MAPP NA19OAR4310289

CAREER AGS-1749261
HDR OAC-1934668
AI Institute ICER-2019758

Explainable / Interpretable AI for Climate Science

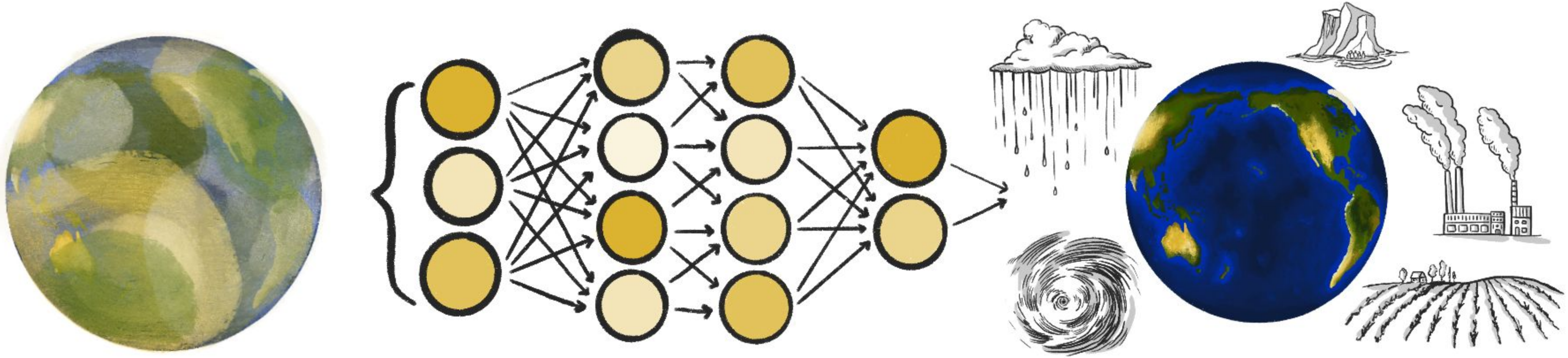
Dr. Elizabeth A. Barnes
Associate Professor
Department of Atmospheric Science
Colorado State University

ATMOSPHERIC SCIENCE
COLORADO STATE UNIVERSITY

WCRP Workshop on Future of Climate Modeling
March 22, 2022

Prof. Elizabeth A. Barnes
eabarnes@colostate.edu
https://barnes.atmos.colostate.edu
@atmosbarnes


MACHINE LEARNING
CLIMATE SCIENCE TOOLS



ML for Climate Science

The field's interest, and research, has exploded in the past ~3 years!

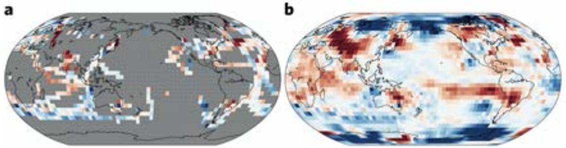
Applications of ML for atmospheric science dates back as far as the 1960's!

Range of applications:

- global weather prediction
- convective & radiative parameterizations
- downscaling
- extreme event detection
- data reconstruction
- weather prediction
- processing of remote sensing data

Climate Reconstruction

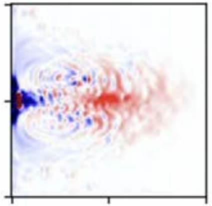
e.g. Kadow et al. (2020), DelSole and Nedza (2020)



Kadow et al. (2020)

Equation discovery

e.g. Zanna and Bolton (2020)



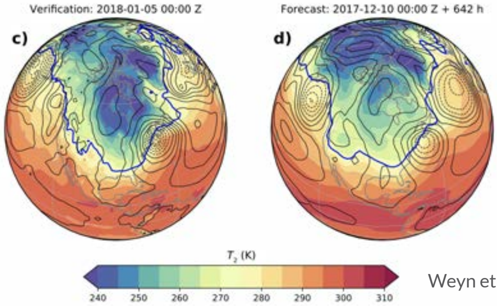
$$\hat{\mathbf{S}}_{\mathbf{u}}^{BT} \approx \kappa_{BT}\overline{\nabla}\cdot\begin{pmatrix}\zeta^2-\zeta D & \zeta\tilde{D}\\ \zeta\tilde{D} & \zeta^2+\zeta D\end{pmatrix}$$

Climate change communication

Mila

Weather Prediction

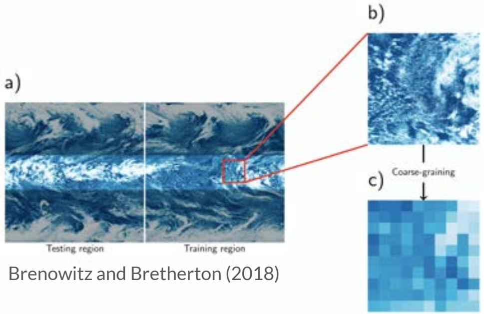
e.g. Gagne et al. (2019); Gagne et al. (2017); Chattopadhyay et al. (2019); Lagerquist et al. (2020)



Weyn et al. (2020)

Convective parameterizations

e.g. Rasp et al. (2018; PNAS); Schneider et al. (2017; GRL); O'Gorman and Dwyer (2018); Beucler et al. (2020; PRL)



Brenowitz and Bretherton (2018)

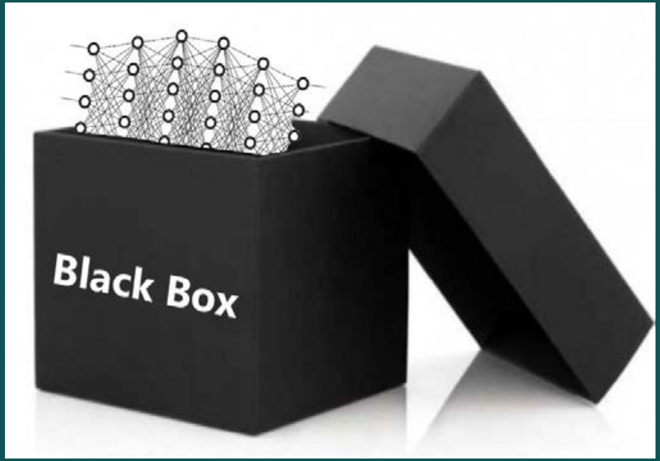
Reasons to use AI for climate science

- Do it better
  - e.g. convective parameterizations in models are not perfect, use ML to make them more accurate
- Do it faster or cheaper
  - e.g. radiation code in models is very slow - use ML methods to speed things up
- Learn something new
  - e.g. go looking for non-linear relationships you didn't know were there

Very relevant for research: may be slower and worse, but can still learn something (more to come on this...)

Opening the "Black Box"

Leveraging advances in
explainable / interpretable AI



https://www.code4thought.eu/index.php/2019/10/03/using-explanations-for-finding-bias-in-black-box-models/

Opening the Black Box

In the past few years multiple papers have come out demonstrating the use of ML explainability methods for geoscience

MAKING THE BLACK BOX MORE TRANSPARENT

Understanding the Physical Implications of Machine Learning

Amy McGovern, Ryan Lagerquist, David John Gagne II, G. Eli Jergensen, Kimberly L. Elmore, Cameron R. Homeyer, and Travis Smith

Machine learning model interpretation and visualization focusing on meteorological domains are introduced and analyzed.

JAMES Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE

Physically Interpretable Neural Networks for the Geosciences: Applications to Earth System Variability

Benjamin A. Toms, Elizabeth A. Barnes, and Imme Ebert-Uphoff

arXiv > physics > arXiv:2103.10005

Physics > Geophysics

[Submitted on 18 Mar 2021]

Neural Network Attribution Methods for Problems in Geoscience: A Novel Synthetic Benchmark Dataset

Antonios Mamalakis, Imme Ebert-Uphoff, Elizabeth A. Barnes

arXiv > physics > arXiv:2202.03407

Physics > Geophysics

[Submitted on 7 Feb 2022]

Investigating the fidelity of explainable artificial intelligence methods for applications of convolutional neural networks in geoscience

Antonios Mamalakis, Elizabeth A. Barnes, Imme Ebert-Uphoff

Convolutional neural networks (CNNs) have recently attracted great attention in geoscience due to their ability to capture non-linear system behavior and extract predictive spatiotemporal patterns. Given their black-box nature however, and the importance of prediction explainability, methods of explainable artificial intelligence (XAI) are gaining popularity as a means to explain the CNN decision-making strategy. Here, we establish an intercomparison of some of the most popular XAI methods and investigate their fidelity in explaining CNN decisions for geoscientific applications. Our goal is to raise awareness of the theoretical limitations of these methods and gain insight into the relative strengths and weaknesses to help guide best practices. The considered XAI methods are first applied to an idealized attribution benchmark, where the ground truth of explanation of the network is known a priori, to help objectively assess their performance. Secondly, we apply XAI to a climate-related prediction setting, namely to explain a CNN that is trained to predict the number of atmospheric rivers in daily snapshots of climate simulations. Our results highlight several important issues of XAI methods (e.g., gradient shattering, inability to distinguish the sign of attribution, ignorance to zero input) that have previously been overlooked in our field and, if not considered cautiously, may lead to a distorted picture of the CNN decision-making strategy. We envision that our analysis will motivate further investigation into XAI fidelity and will help towards a cautious implementation of XAI in geoscience, which can lead to further exploitation of CNNs and deep learning for prediction problems.

BAMS

RESEARCH ARTICLE | 31 AUGUST 2020

Evaluation, Tuning and Interpretation of Neural Networks for Working with Images in Meteorological Applications

Imme Ebert-Uphoff; Kyle Hilburn

Bull. Amer. Meteor. Soc. 1–49.

https://doi.org/10.1175/BAMS-D-20-0097.1

Capsule:

This article discusses strategies for the development of neural networks (aka *deep learning*) for meteorological applications. Topics include evaluation, tuning and interpretation of neural networks for working with meteorological images.

Abstract

The method of neural networks (aka deep learning) has opened up many new opportunities to utilize remotely sensed images in meteorology. Common applications include image classification, e.g., to determine whether an image contains a tropical cyclone, and image-to-image translation, e.g., to emulate radar imagery for satellites that only have passive channels. However, there are yet many open questions regarding the use of neural networks for working with meteorological images, such as best practices for evaluation, tuning and interpretation. This article highlights several strategies and practical considerations for neural network development that have not yet received much attention in the meteorological community, such as the concept of receptive fields, underutilized meteorological performance measures, and methods for neural network interpretation, such as synthetic experiments and layer-wise relevance propagation. We also consider the process of neural network interpretation as a whole, recognizing it as an iterative meteorologist-driven discovery process that builds on experimental design and hypothesis generation and testing. Finally, while most work on neural network interpretation in meteorology has so far focused on networks for image classification tasks, we expand the focus to also include networks for image-to-image translation.

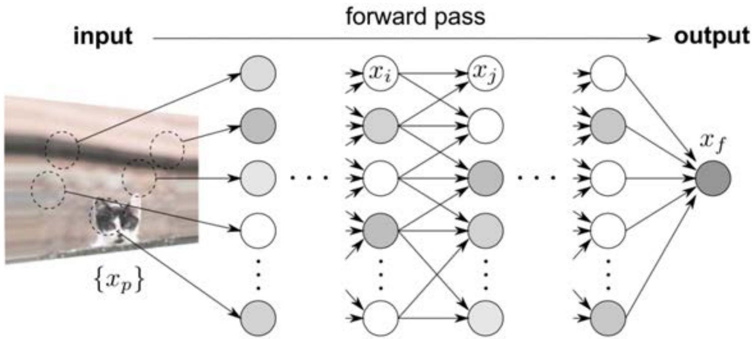
XAI Attribution Methods

Attribution methods produce a heatmap of the most relevant regions of the input for each prediction

Attribution heatmaps are largely consistent with how many climate scientists pose questions

Prediction

of 1 sample



Pr(cat)=.8

e.g. Montavon et al. (2017), Pattern Recognition; Montavon et al. (2018), Digital Signal Processing

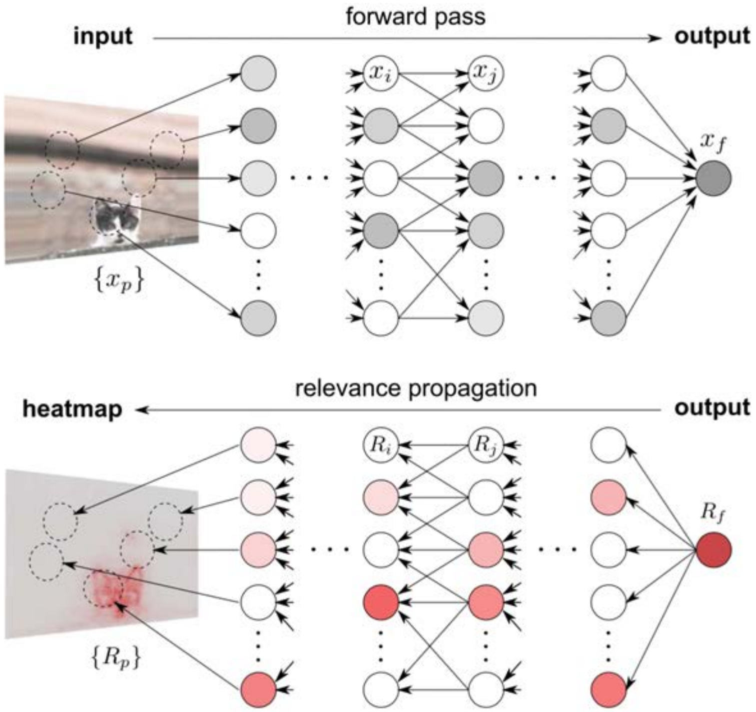
XAI Attribution Methods

Attribution methods produce a heatmap of the most relevant regions of the input for each prediction

Attribution heatmaps are largely consistent with how many climate scientists pose questions

Prediction
of 1 sample

Attribution
of 1 sample



Pr(cat)=.8

Pr(cat)=.8

Montavon et al. (2017), Pattern Recognition; Montavon et al. (2018), Digital Signal Processing

Kirsten Mayer

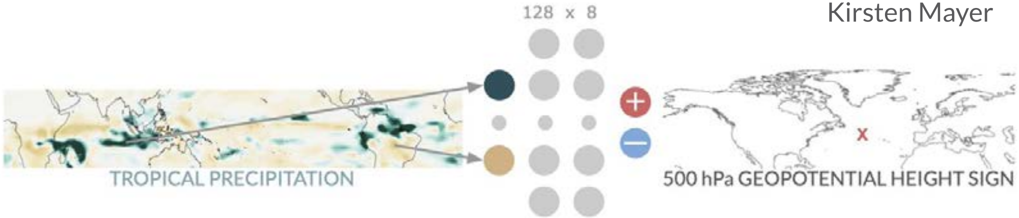
XAI Attribution Methods

Attribution methods produce a heatmap of the most relevant regions of the input for each prediction

Attribution heatmaps are largely consistent with how many climate scientists pose questions

Prediction

of 1 sample



Mayer and Barnes (2022)

XAI Attribution Methods

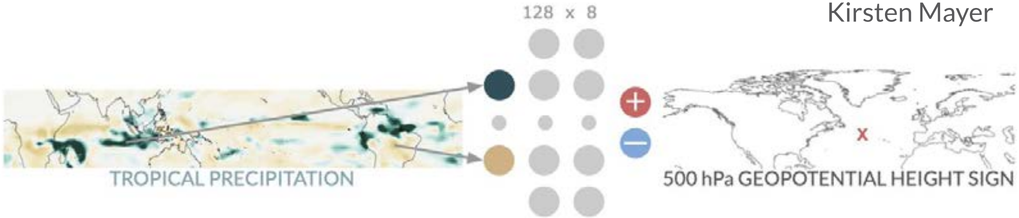
Attribution methods produce a heatmap of the most relevant regions of the input for each prediction

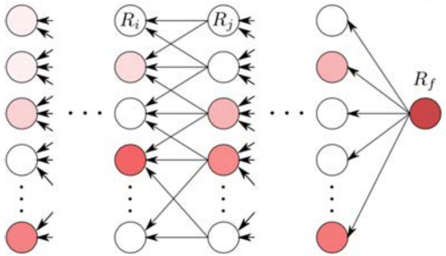
Attribution heatmaps are largely consistent with how many climate scientists pose questions

Kirsten Mayer

Prediction

of 1 sample





XAI

Mayer and Barnes (2022)

XAI Attribution Methods

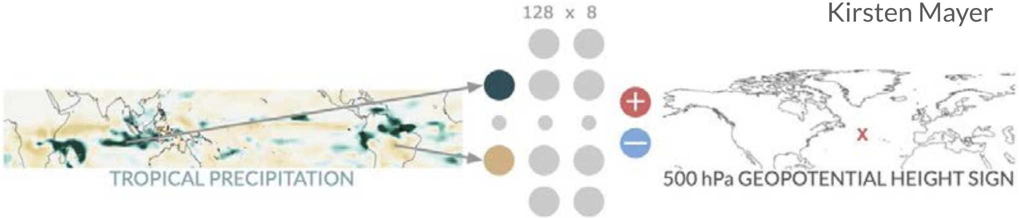
Attribution methods produce a heatmap of the most relevant regions of the input for each prediction

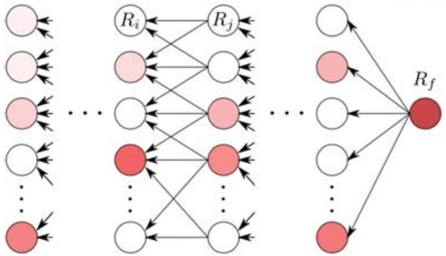
Attribution heatmaps are largely consistent with how many climate scientists pose questions

Kirsten Mayer

Prediction

of 1 sample

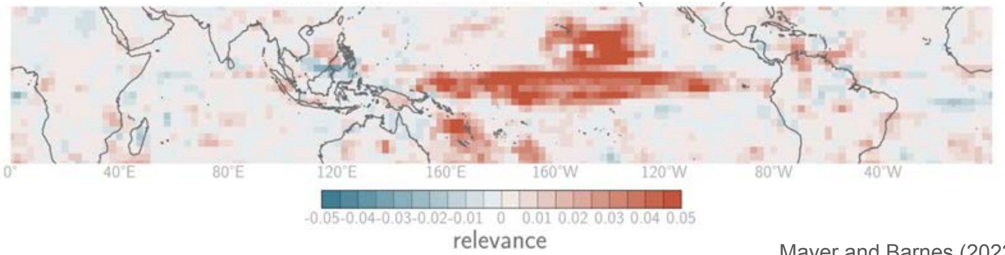




XAI

Attribution

of 1 sample



Mayer and Barnes (2022)

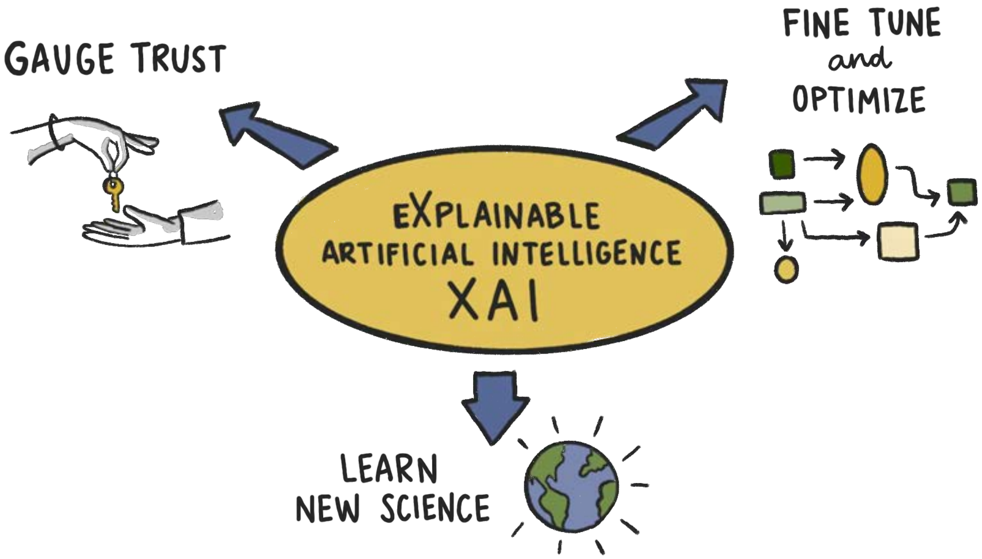
COLORADO STATE UNIVERSITY

Why you should care about XAI

As scientists our ultimate goal is to understand “why?”. But even if you don’t care “why” you should still care about XAI.

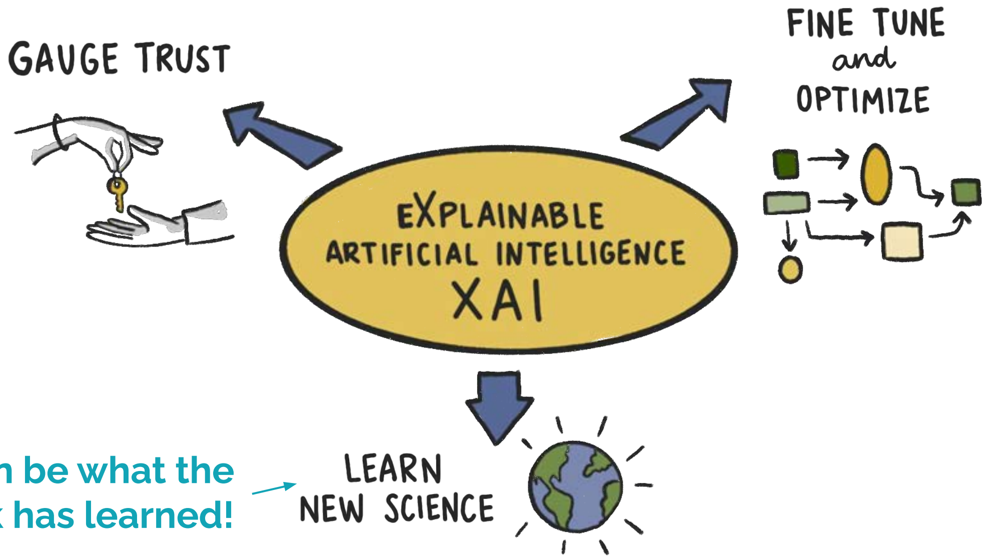
Why you should care about XAI

As scientists our ultimate goal is to understand “why?”. But even if you don’t care “why” you should still care about XAI.



Why you should care about XAI

As scientists our ultimate goal is to understand “why?”. But even if you don’t care “why” you should still care about XAI.



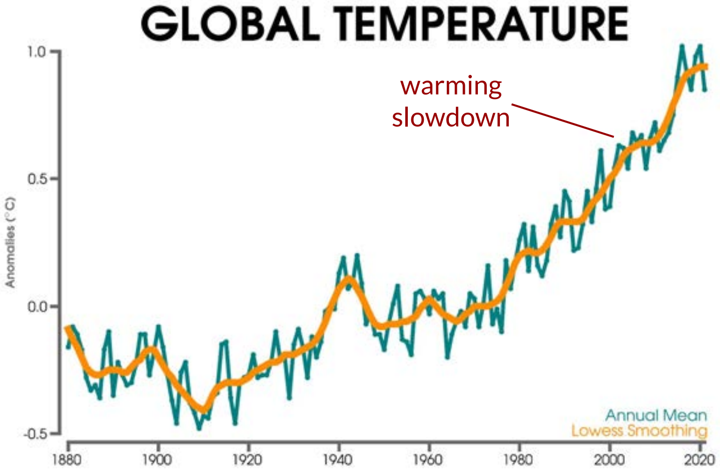
The science can be what the network has learned!

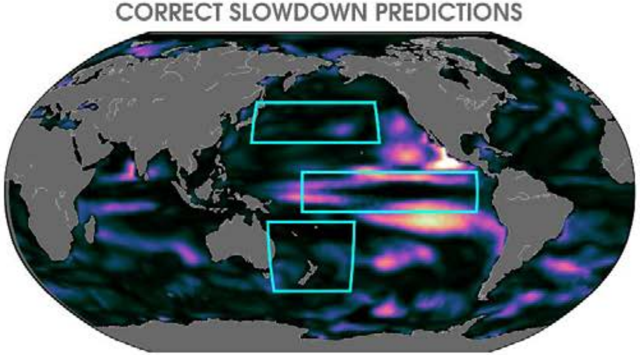
XAI for climate change & variability

Train a neural network to predict temporary slowdowns in global mean surface temp.

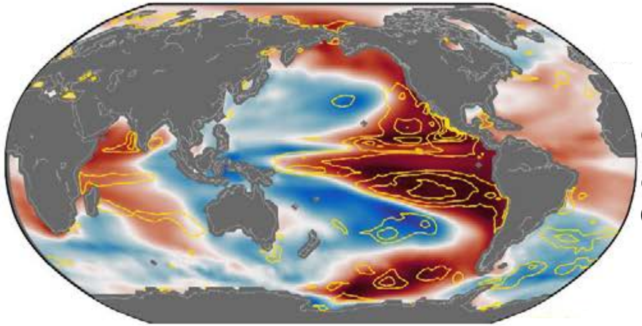

Dr. Zack Labe

- ANN learns patterns of upper ocean heat content associated with decadal slowdowns in both climate model data and observations
- XAI reveals the ANN is learning off-equatorial patterns of anomalous ocean heat content that resemble **transitions in the phase of the Interdecadal Pacific Oscillation**





XAI

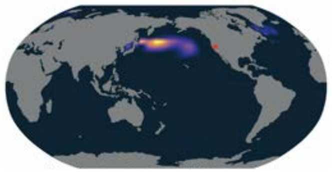


ocean heat content
(0-100 m)

Labe, Z.M. and E.A. Barnes (2021),
https://doi.org/10.1002/essoar.10508874.1

Other example uses of XAI @ CSU (only a subset)

Subseasonal-to-decadal predictability




Emily Gordon


Zane Martin

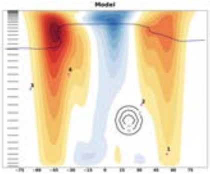

Kirsten Mayer


Jack Cahill


Zaibeth Carlo Frontera

Exploring subseasonal-to-decadal climate dynamics with implications for prediction, scientific mechanisms, and basic theory

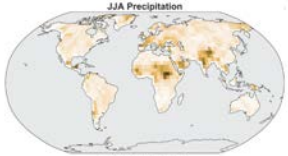
Forced response of midlatitude dynamics




Charlie Connolly

Understanding basic general circulation responses to climate forcings *[figure from Baker et al. 2017]*

Indicator patterns of forced change



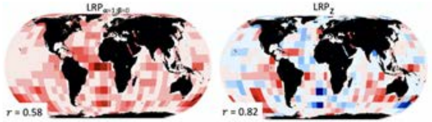

Zack Labe


Jamin Rader


Prof. James Hurrell

Learn non-linear, time-evolving patterns of forced change in climate simulations and observations

XAI benchmarking & robust predictions




Antonios Mamalakis


Dr. Imme Ebert-Uphoff


Nico Gordillo

Develop robust AI methods and benchmarks for XAI method evaluation and comparison

Exciting Frontiers

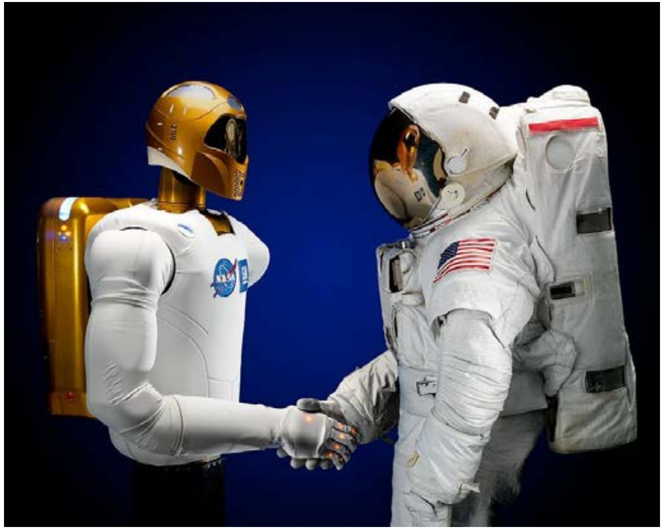
#1

Knowledge-guided machine learning

Continue fusing scientific knowledge and AI for climate science

- the availability of extensive existing knowledge
- the desire of Earth scientists to gain scientific insights rather than just “get numbers” from an algorithm
- the high complexity of the Earth system
- the limited sample size and lack of reliable labels in many Earth science applications
- improves transparency and trustworthiness

Make physics and ML work together



https://www.pxfuel.com/en/free-photo-oadru

For more discussion, see Ebert-Uphoff et al. (2019)
Karpatne et al. (2017)

#2 Transfer learning

Leverage imperfect climate model output through a transfer learning framework

For many earth science applications we have very little observational data

...BUT...

We have thousands of years of imperfect earth system model simulations from which ML tools can learn.

Step 1: Train ML model with climate model simulations

Step 2: Update the ML model with data from observations

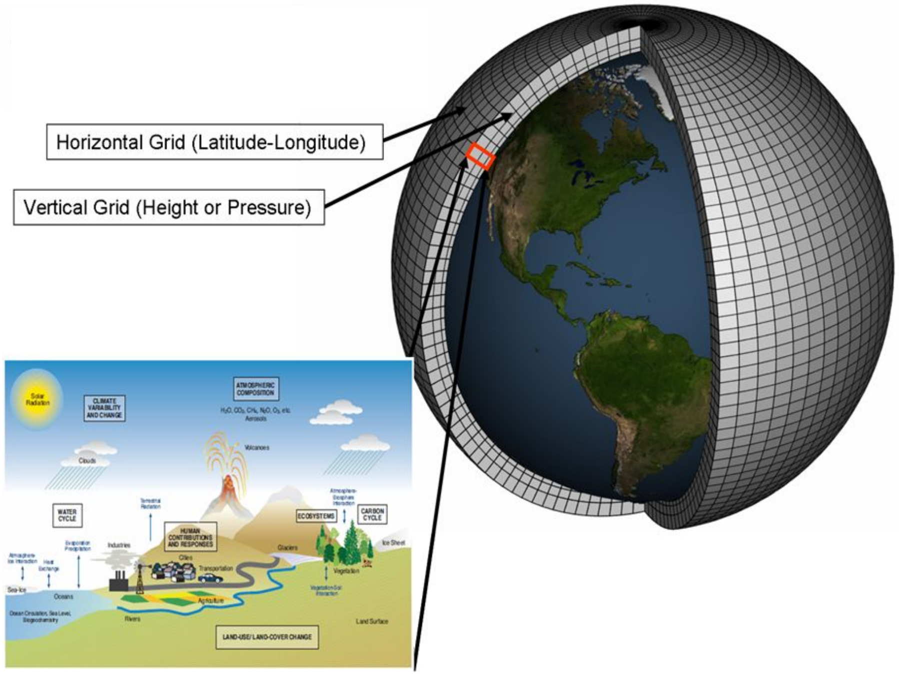
+

End Result: A trained prediction model that leverages dynamical simulations but applies better to the real world

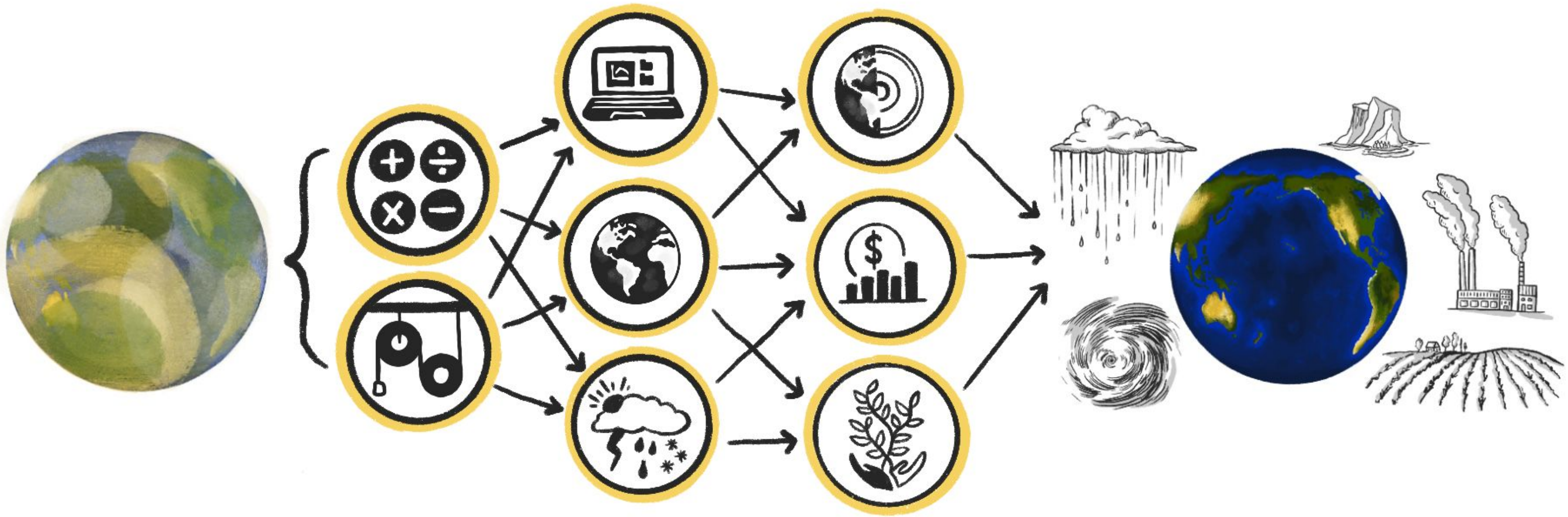
#3 Improve climate projections

Bring ML methods into the building, evaluation and use of climate model projections

- There is great promise for improving climate models through ML-developed **convective/radiative parameterizations**
- ML for model **comparisons** and **evaluation** against observations.
- ML to explore **bias correcting / transforming** climate model projections to **narrow uncertainties**



Climate science requires the **mixing of knowledge** from many fields. And ultimately we want more than just a prediction - we want to know "**why?**"



Explainable / interpretable ML is a game changer for climate research.