		Trade		Transport	
	P	W	P	W	P	W	P	W
SSP1	Medium/Low Risk	Low Risk	Medium/Low Risk	Low Risk	Medium/Low Risk	Low Risk	Medium/Low Risk	Low Risk
SSP2	Medium Risk	Medium/Low Risk	Medium Risk	Low Risk	High Risk	Medium Risk	Medium Risk	Medium/Low Risk
SSP3	High Risk	Medium Risk	Medium Risk	Low Risk	Very High Risk	High Risk	Medium Risk	Medium/Low Risk
SSP4	Very High Risk	Medium Risk	Medium Risk	Low Risk	Very High Risk	Medium Risk	Medium Risk	Low Risk
SSP5	Medium/Low Risk	Low Risk	Medium/Low Risk	Low Risk	Medium/Low Risk	Low Risk	Medium Risk	Low Risk

(P: poorer nations, W: wealthier nations)

Key
Low Risk
Medium/Low Risk
Medium Risk
High Risk
Very High Risk

Figure ES- 5. Relative risks to food availability for different SSPs. *The risks to food availability would be lowest under the economic conditions described in in SSP 1 and SSP 5 for a given scenario of climate change, with poorer nations being at higher risk across all food production, distribution and trade categories for all SSPs. Shading represents higher or lower risks for each SSP from climate change. Risks reflect the informed judgment of the authors of this report based on the available literature.*

Climate change will impact production and distribution of food

Impacts on health

Impact of climate change on food production could cause over 500,000 extra deaths in 2050

03 March 2016

https://www.oxfordmartin.ox.ac.uk/news/201601-climate-food-production/

- Lead to average per-person reductions in food availability of 3.2% (99 kcal per day), in fruit and vegetable intake of 4.0% (14.9g per day), and red meat consumption of 0.7% (0.5g per day).
- Almost three-quarters of all climate-related deaths expected to occur in China (248,000) and India (136,000).

Accurate knowledge of crops will be crucial for planning

For example, during the COVID-19 pandemic, the government of the Togolese Republic used satellite imagery to identify crop locations and distribute aid to increase production

> *“We present results for this method in Togo, where we delivered a high-resolution (10 m) cropland map in under 10 days to facilitate rapid response to the COVID-19 pandemic by the Togolese government. This demonstrated a successful transition of machine learning applications research to operational rapid response in a real humanitarian crisis.”*

KDD ’20 Humanitarian Mapping Workshop, August 24, 2020, San Diego, CA

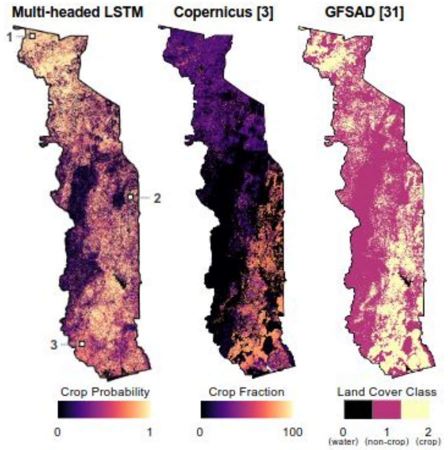


Figure 3: Cropland maps of Togo generated using our method, Buchhorn et al. [4], and Xiong et al. [33]. White boxes indicate locations of insets shown in Figure 4.

https://arxiv.org/pdf/2006.16866.pdf

Sentinel-1

We have previously seen work that uses data from the Sentinel-2 (optical) satellite. The current paper uses Sentinel-1 data, which is an “active sensing” satellite

Sentinel-1 Terminology

"Synthetic Aperture Radar"

Transmits pulses and listens for echoes, called backscatter.

Phase of the backscatter is used to determine the distance from the sensor to a target

Amplitude provides information about the roughness, geometry, wetness, etc of the Earth at that location

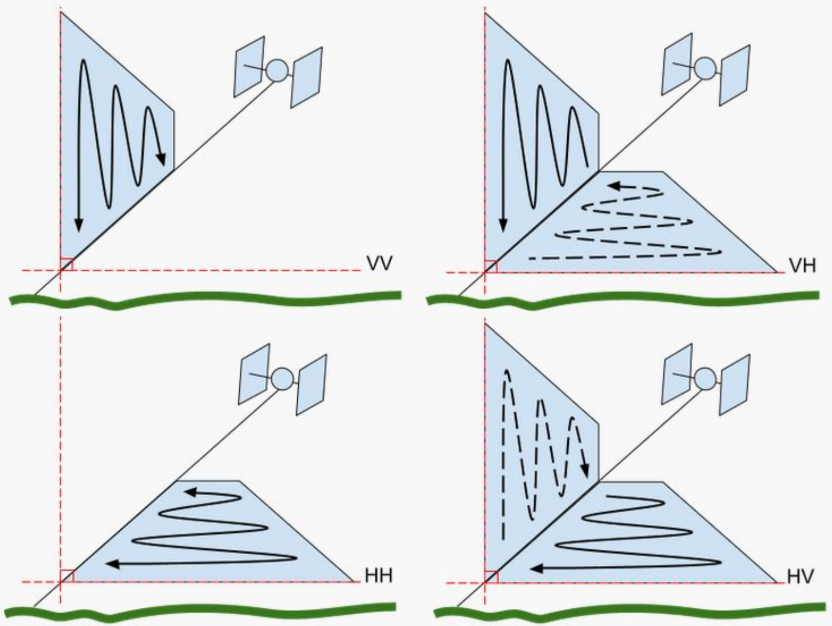
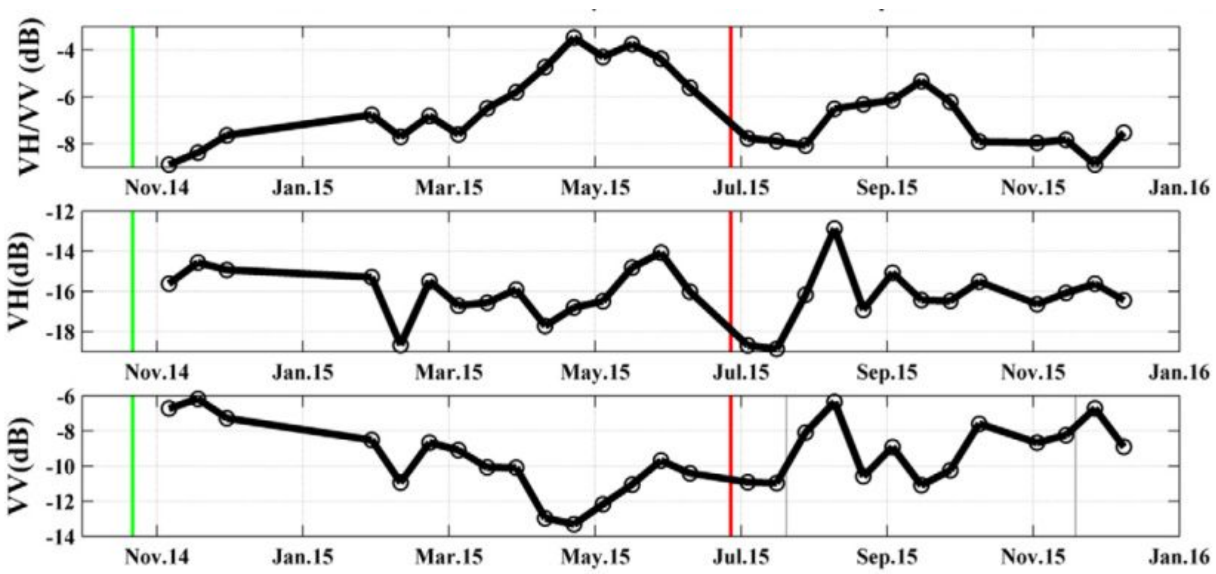


Figure 2: SAR signals are transmitted and received either vertically (V) or horizontally (H). This gives the potential for four different polarization combinations (transmit listed first, receive second): VV, VH, HH, and HV. Credit: ASF

https://hyp3-docs.asf.alaska.edu/guides/introduction_to_sar/

SAR values over time



Time Series Data

Data that

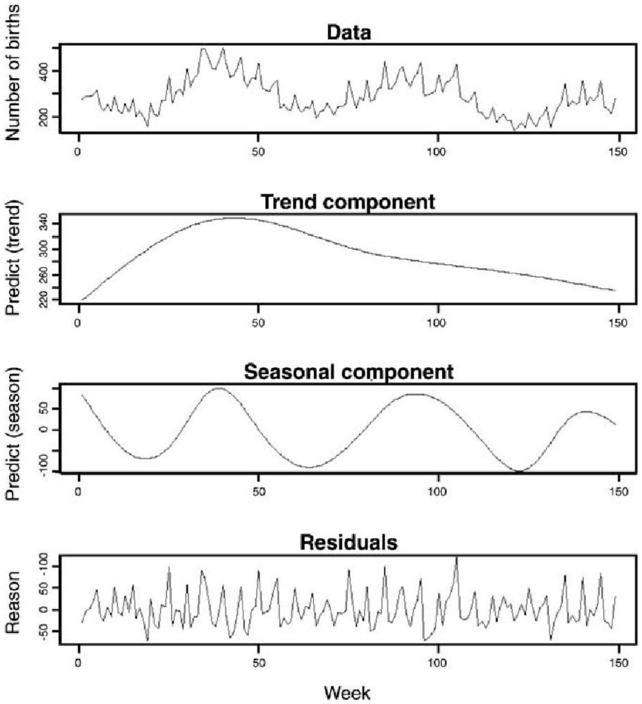
-is sampled at regular intervals

-has some discernible trends

-can be univariable

-a type of “sequence” data

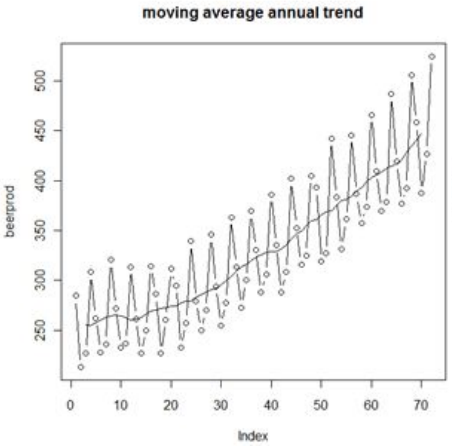
Have we seen any time series data in this class?

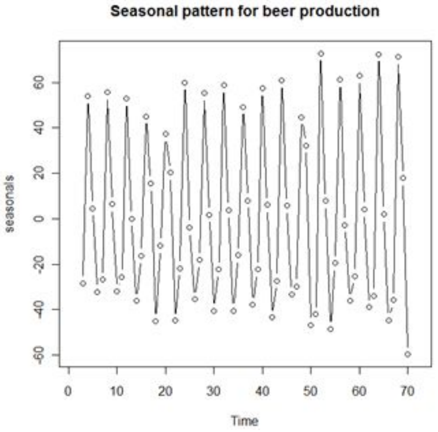


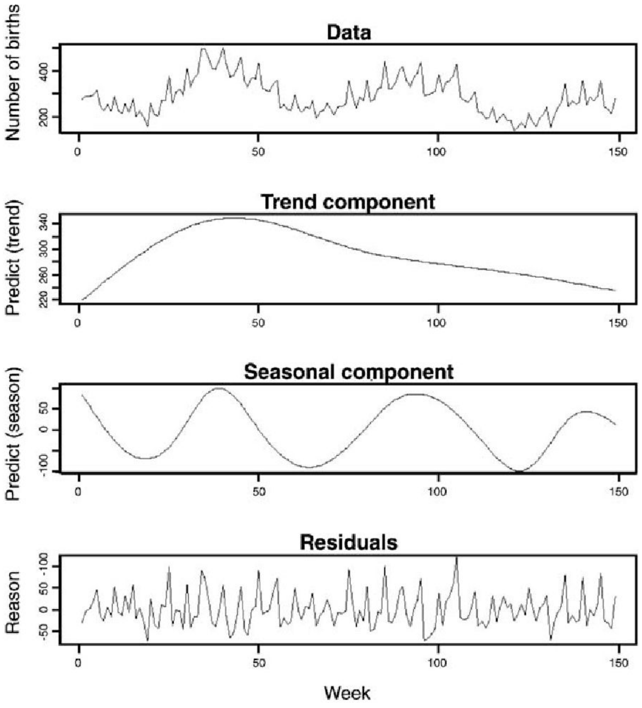
Zeger et al 2016

What you can do with time series data:

Identify trends and patterns with smoothing and subtraction







stat.psu.edu

What you can do with time series data:

“Auto” regression for prediction

Autoregressive (AR) Models

A common approach for modeling univariate time series is the autoregressive (AR) model:

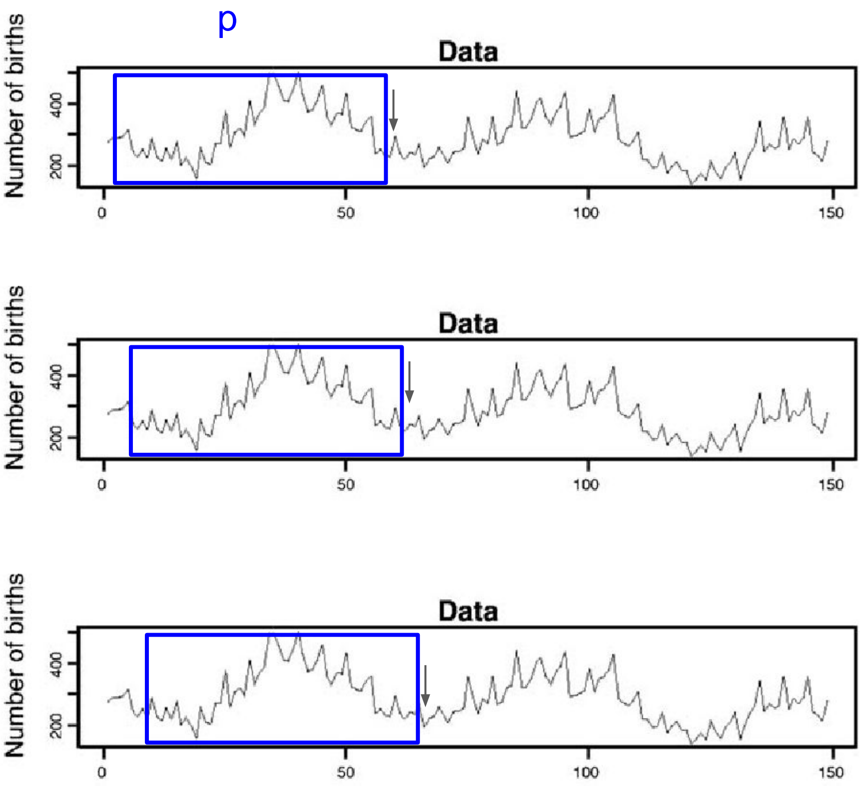
$$X_t = \delta + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-1} + A_t$$

where X_t is the time series, A_t is white noise, and

$$\delta = \left(1 - \sum_{i=1}^{p} \phi_i\right) \mu \,,$$

with μ denoting the process mean.

An autoregressive model is simply a linear regression of the current value of the series against one or more prior values of the series. The value of p is called the order of the AR model.



https://www.itl.nist.gov/div898/handbook/pmc/section4/pmc444.htm

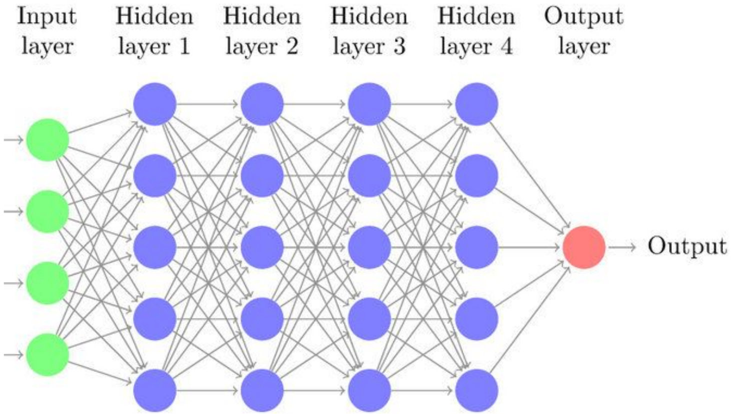
What you can do with time series data:

Put them into artificial neural networks for prediction, classification, etc

All time points go into the model at once (concatenated):

(2) pre-disaster image

(1) post-disaster image

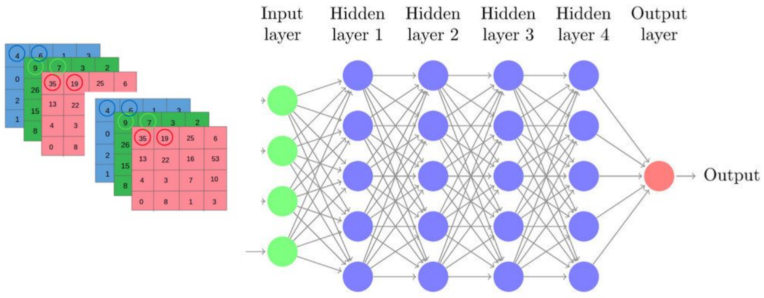


What you can do with time series data:

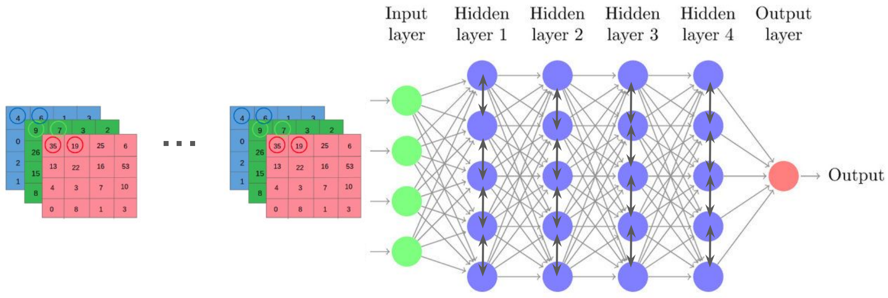
Put them into artificial neural networks for prediction, classification, etc

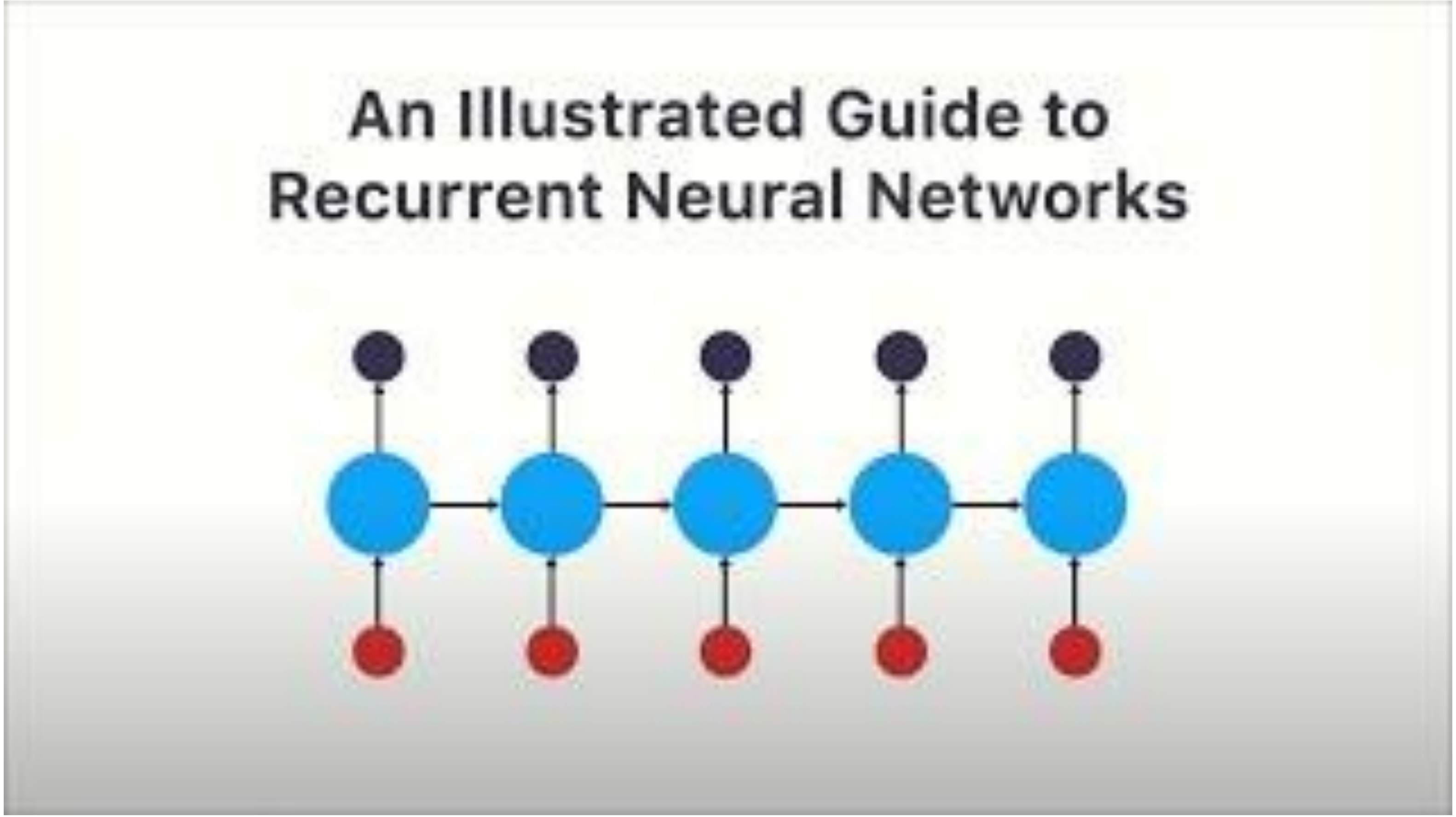
Or time points go in one at a time with **recurrent neural networks**

Feedforward Network:



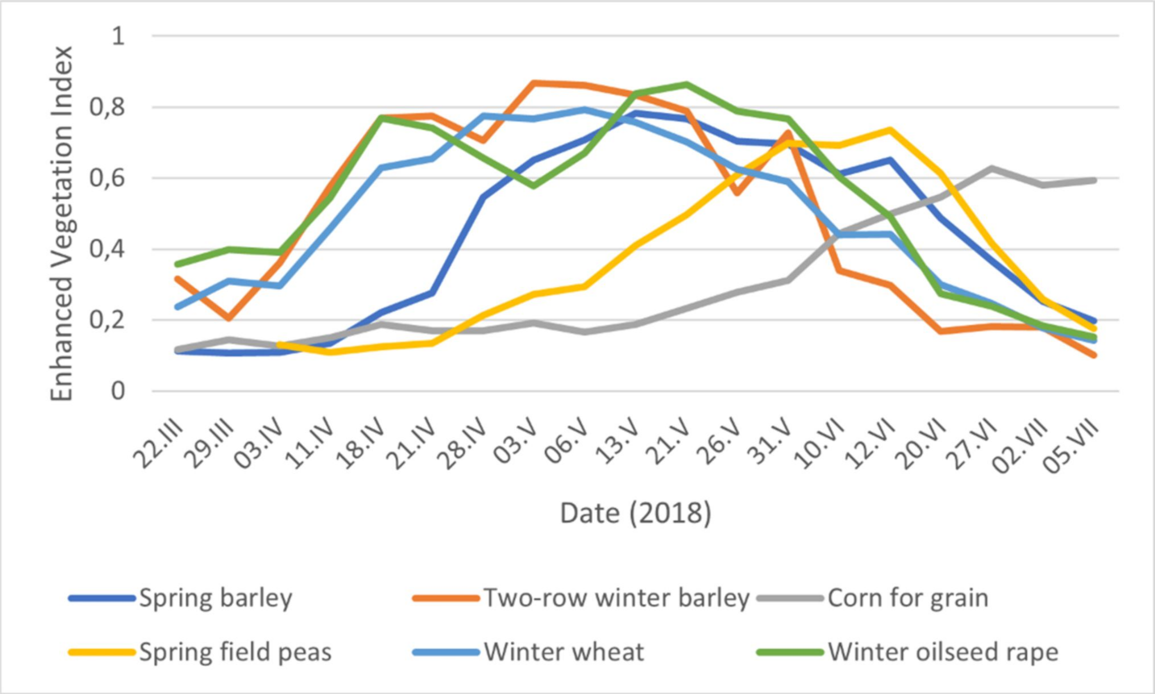
Recurrent Network:





Recurrent neural networks have **memory**

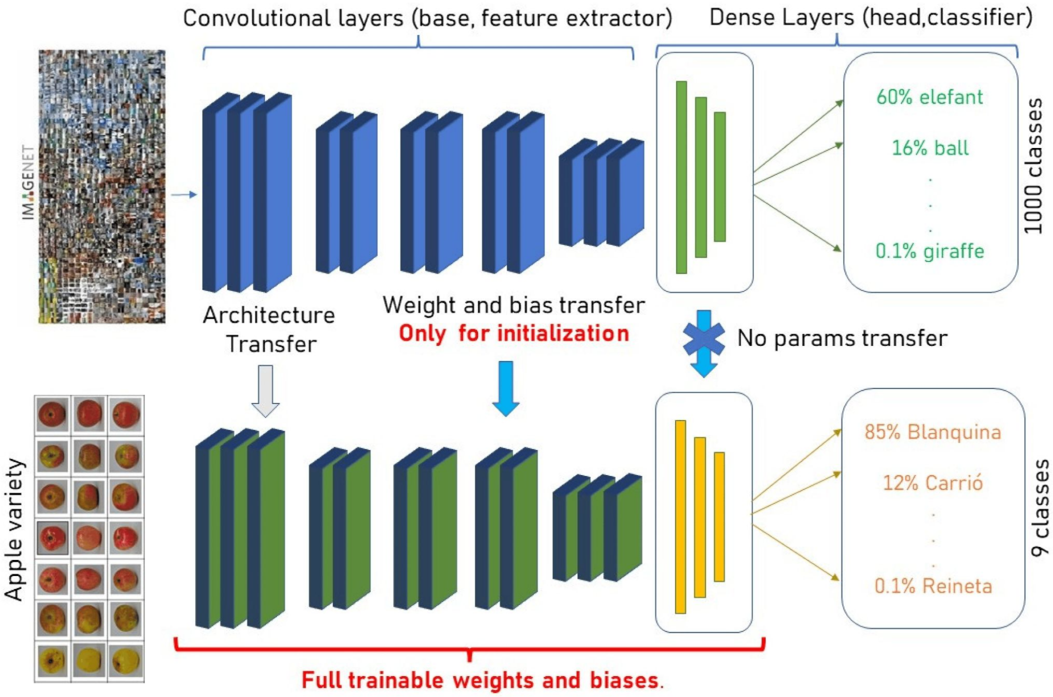
If our goal is to identify a specific crop, why do we want time series data?



Crops have specific temporal signatures

Snevajs et al

What if we don't have a lot of labeled data for the region we want to study?



Cortés et al

Transfer Learning: Train a neural network on a large related dataset and then adapt it for your specific task

Fine-tuning: Adapt to your specific task by making small changes to the weights of your pre-trained network

Frozen weights: Weights that aren't changed during fine tuning