Chihuahua or Blueberry Muffin?

Intro to Machine Learning (ML) for Earth Scientists

Day 2

Lauren Hoffman
LPHYS2268, Winter 2025

Lecture 1: What is the machine learning?

- Motivation
- Case study
- Types of ML
- How it works
- *Tutorial: Building a ML model*

Lecture 2: ML Applications in Earth and Climate Science

- Applications
- Challenges
- *Activity: literature dive*

https://github.com/lahoffman/ml_lectures_LPHYS2268

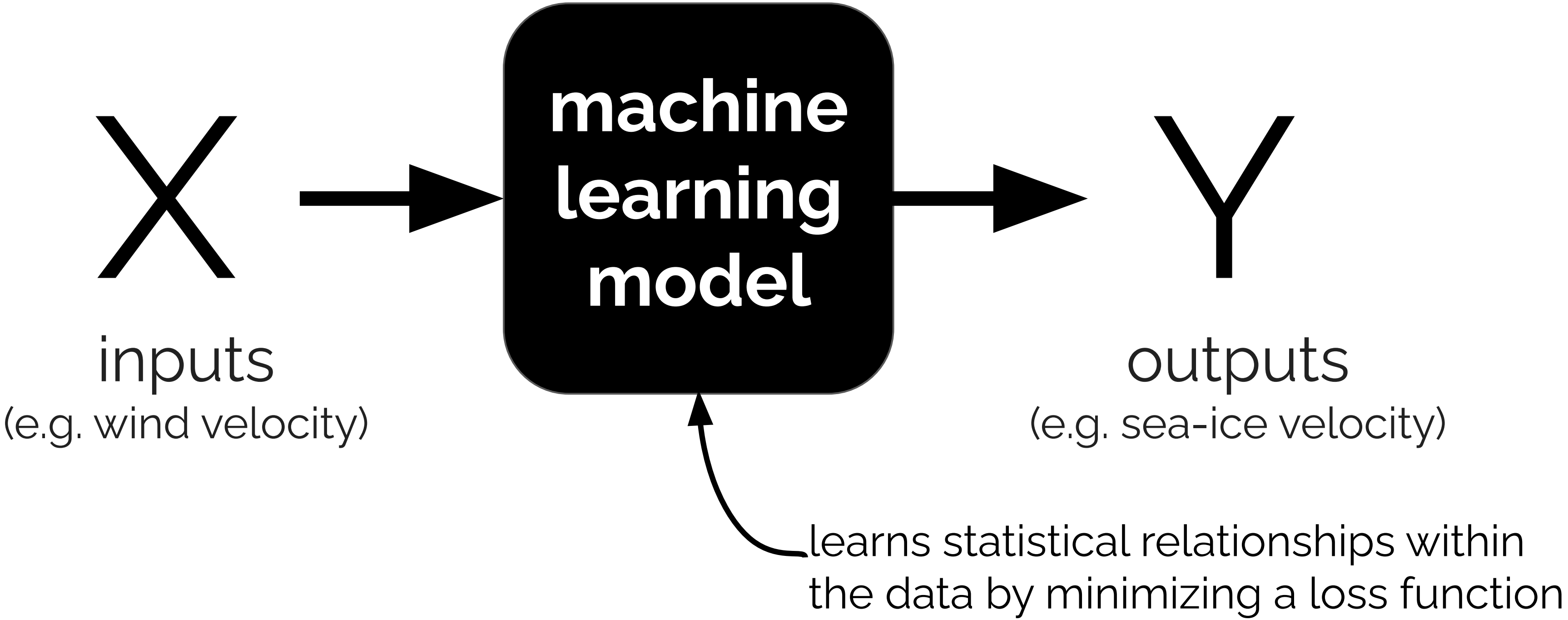
The take-home messages from Day 1...

ML models learn statistical relationships in the data, whereas numerical models are based on prescribed physical equations.

There are different machine learning types (supervised, unsupervised, reinforcement), problems (regression, classification), and algorithms (linear regression, neural networks).

ML models are trained to learn the weights and biases by minimizing a loss function.

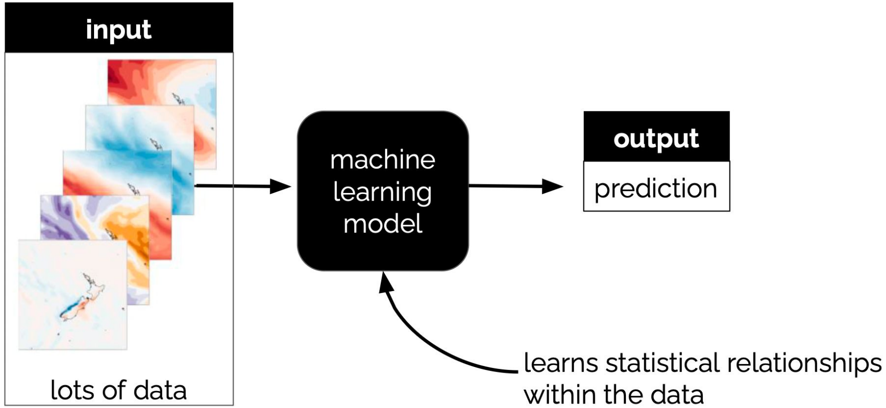
Machine learning models learn statistical relationships between the inputs and outputs.



Train Validate Test

ML models learn statistical relationships in data, whereas numerical models are based on prescribed physics.

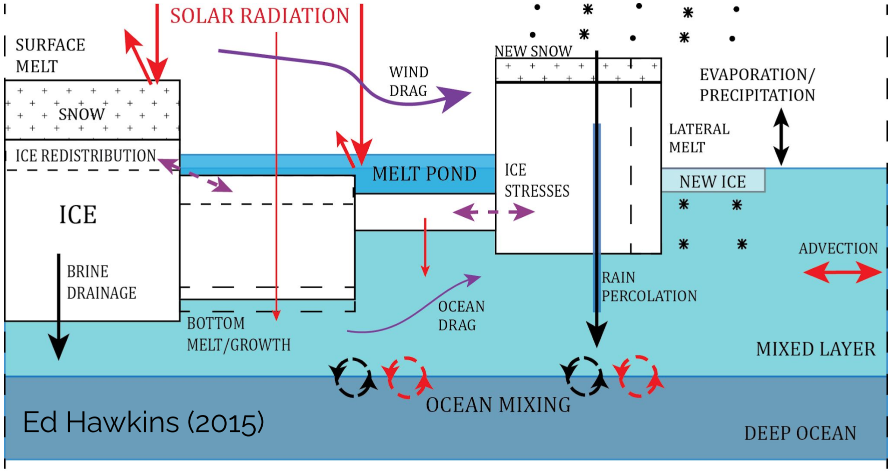
Machine Learning Models



Rampal et al. (2024)

ML models *learn from data* to recognize patterns and make predictions based on those patterns.

Numerical Models



Numerical models are *prescribed dynamical equations* and predict based on physics.

Machine Learning Models

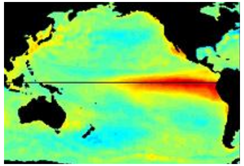
- Lack physical process knowledge
  - *there are efforts to improve this*
- The ‘black box’ is difficult to interpret
  - *there are efforts to improve this*
- Rely on statistical interactions in data
- Computationally efficient

Numerical Models

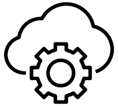
- Rooted in physics
- Interpretable
- Rely on parameterizations for small-scale processes
  there are efforts to improve this numerically ***and with ML***
- Computationally inefficient

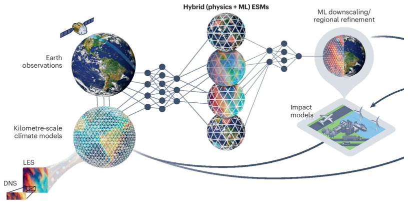
There are an number of applications for ML in Earth and climate sciences, including (but not limited to)...

Sources of predictability for modes of climate variability

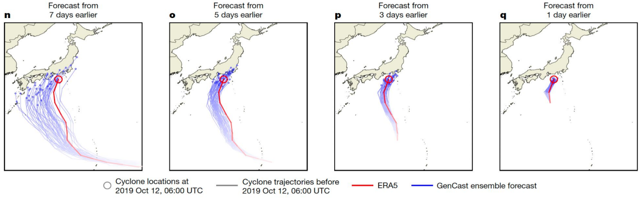


Earth System Modeling



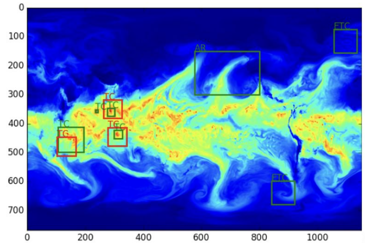


Extreme weather and climate prediction

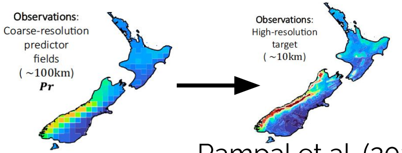


Price et al. (2024)

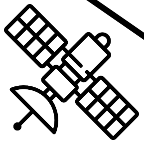
Feature Detection



Racah et al. (2017)



Rampal et al. (2024)

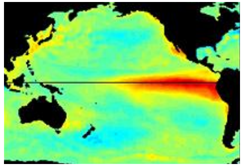


Satellite Data & Observations

Sources of Predictability

Climate Modes

Causal analysis



Forecasting

Short-term weather forecasting

Extreme Weather Events

Climate Forecasting

Satellite Data & Observations

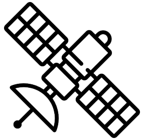
Interpolation

Global Reconstruction

Variable Representation

Downscaling

Merging Satellite Data



Synthetic data

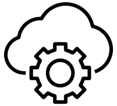
Sensor placement

Feature Detection

Extremes

Pattern Recognition

Earth System Modeling



Emulators

Model Calibration and Validation

Equation Discovery

Uncertainty Quantification

Sources of uncertainty

Parameterizations

Bias Correction

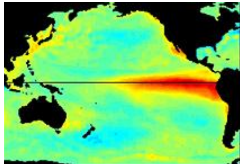
Data Assimilation

Climate Model analysis and benchmarking

Sources of Predictability

Climate Modes

Causal analysis



Forecasting

Short-term weather forecasting

Extreme Weather Events

Climate Forecasting

Satellite Data & Observations

Interpolation

Global Reconstruction

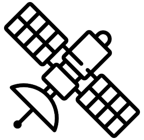
Variable Representation

Downscaling

Merging Satellite Data

Synthetic data

Sensor placement



Feature Detection

Extremes

Pattern Recognition

Earth System Modeling

Emulators

Model Calibration and Validation

Equation Discovery

Uncertainty Quantification

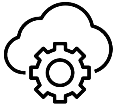
Sources of uncertainty

Parameterizations

Bias Correction

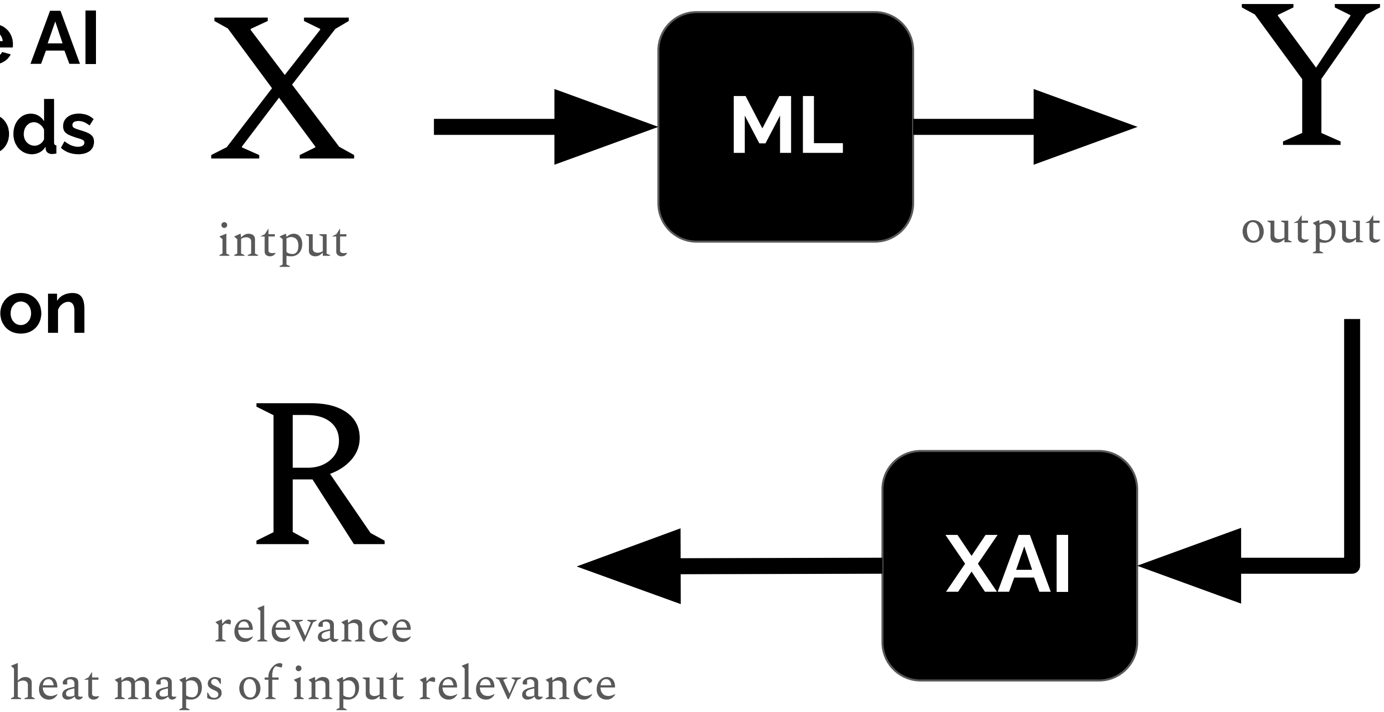
Data Assimilation

Climate Model analysis and benchmarking



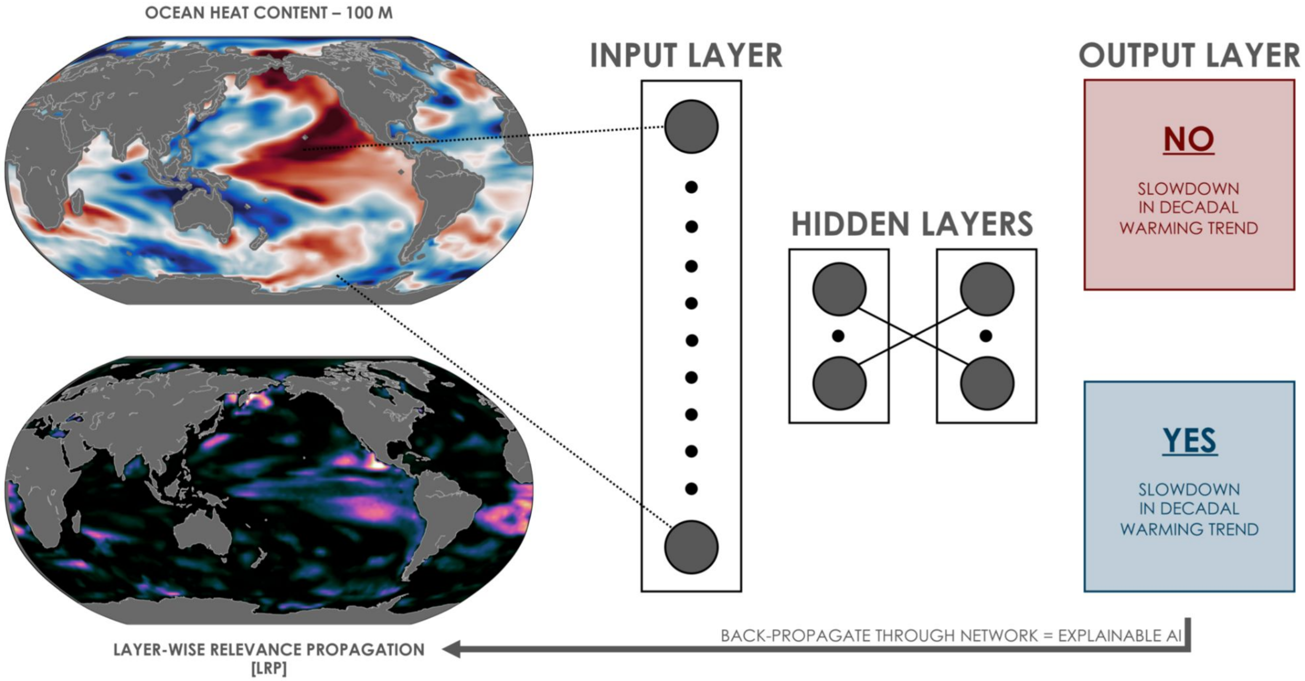
Sources of Predictability

Explainable AI (XAI) methods show us an interpretation of what the model has learned.



Sources of Predictability

A neural network predicts slowdowns in decadal warming trends from maps of ocean heat content.

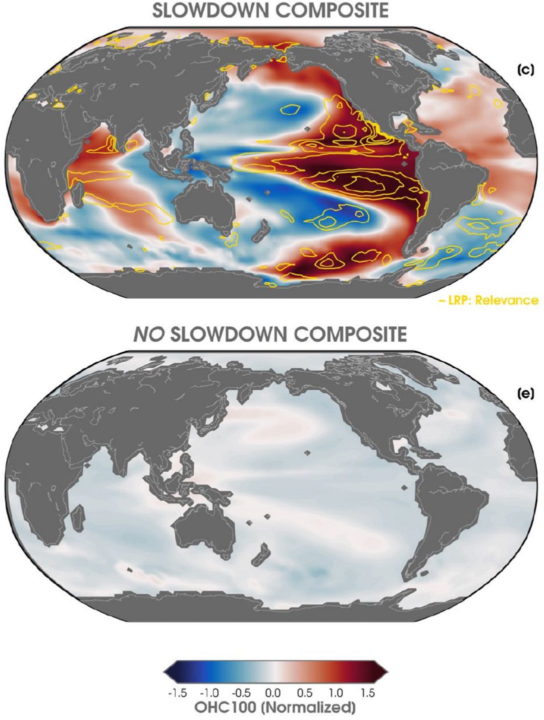


Labe & Barnes (2022)

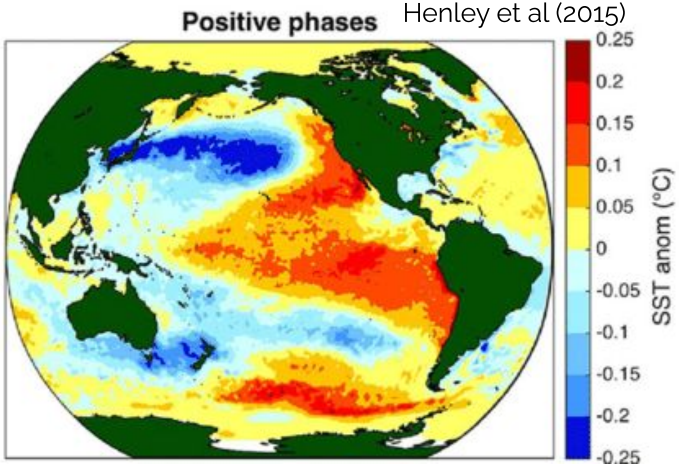
Sources of Predictability

Composite maps of OHC where the NN correctly predicted a 'slowdown' resemble the positive phase of the IPO.

Labe & Barnes (2022)



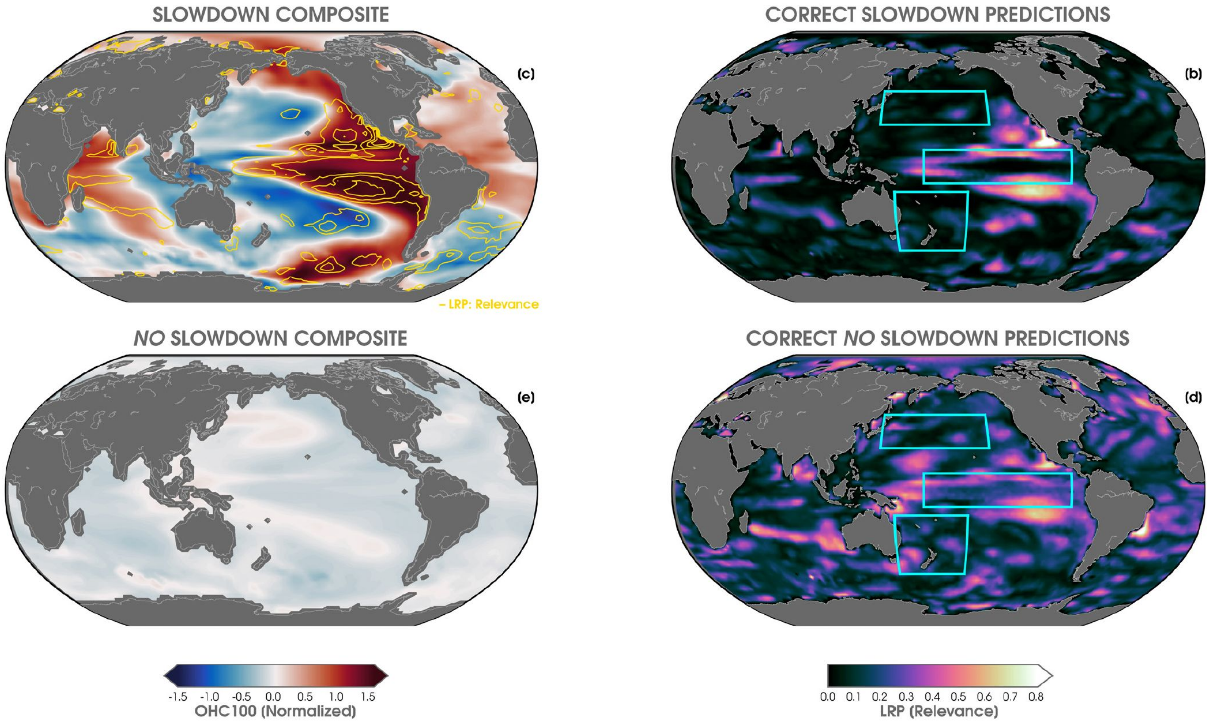
Henley et al (2015)



SST in the positive phase of the Interdecadal Pacific Oscillation (IPO)

Sources of Predictability

eXplainable AI shows where the input maps were relevant for predicting slowdown events.

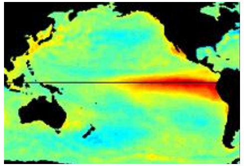


Labe & Barnes (2022)

Sources of Predictability

Climate Modes

Causal analysis



Feature Detection

Extremes

Pattern Recognition

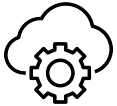
Forecasting

Short-term weather forecasting

Extreme Weather Events

Climate Forecasting

Earth System Modeling



Emulators

Model Calibration and Validation

Equation Discovery

Uncertainty Quantification

Sources of uncertainty

Parameterizations

Bias Correction

Data Assimilation

Climate Model analysis and benchmarking

Satellite Data & Observations

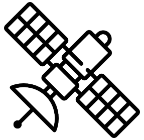
Interpolation

Global Reconstruction

Variable Representation

Downscaling

Merging Satellite Data

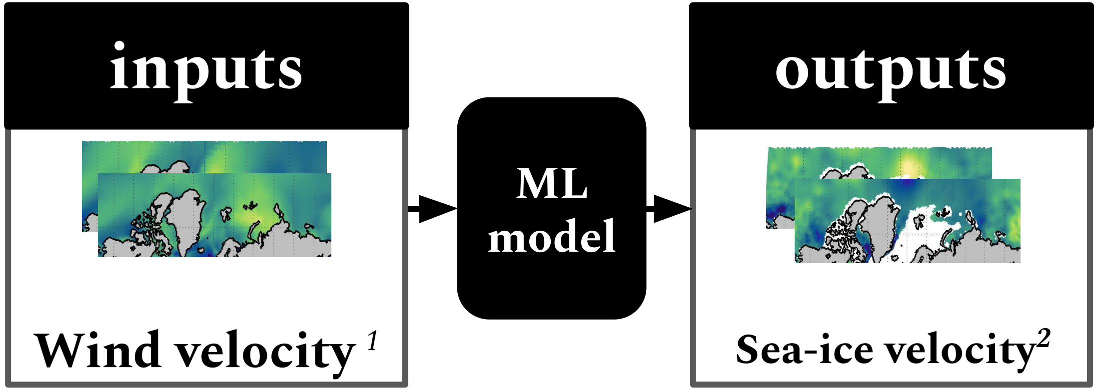


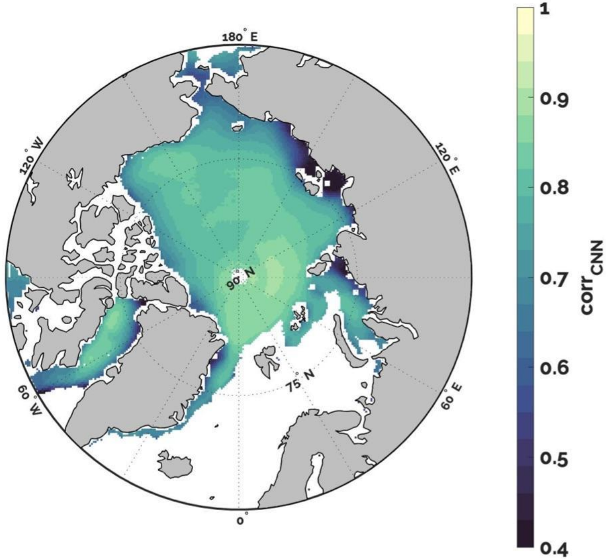
Synthetic data

Sensor placement

Forecasting

A convolutional neural network (CNN) makes skillful one-day predictions of Arctic sea ice motion.



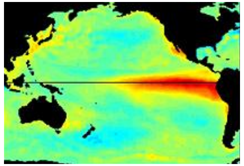


Hoffman et al. *AIES.* (2023).

Sources of Predictability

Climate Modes

Causal analysis



Forecasting

Short-term weather forecasting

Extreme Weather Events

Climate Forecasting

Satellite Data & Observations

Interpolation

Global Reconstruction

Variable Representation

Downscaling

Merging Satellite Data

Synthetic data

Sensor placement

Feature Detection

Extremes

Pattern Recognition

Earth System Modeling

Emulators

Model Calibration and Validation

Equation Discovery

Uncertainty Quantification

Sources of uncertainty

Parameterizations

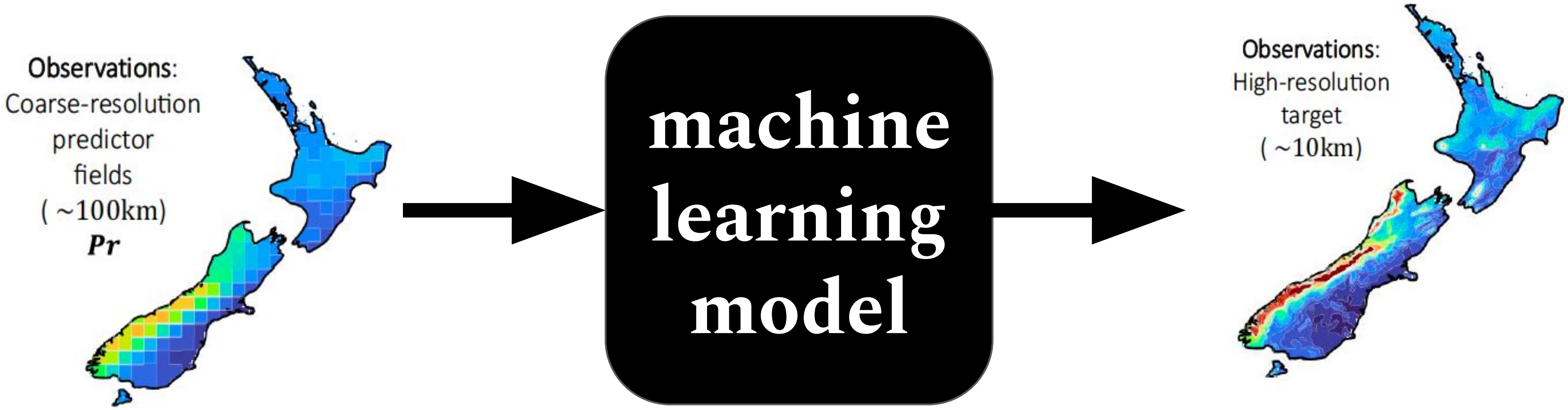
Bias Correction

Data Assimilation

Climate Model analysis and benchmarking

Satellite Data & Observations: Downscaling

Super Resolution (SR) downscaling maps a coarse resolution target to a high resolution target.

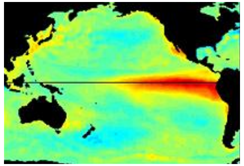


Rampal et al. (2024)

Sources of Predictability

Climate Modes

Causal analysis



Forecasting

Short-term weather forecasting

Extreme Weather Events

Climate Forecasting

Satellite Data & Observations

Interpolation

Global Reconstruction

Variable Representation

Downscaling

Merging Satellite Data

Synthetic data

Sensor placement

Feature Detection

Extremes

Pattern Recognition

Earth System Modeling

Emulators

Model Calibration and Validation

Equation Discovery

Uncertainty Quantification

Sources of uncertainty

Parameterizations

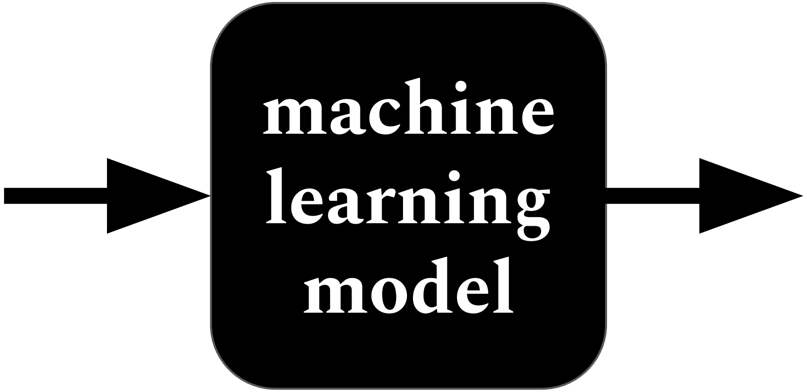
Bias Correction

Data Assimilation

Climate Model analysis and benchmarking

Feature Detection: Cloud Types

Machine learning model are trained to detect cloud types from a collection of images.



Strato-cumulus [80%]

Cumulus [15%]

Cirrus [5%]

Cloud Types

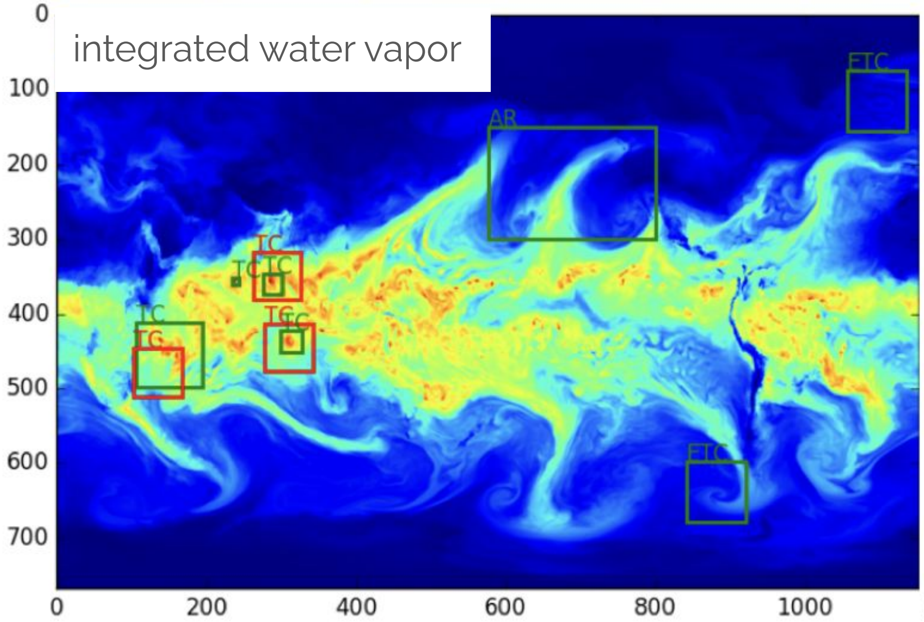
Lee et al. (1990)

Feature Detection: Extreme Weather

Supervised or unsupervised learning can detect patterns pertaining to extreme weather events.

Extreme Weather Detection

Racah et al. (2017)

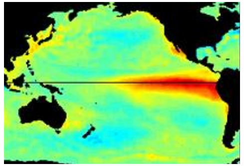


TC: tropical cyclone; AR: atmospheric river ; ETC: extratropical cyclone; ground truth; prediction

Sources of Predictability

Climate Modes

Causal analysis



Feature Detection

Extremes

Pattern Recognition

Forecasting

Short-term weather forecasting

Extreme Weather Events

Climate Forecasting

Satellite Data & Observations

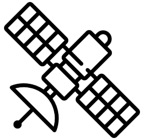
Interpolation

Global Reconstruction

Variable Representation

Downscaling

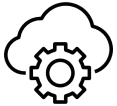
Merging Satellite Data



Synthetic data

Sensor placement

Earth System Modeling



Emulators

Model Calibration and Validation

Equation Discovery

Uncertainty Quantification

Sources of uncertainty

Parameterizations

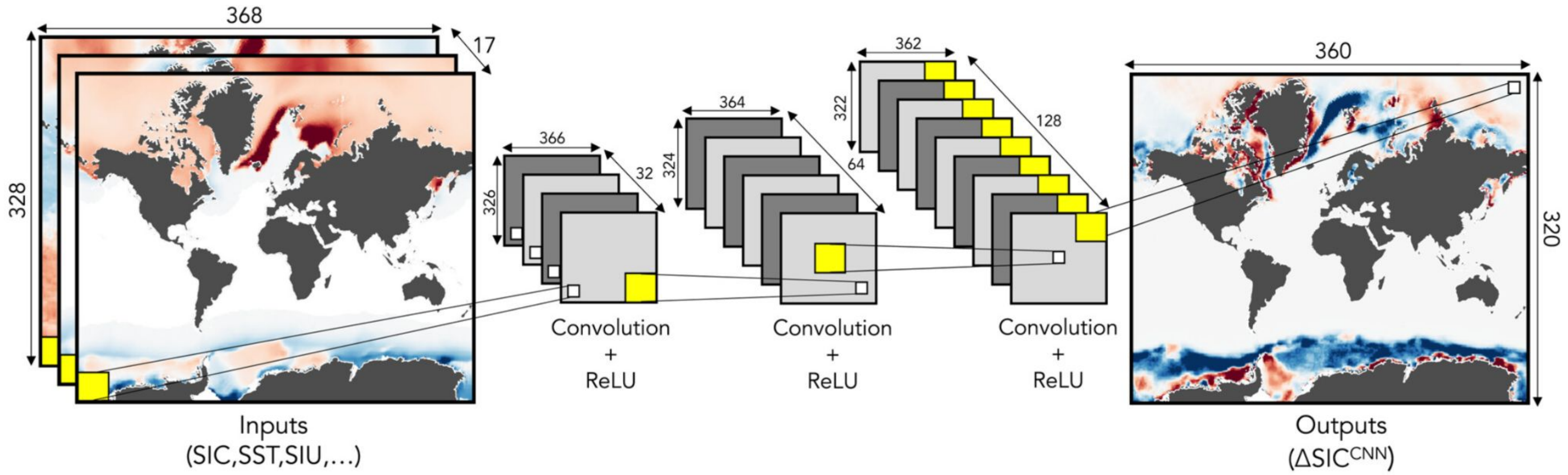
Bias Correction

Data Assimilation

Climate Model analysis and benchmarking

Earth System Modeling

Machine learning models trained to predict data assimilation errors in sea ice concentration (known as increments, ΔSIC) could help correct model bias.

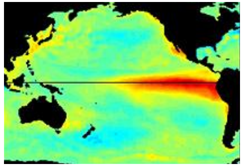


Gregory et al. (2024)

Sources of Predictability

Climate Modes

Causal analysis



Forecasting

Short-term weather forecasting

Extreme Weather Events

Climate Forecasting

Satellite Data & Observations

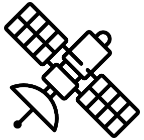
Interpolation

Global Reconstruction

Variable Representation

Downscaling

Merging Satellite Data



Synthetic data

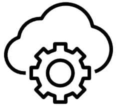
Sensor placement

Feature Detection

Extremes

Pattern Recognition

Earth System Modeling



Emulators

Model Calibration and Validation

Equation Discovery

Uncertainty Quantification

Sources of uncertainty

Parameterizations

Bias Correction

Data Assimilation

Climate Model analysis and benchmarking

But incorporating ML into Earth and climate science still has its challenges…

Challenges
Data Availability
Robustness on out-of-distribution samples
Interpretability
Physical Inconsistency
Uncertainty Quantification

But incorporating ML into Earth and climate science still has its challenges...

Challenges
Data Availability
Robustness on out-of-distribution samples
Interpretability
Physical Inconsistency
Uncertainty Quantification

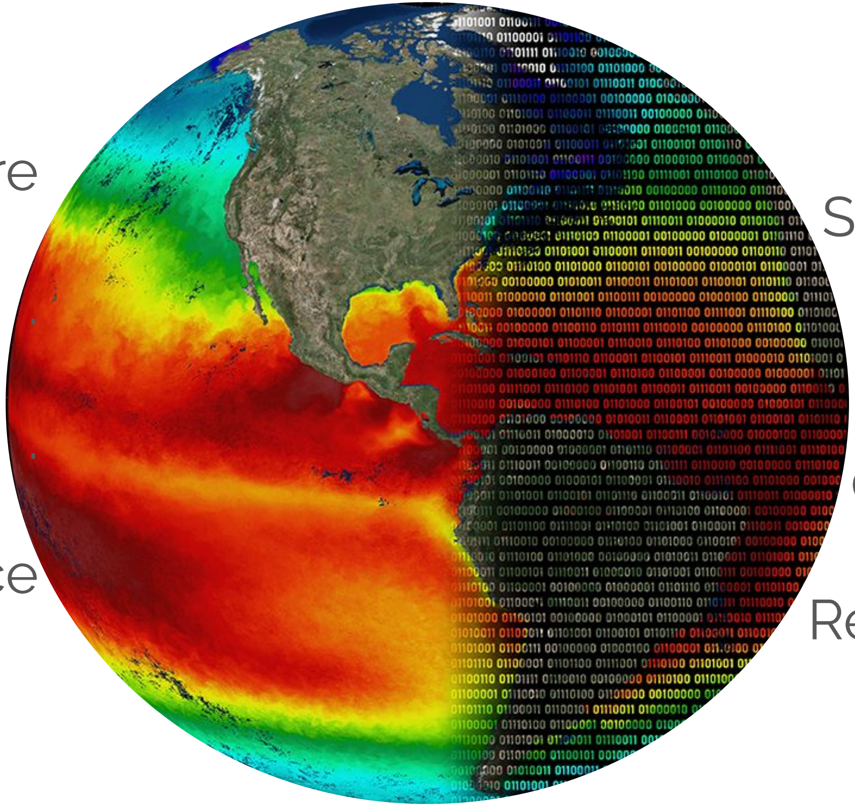
Data Availability: it is impossible to measure everything →
Climate Model Output can be used to train a ML model.

Atmosphere

Ocean

Land

Ice



Satellite

In situ

Climate models

Reanalysis products

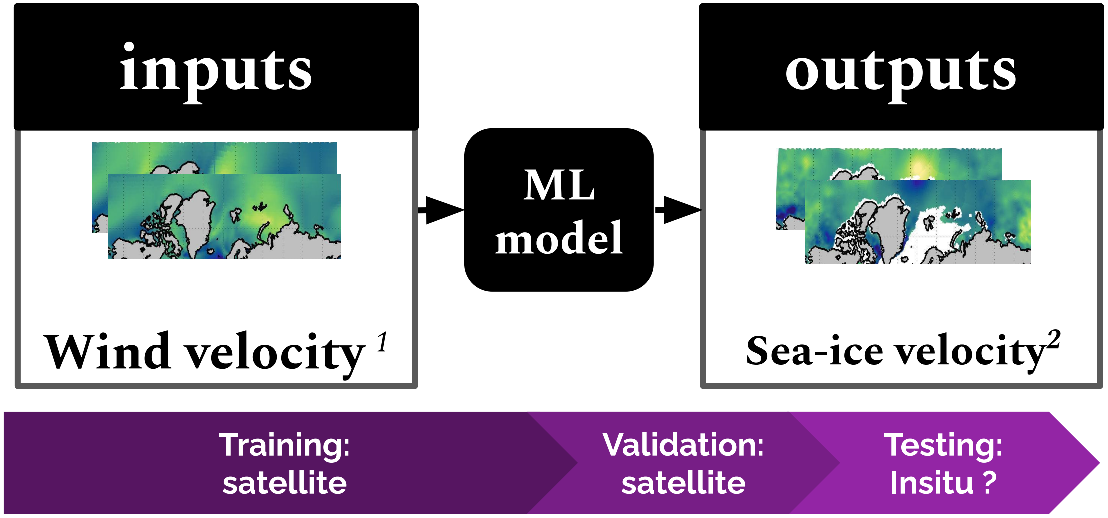
ML ❤ DATA

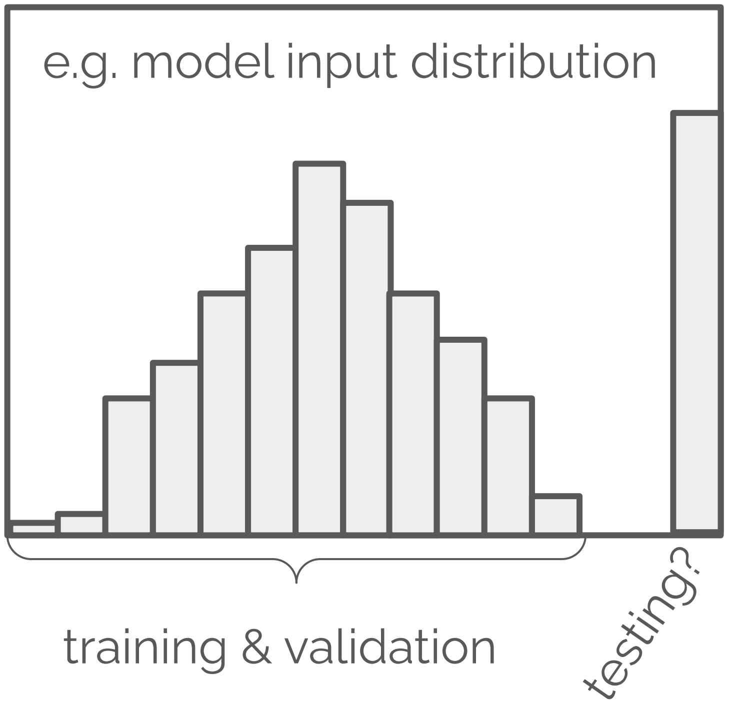
Image: Courtesy of Kate Culpepper with design elements provided by Esri, HERE, Garmin, FAO, NOAA, USGS, EPA.

But incorporating ML into Earth and climate science still has its challenges...

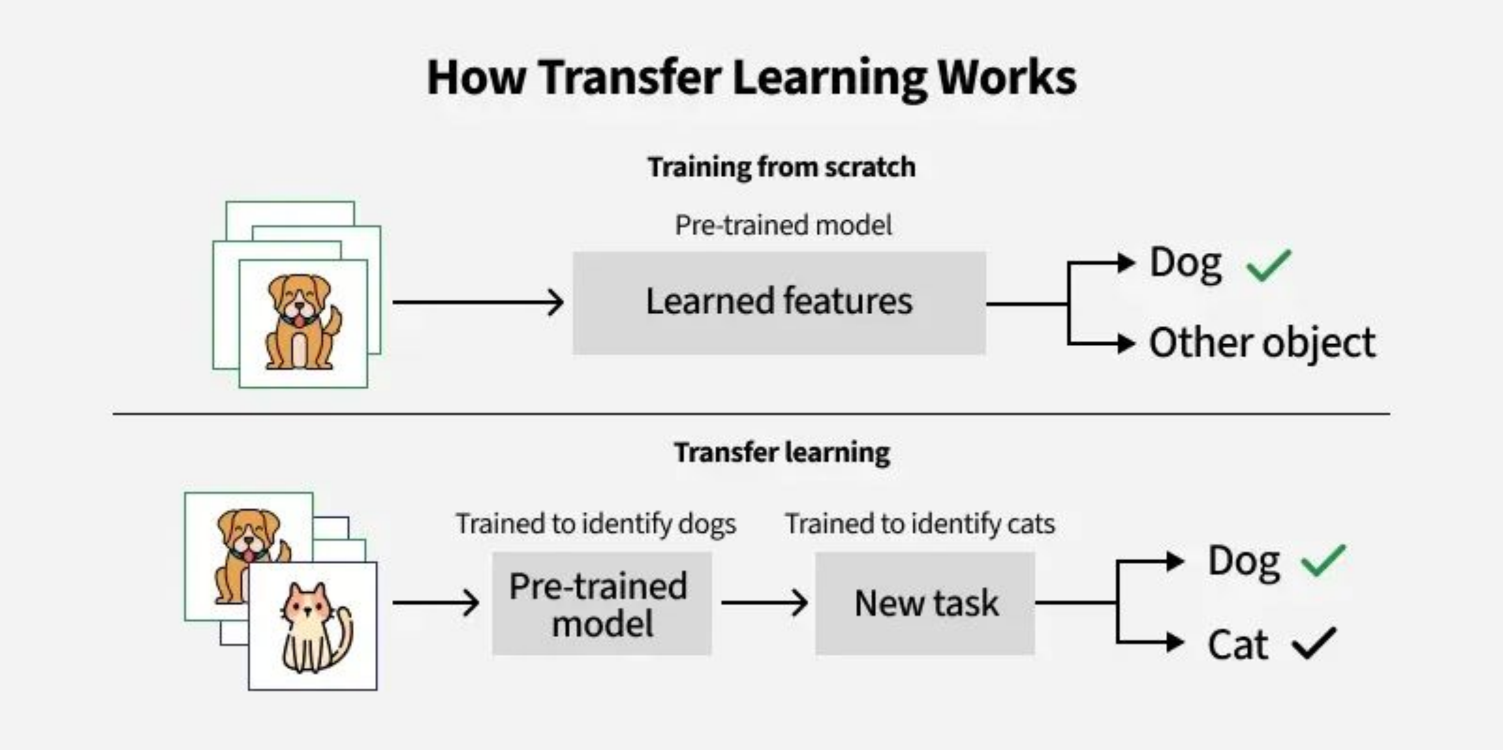
Challenges
Data Availability
Robustness on out-of-distribution samples
Interpretability
Physical Inconsistency
Uncertainty Quantification

Robustness on Out-of-Distribution Samples:
ML models often have difficulty generalizing.





Robustness on Out-of-Distribution Samples →



Transfer Learning uses a pre-trained model and new data to learn more information (i.e. further improve loss function).

But incorporating ML into Earth and climate science still has its challenges…

Challenges
Data Availability
Robustness on out-of-distribution samples
Interpretability
Physical Inconsistency
Uncertainty Quantification

Interpretability:
ML models can learn the right thing for the wrong reasons.

Clever Hans: the horse that can do math!

Or can he....

Interpretability → eXplainable AI (XAI)

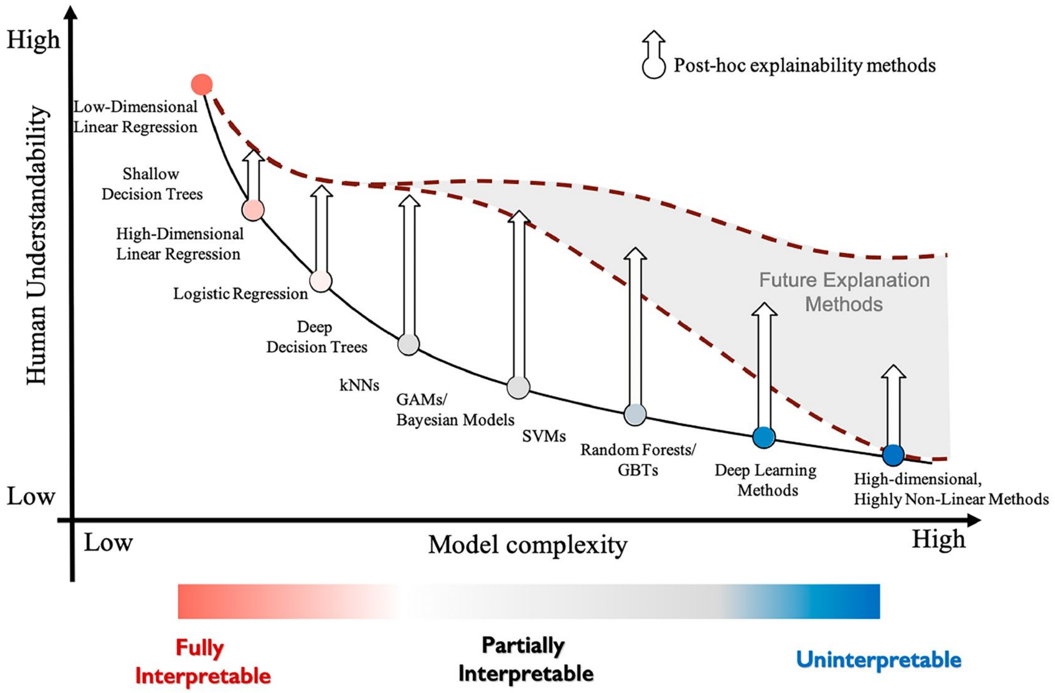


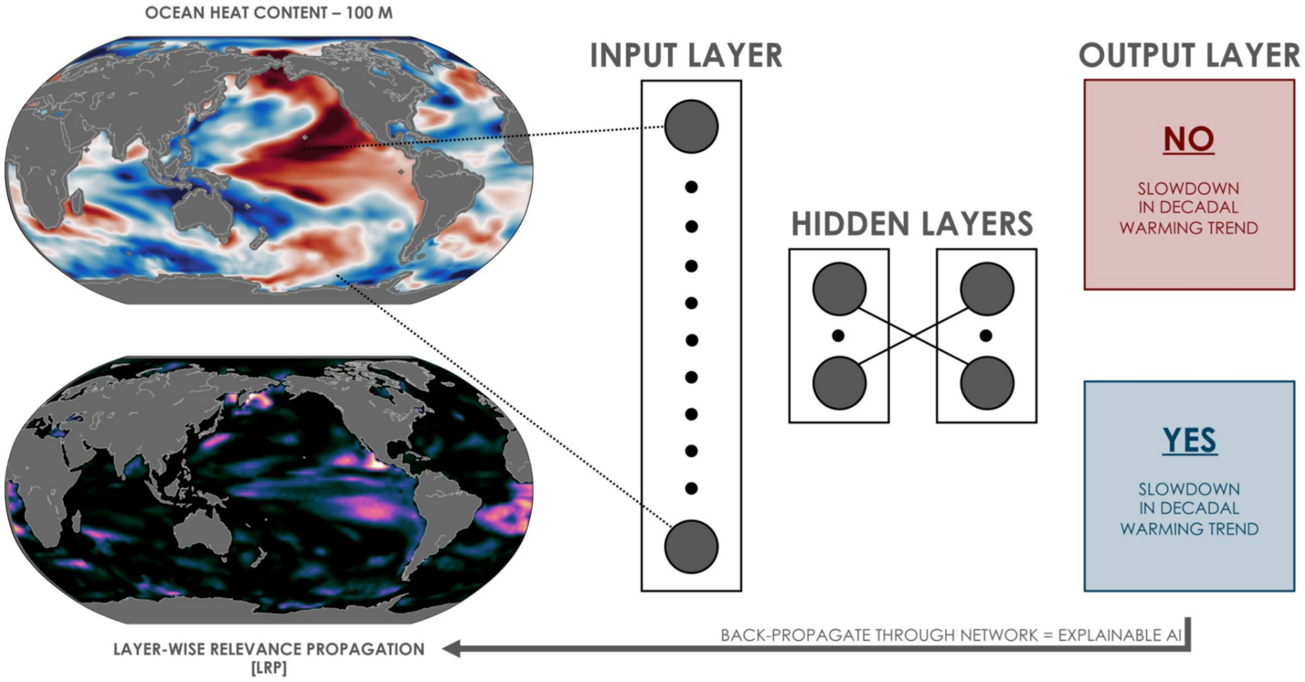
FIG. 1. Illustration of the relationship between understandability and model complexity. Fully interpretable models have high intrinsic understandability, while partially interpretable or simpler black box models have the most to gain from explainability methods. With increased dimensionality and nonlinearity, explainability methods can improve understanding. Still, there is considerable uncertainty about the ability of future explanation methods to improve the understandability of high-dimensional, highly nonlinear methods.

Models that are more complex tend to be less interpretable.

Flora et al. (2024)

Interpretability → eXplainable AI (XAI)

A neural network predicts slowdowns in decadal warming trends from maps of ocean heat content. ***eXplainable AI shows where the inputs were relevant.***

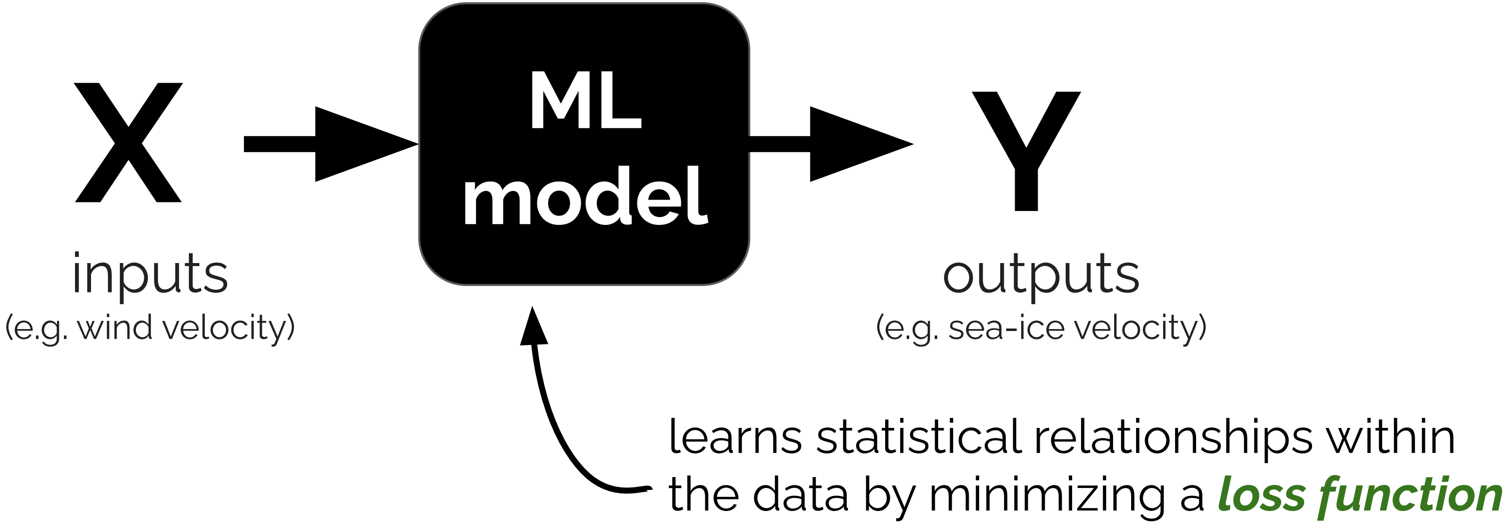


Labe & Barnes (2022)

But incorporating ML into Earth and climate science still has its challenges…

Challenges
Data Availability
Robustness on out-of-distribution samples
Interpretability
Physical Inconsistency
Uncertainty Quantification

Physical Inconsistency → Physics-Informed ML



We can apply a custom loss function that includes information about physics (i.e. conservation of mass, etc.)

But incorporating ML into Earth and climate science still has its challenges...

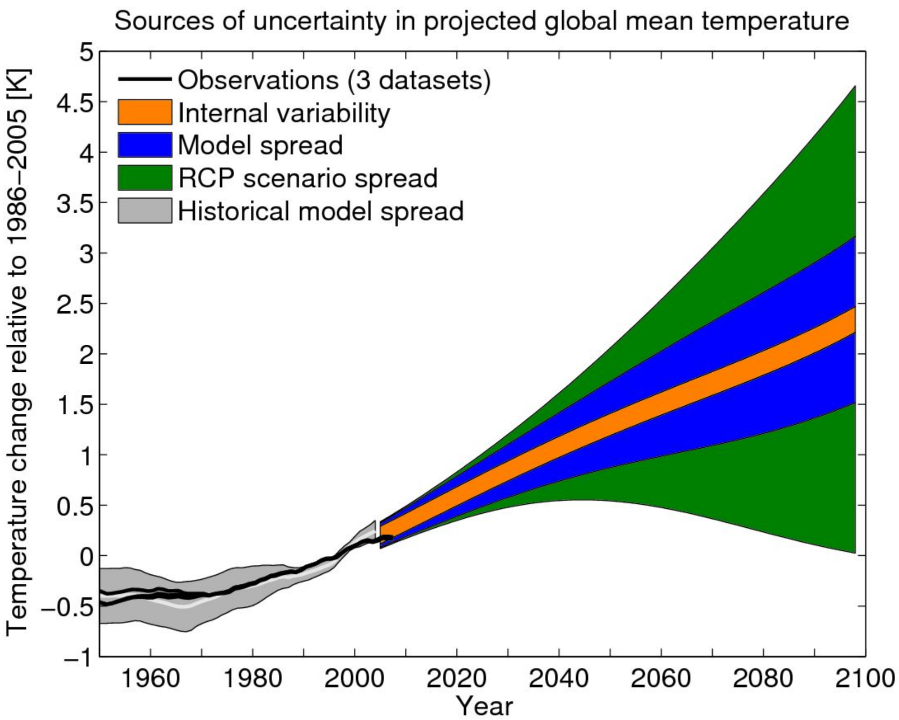
Challenges
Data Availability
Robustness on out-of-distribution samples
Interpretability
Physical Inconsistency
Uncertainty Quantification

Uncertainty Quantification

1. **Internal Variability:**
   Our knowledge of the current state is imperfect

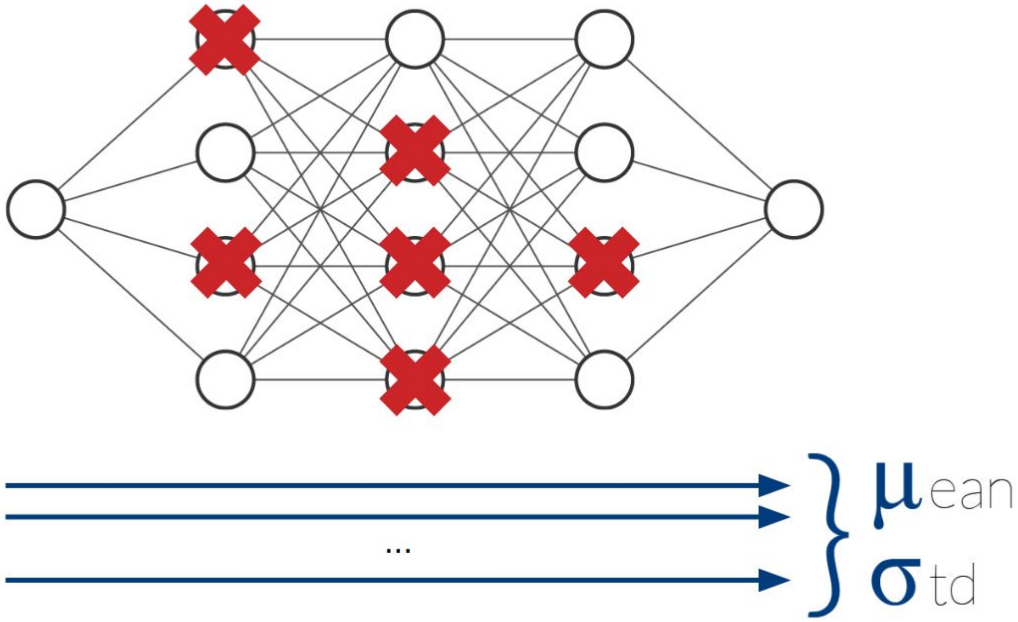
2. **Model Spread:**
   Our representation of nature is imperfect

3. **Scenario Spread:**
   Our knowledge on future boundary conditions is imperfect



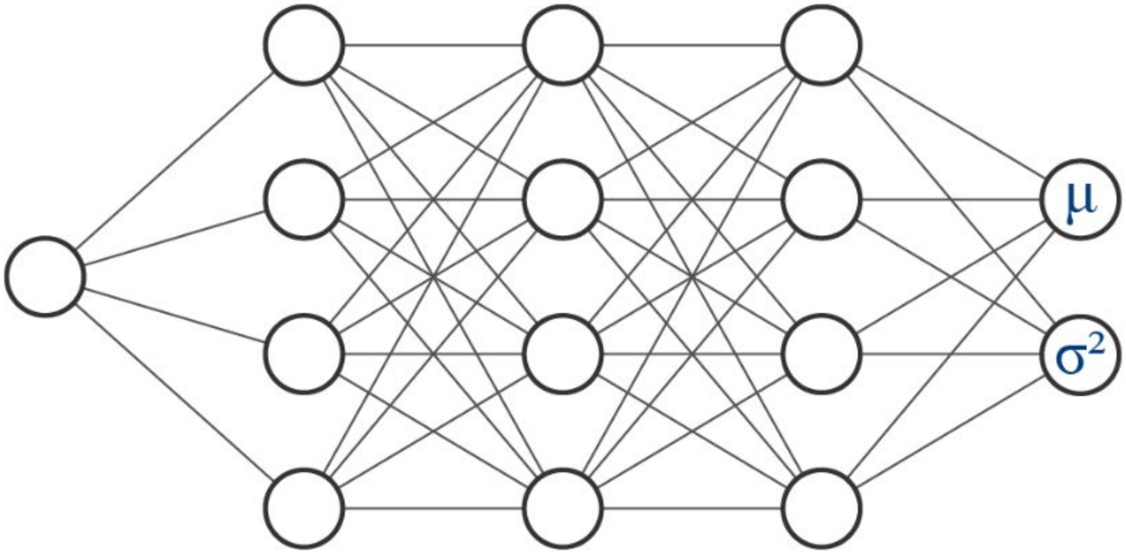
Ed Hawkings

Uncertainty Quantification → Monte Carlo Dropout



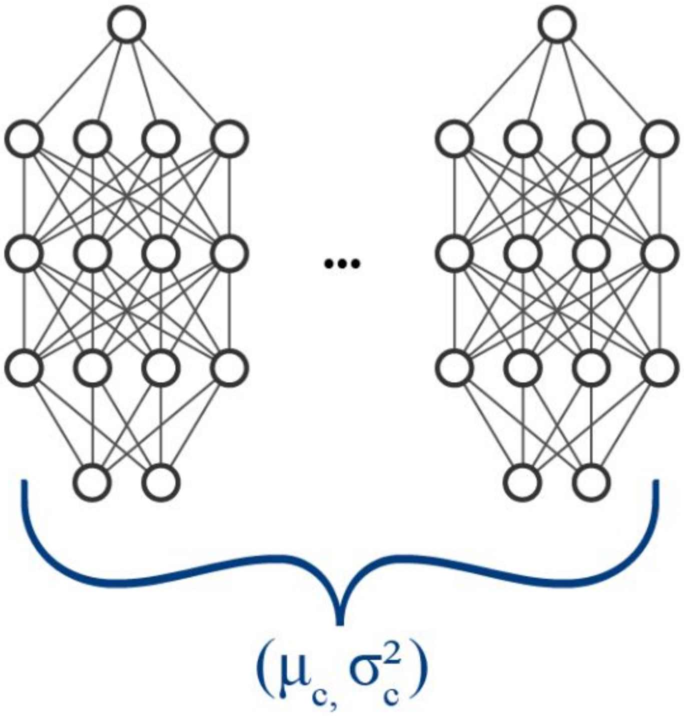
During Monte Carlo Dropout ***nodes are randomly disabled during training*** so that for each epoch a random subset of weights and biases is evaluated and adjusted.

Uncertainty Quantification → Distributional Parameter Estimation



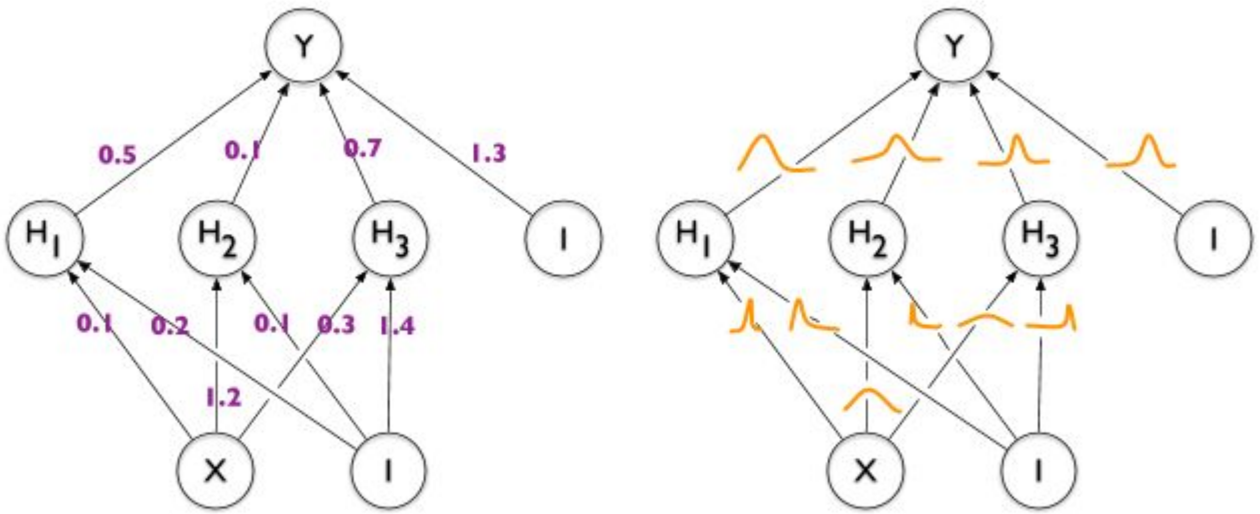
A neural network predicts the mean and standard deviation, rather than a deterministic value.

Uncertainty Quantification → Ensemble Averaging



Train an ensemble of neural networks with different initializations.

Uncertainty Quantification → Bayesian Neural Networks (BNNs)



In BNNs, the weights are distributions not deterministic values.

But incorporating ML into Earth and climate science still has its challenges...

Challenges	Solutions
Data Availability	Climate model output
Robustness on out-of-distribution samples	Transfer learning
Interpretability	eXplainable AI
Physical Inconsistency	Physics-informed ML
Uncertainty Quantification	BNNs, Ensembles, Dropout, etc.

The take-home messages...

There are many different applications for ML in Earth and climate sciences:

- Sources of predictability
- Forecasting
- Satellite data & observations
- Feature detection
- Earth System Modeling

Some of the challenges that we are currently facing applying ML in climate science have promising solutions that warrant further research:

- Data availability → use climate model output
- Robustness on out-of-distribution samples → transfer learning
- Interpretability → eXplainable AI
- Physical Inconsistency → Physics informed ML
- Uncertainty quantification → Bayesian NN, problem set-up

Lecture 1: What is the machine learning?

- Motivation
- Case study
- Types of ML
- How it works
- *Tutorial: Building a ML model*

Lecture 2: ML Applications in Earth and Climate Science

- Applications
- Challenges
- *Activity: literature dive*

https://github.com/lahoffman/ml_lectures_LPHYS2268