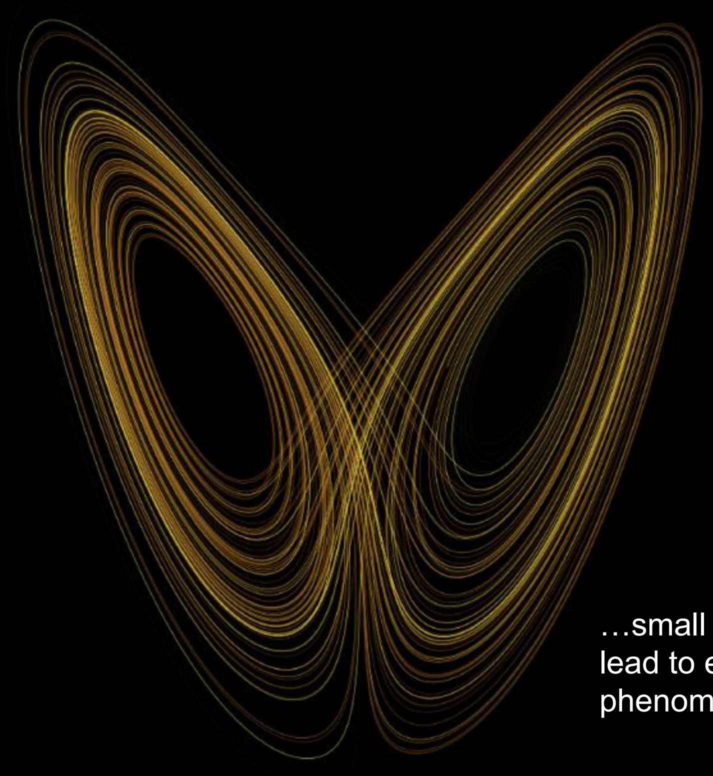
AGS CLD CAREER Award
Graduate Research Fellowship
Postdoctoral Fellowship

Benefits of saying "I Don't Know" when analyzing and modeling the climate system with ML

Prof. Elizabeth A. Barnes

STAMPS - Carnegie Melon University
December 3, 2021

CHAOS



When the flap of a butterfly's wings in Brazil sets off a tornado in Texas.
- Edward Lorenz (1972)

…small differences in the initial positions may lead to enormous differences in the final phenomena. ***Prediction becomes impossible.***
- Henri Poincare (1903)

The earth system is exceedingly complex and often chaotic in nature, making prediction incredibly challenging.

We cannot expect to make perfect predictions all of the time...

Forecasts of Opportunity*

certain conditions lead to more predictable behaviour than others

Beyond the weather timescale we must look for specific states of the earth system, i.e. "opportunities", that lead to enhanced predictable behavior.

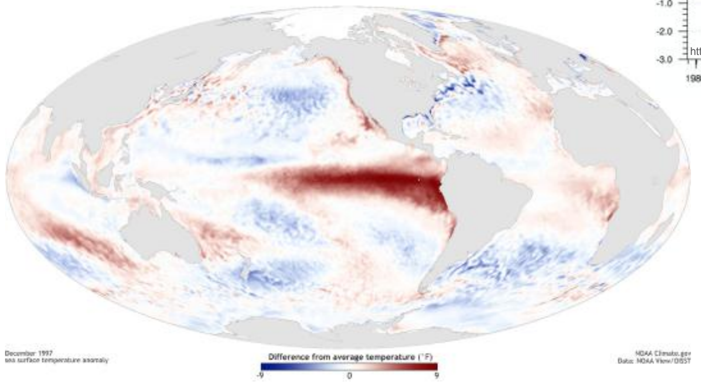
** also known as "State-dependent predictability"*
See also Mariotti et al. (2020) and also Albers and Newman (2019)

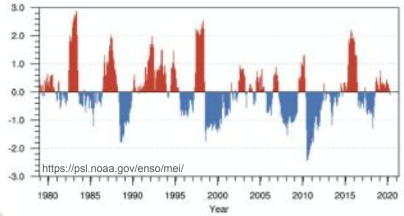
COLORADO STATE UNIVERSITY

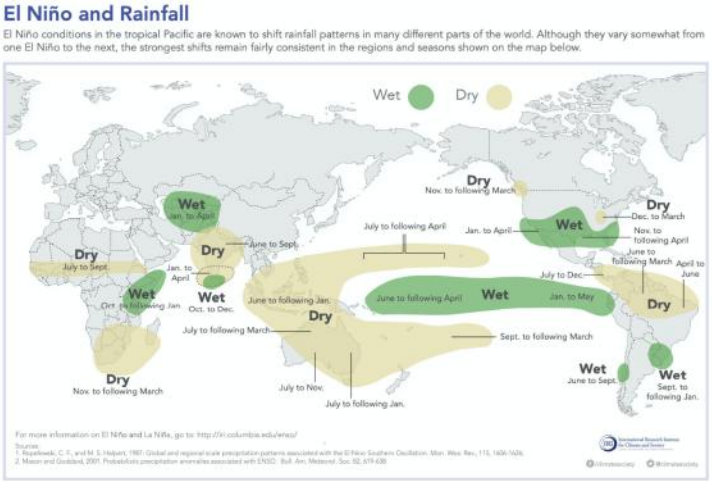
El Nino Southern Oscillation [ENSO]

- Timescales of seasons-to-years
- Long-studied tropical phenomenon that, **when active**, impacts weather across the globe

Sea surface temperature anomalies
Dec. 1997 El Nino







Finding forecasts of opportunity

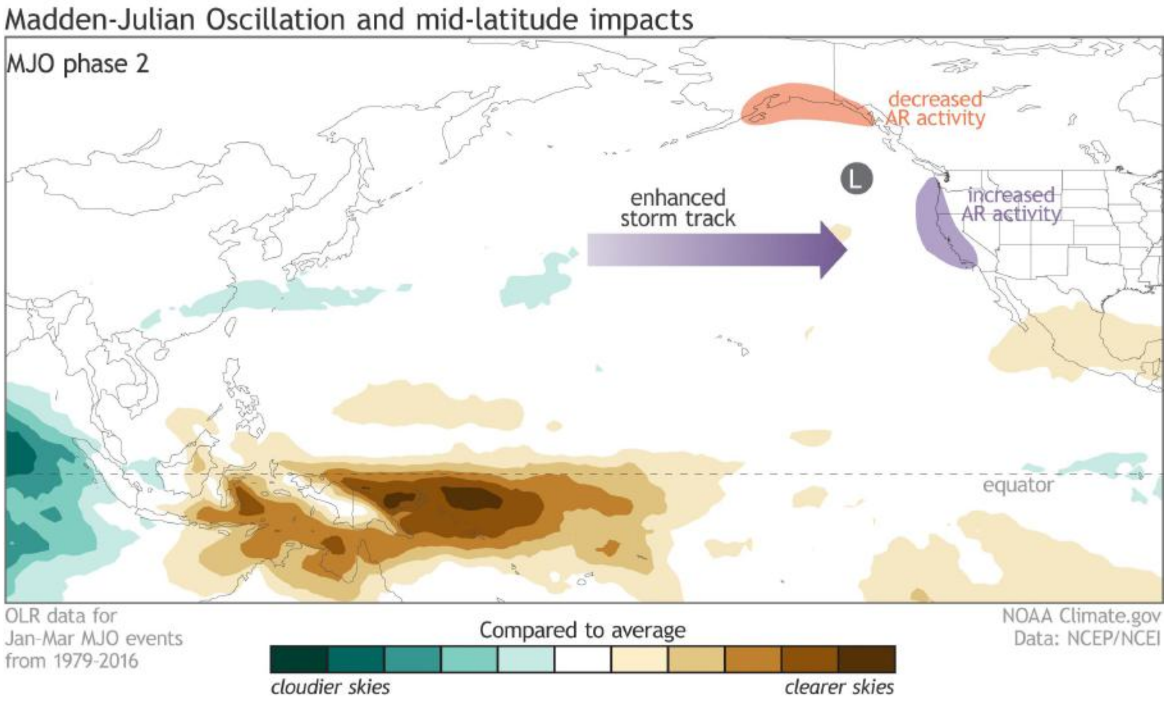
1. **When?** Under what conditions do we have skillful forecasts of opportunity?

2. **Why?** Where is this predictability coming from?

3. **How** do we leverage these opportunities?

Global teleconnections

Climate phenomena can influence weather across the globe via atmospheric teleconnections



COLORADO STATE UNIVERSITY

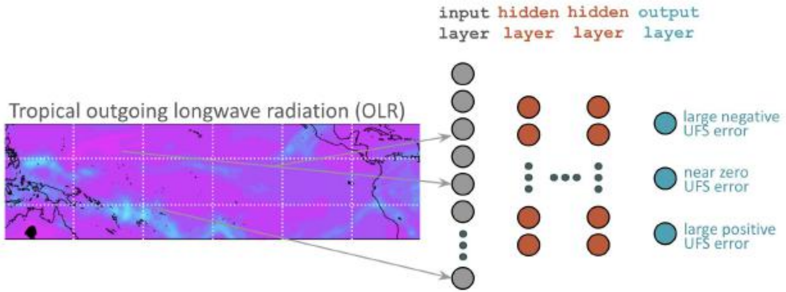
Post-processing by predicting forecast errors

Identifying state-dependent forecasts errors in the NOAA Unified Forecast System

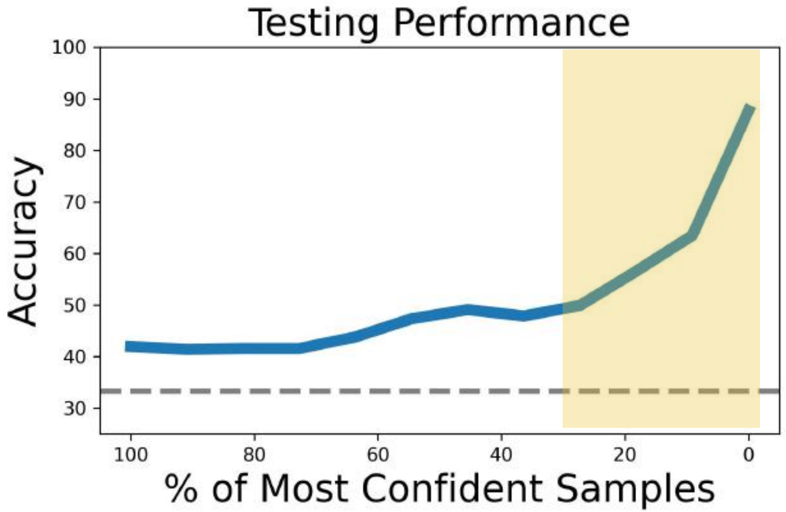
Network Task: Train a neural network to ingest daily maps of outgoing longwave radiation (OLR) to predicting the 5-day averaged precipitation error elsewhere at 10-14 day lead

Forecasts of Opportunity: Confident predictions lead to more accurate predictions = forecasts of opportunity

Explainable AI Approach: Learn tropical patterns of variability that lead to predictable forecast errors



Jack Cahill



COLORADO STATE UNIVERSITY

Preliminary work by Jack Cahill
Co-advised by E. Barnes and E. Maloney, CSU

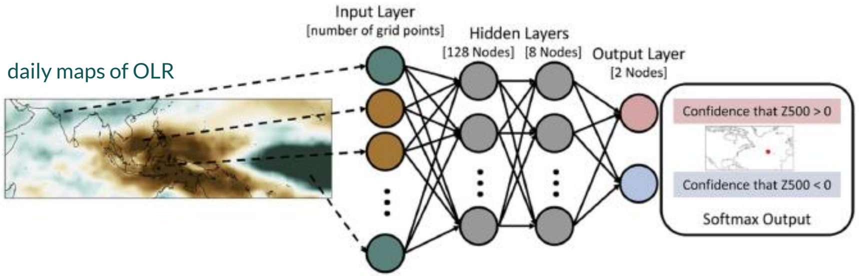
Subseasonal weather prediction

Exploring how tropical information can lend predictability to midlatitude circulation on S2S timescales

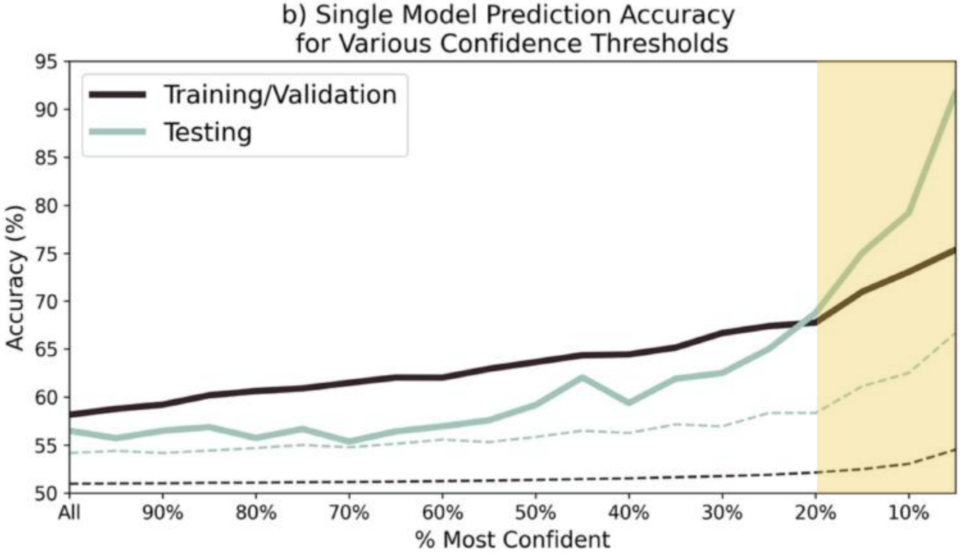
Network Task: Train a neural network to ingest daily maps of outgoing longwave radiation (OLR) to predict the sign of the subseasonal circulation anomalies over the North Atlantic 22 days in advance

Forecasts of Opportunity: Confident predictions lead to more accurate predictions = forecasts of opportunity

Explainable AI Approach: Learn tropical patterns of variability that lead to enhanced predictability of midlatitude weather on subseasonal timescales



Kirsten Mayer



Mayer and Barnes (2021)

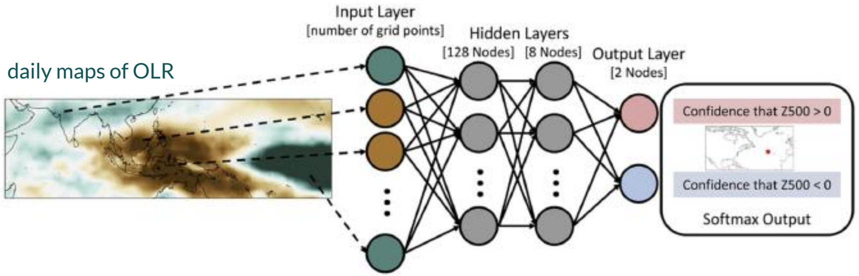
Subseasonal weather prediction

Exploring how tropical information can lend predictability to midlatitude circulation on S2S timescales

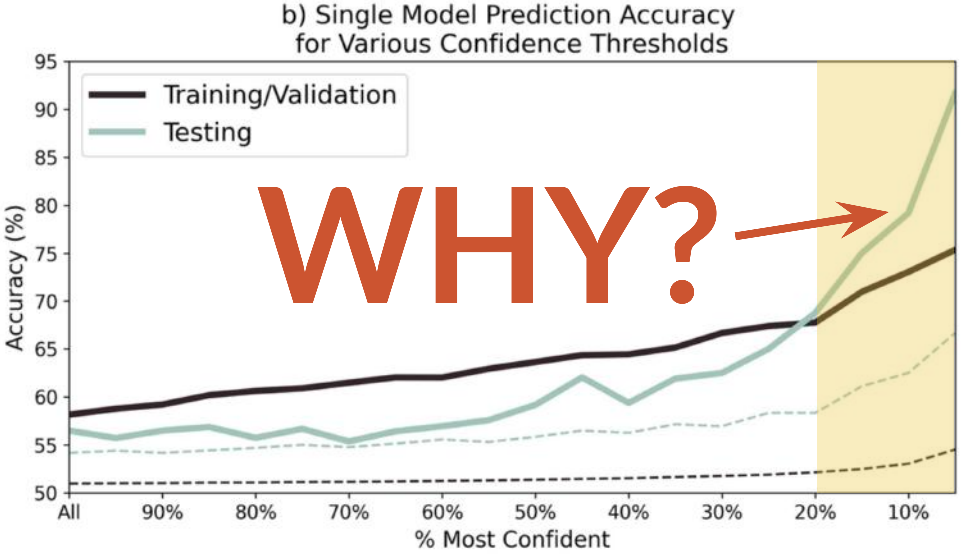
Network Task: Train a neural network to ingest daily maps of outgoing longwave radiation (OLR) to predict the sign of the subseasonal circulation anomalies over the North Atlantic 22 days in advance

Forecasts of Opportunity: Confident predictions lead to more accurate predictions = forecasts of opportunity

Explainable AI Approach: Learn tropical patterns of variability that lead to enhanced predictability of midlatitude weather on subseasonal timescales



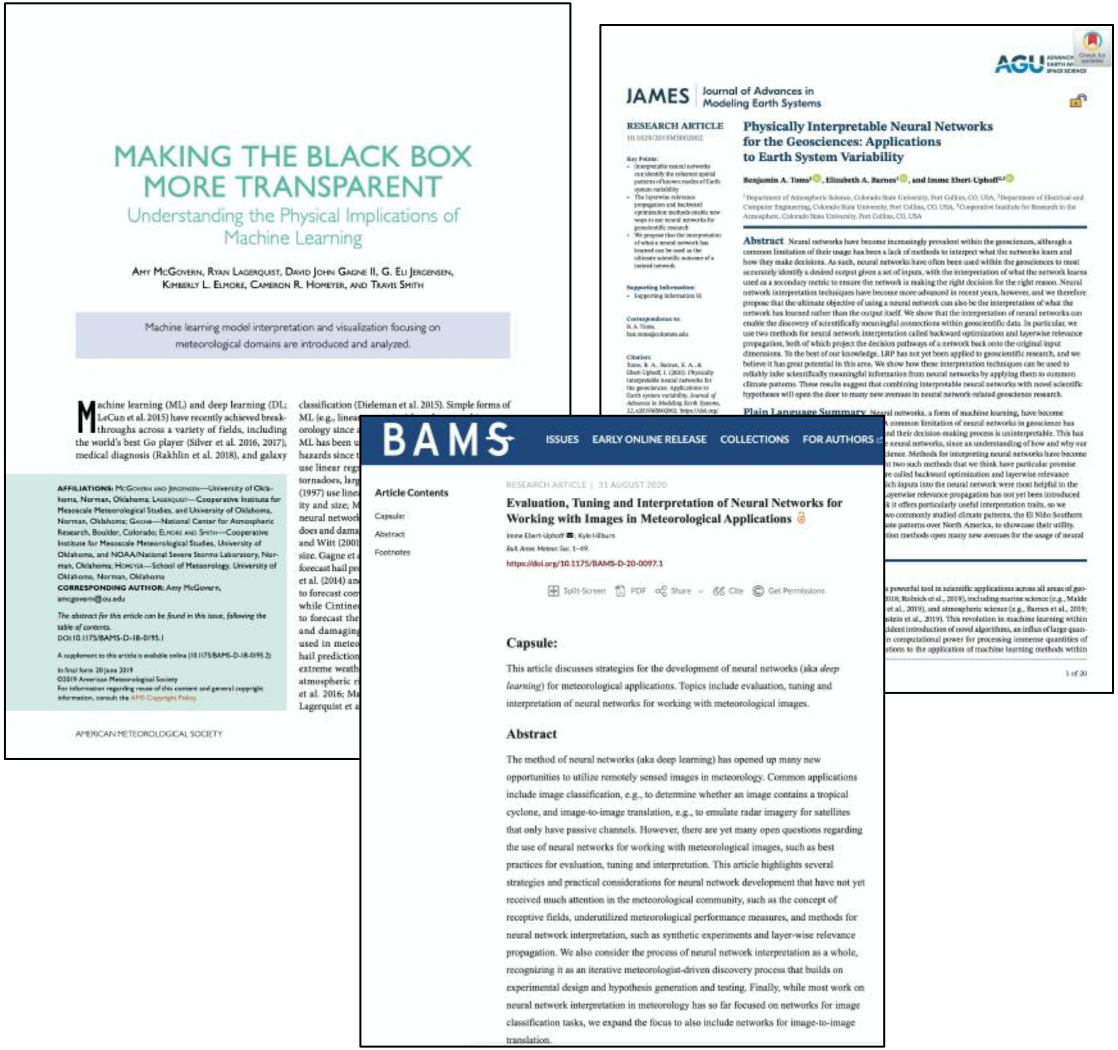
Kirsten Mayer



Mayer and Barnes (2021)

Opening the Black Box with XAI

In the past few years multiple papers have come out demonstrating the use of explainable AI (XAI) 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.

AMERICAN METEOROLOGICAL SOCIETY

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

BAMS

ISSUES EARLY ONLINE RELEASE COLLECTIONS FOR AUTHORS

Article Contents

Capsule:

Abstract

Footnotes

RESEARCH ARTICLE | 31 AUGUST 2020

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

Imme Ebert-Uphoff; Kyle Hilburn

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.

COLORADO STATE UNIVERSITY

Layerwise Relevance Propagation

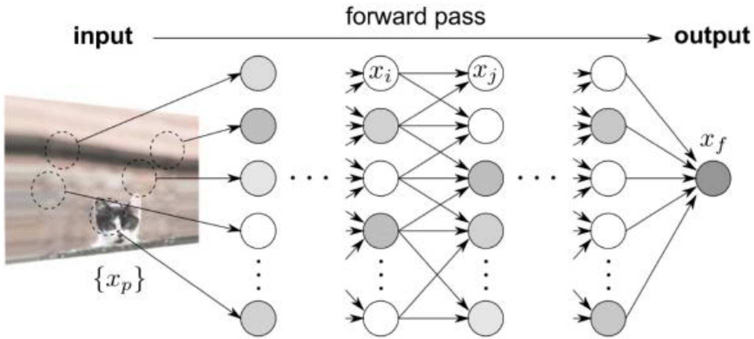
LRP is an XAI method that produces a heatmap of the most relevant regions of the input for each prediction

LRP is largely consistent with how many climate scientists analyze and interpret data methods

While many visualization tools are coming out of the computer science community, LRP has been most useful for our group thus far

Prediction

of 1 sample



Pr(cat)=.8

Layerwise Relevance Propagation

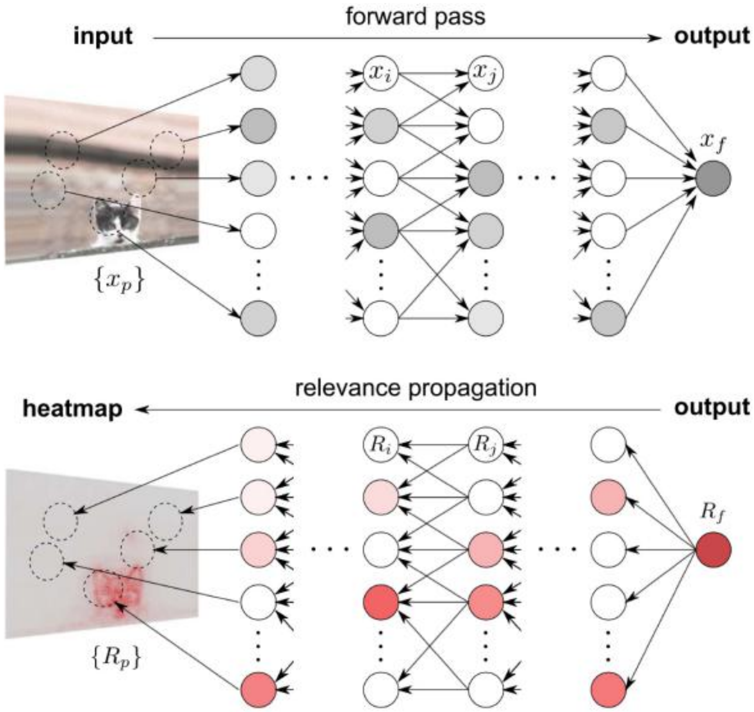
LRP is an XAI method that produces a heatmap of the most relevant regions of the input for each prediction

LRP is largely consistent with how many climate scientists analyze and interpret data methods

While many visualization tools are coming out of the computer science community, LRP has been most useful for our group thus far

Prediction
of 1 sample

LRP
of 1 sample



Pr(cat)=.8

Pr(cat)=.8

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

Why use XAI?

One use of LRP is to ensure the right answers for the right reasons

Example Task:
Decide whether there is a horse in a given image.

Interpretable AI Approach:
What strategy did the network use? Is it focusing on the right things?

Lapuschkin et al. “Unmasking Clever Hans Predictors and Assessing What Machines Really Learn.” Nature Communications, vol. 10, no. 1, Mar. 2019, p. 1096, doi:10.1038/s41467-019-08987-4.

Why use XAI?

One use of LRP is to ensure the right answers for the right reasons

Example Task:
Decide whether there is a horse in a given image.

Interpretable AI Approach:
What strategy did the network use? Is it focusing on the right things?

Lapuschkin et al. “Unmasking Clever Hans Predictors and Assessing What Machines Really Learn.” Nature Communications, vol. 10, no. 1, Mar. 2019, p. 1096, doi:10.1038/s41467-019-08987-4.

Why use XAI?

#1 Identify problematic strategies

#2 Evaluate trust

#3 Choose the approach

#4 Learn something new

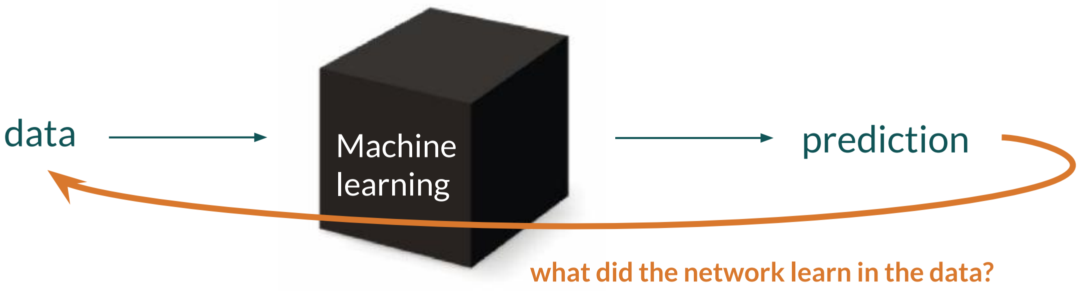
Why use XAI?

#1 Identify problematic strategies

#2 Evaluate trust

#3 Choose the approach

#4 Learn something new



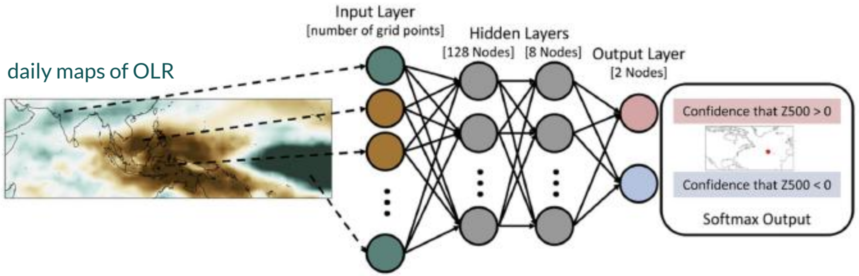
Subseasonal weather prediction

Exploring how tropical information can lend predictability to midlatitude circulation on S2S timescales

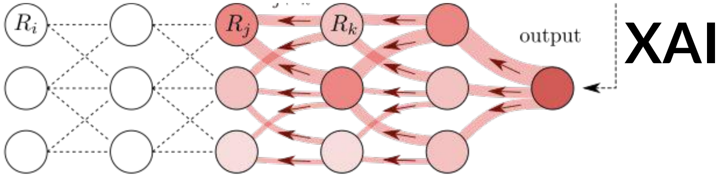
Network Task: Train a neural network to ingest daily maps of outgoing longwave radiation (OLR) to predict the sign of the subseasonal circulation anomalies over the North Atlantic 22 days in advance

Forecasts of Opportunity: Confident predictions lead to more accurate predictions = forecasts of opportunity

Explainable AI Approach: Learn tropical patterns of variability that lead to enhanced predictability of midlatitude weather on subseasonal timescales



Kirsten Mayer



apply explainable AI methods (e.g. post-hoc attribution approaches) to create a heatmap of relevant regions in the input for the network's prediction

COLORADO STATE UNIVERSITY

Mayer and Barnes (2021)
Bach et al. (2015)

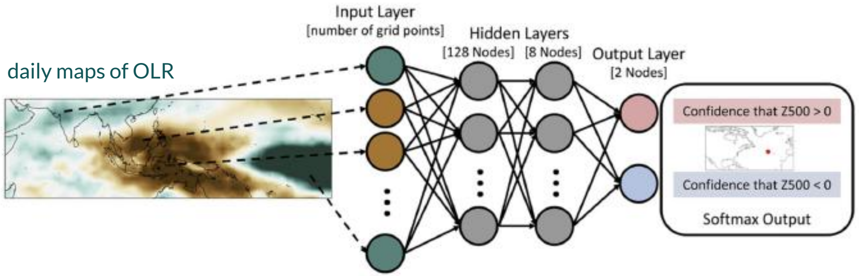
Subseasonal weather prediction

Exploring how tropical information can lend predictability to midlatitude circulation on S2S timescales

Network Task: Train a neural network to ingest daily maps of outgoing longwave radiation (OLR) to predict the sign of the subseasonal circulation anomalies over the North Atlantic 22 days in advance

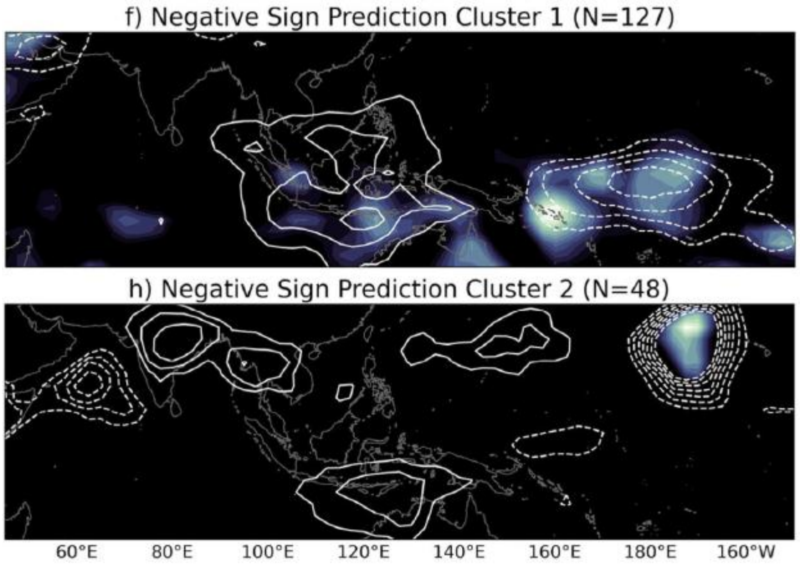
Forecasts of Opportunity: Confident predictions lead to more accurate predictions = forecasts of opportunity

Explainable AI Approach: Learn tropical patterns of variability that lead to enhanced predictability of midlatitude weather on subseasonal timescales



Kirsten Mayer

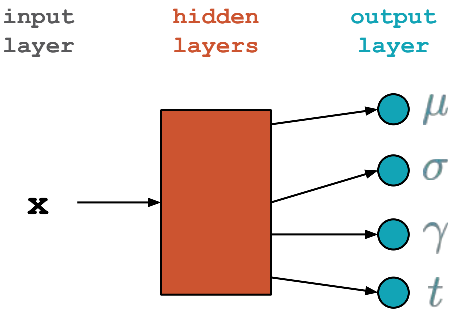
LRP: OLR patterns that lead to accurate & confident Z500<0 predictions over the North Atlantic

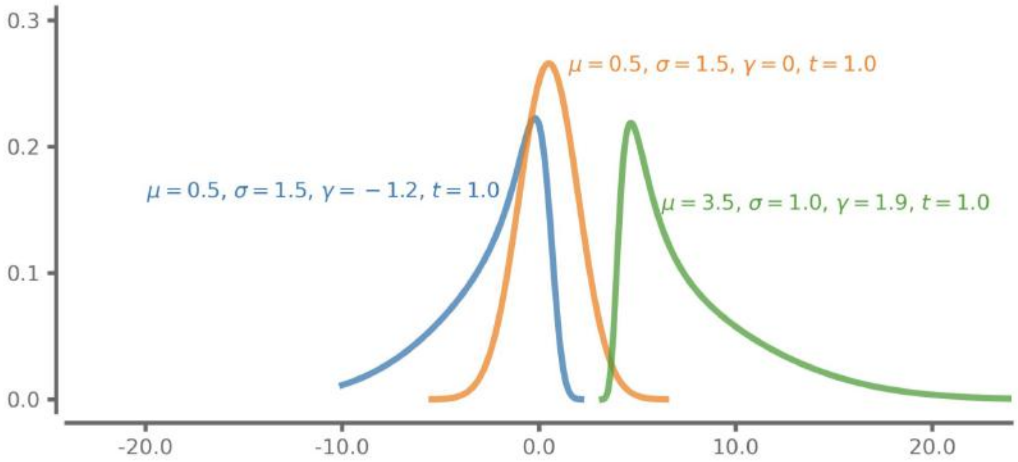


Mayer and Barnes (2021)

What about regression problems?

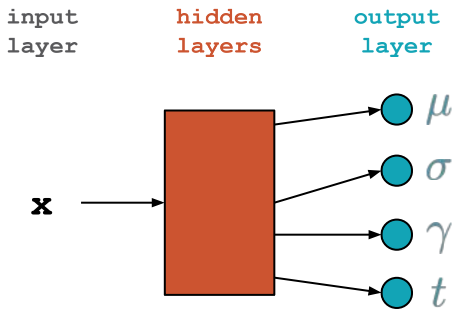
Simple uncertainty for neural network regression tasks

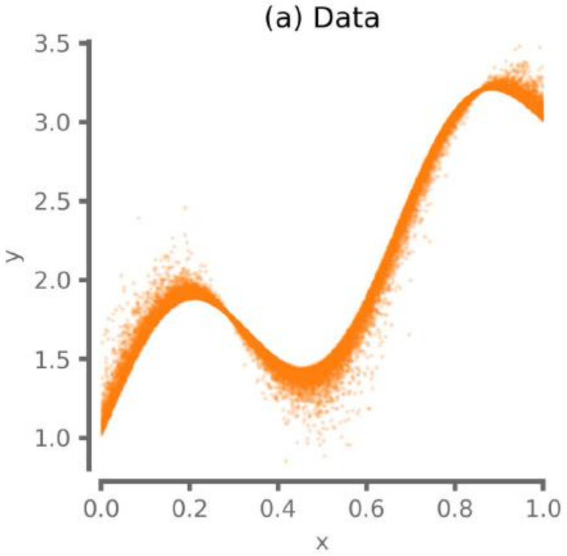




trained with -log(p), based on maximum likelihood estimation
Barnes, Barnes and Gordillo (2021)

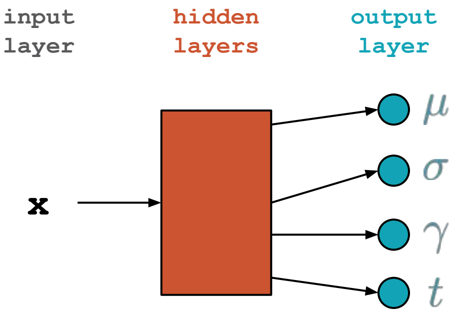
Simple uncertainty for neural network regression tasks

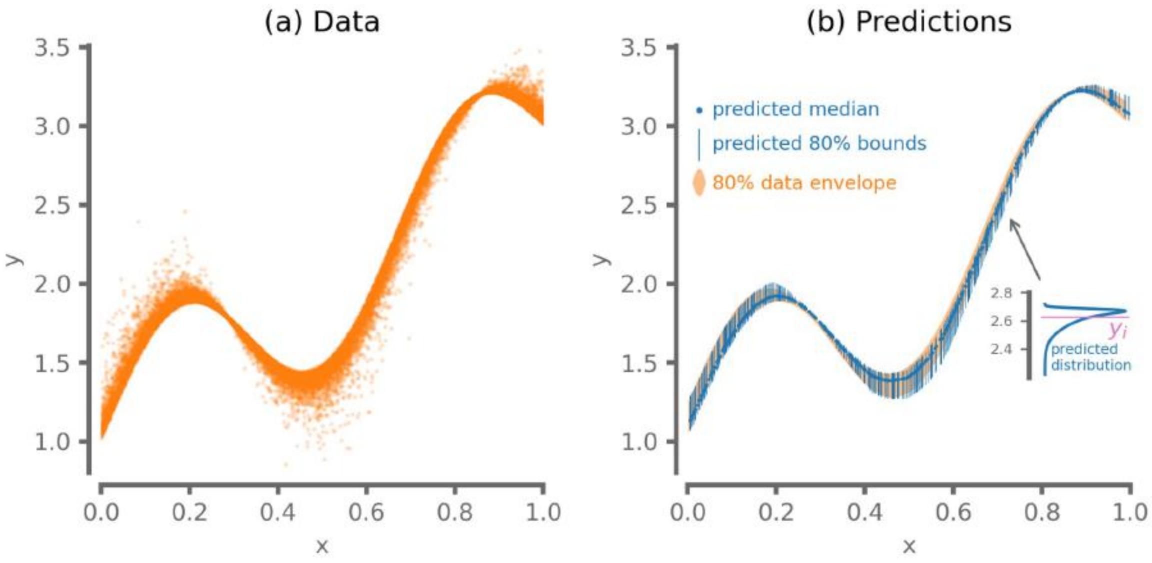




trained with -log(p), based on maximum likelihood estimation
Barnes, Barnes and Gordillo (2021)

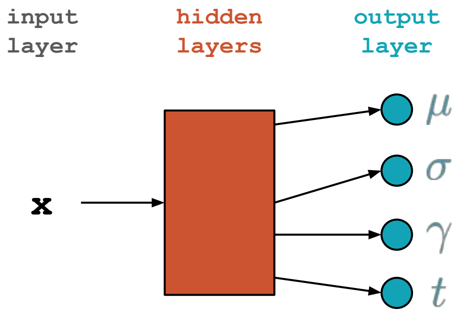
Simple uncertainty for neural network regression tasks

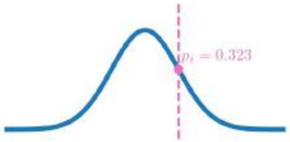




trained with -log(p), based on maximum likelihood estimation
Barnes, Barnes and Gordillo (2021)

The loss function



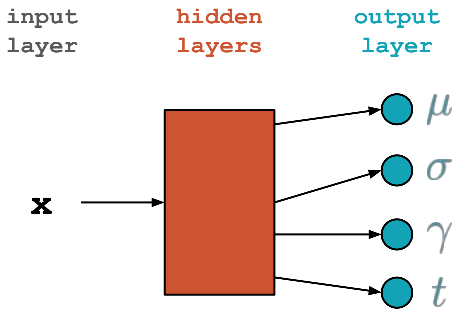


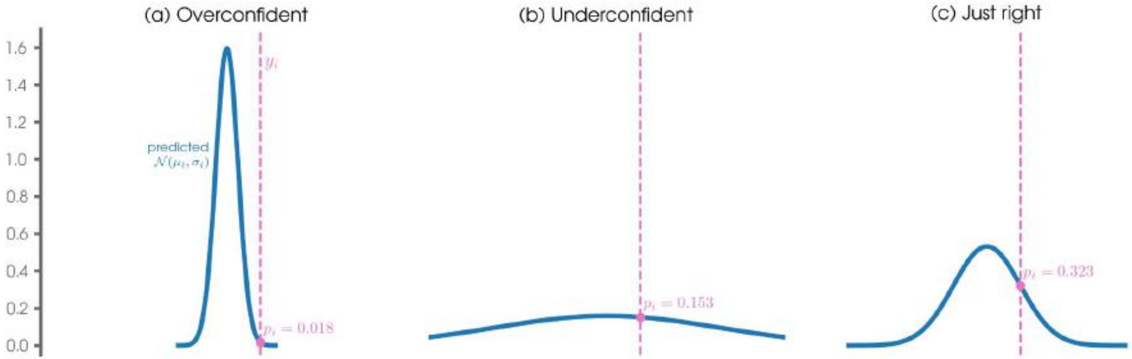
$$p_i = \mathcal{P}(y_i | \mu, \sigma, \gamma, \tau)$$

$$\mathcal{L}(x_i) = -\log p_i$$

trained with -log(p), based on maximum likelihood estimation
Barnes, Barnes and Gordillo (2021)

The loss function





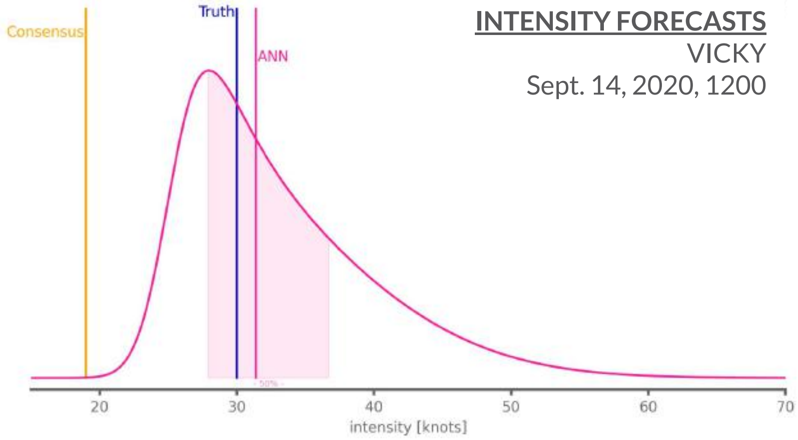
$$p_i = \mathcal{P}(y_i | \mu, \sigma, \gamma, \tau)$$

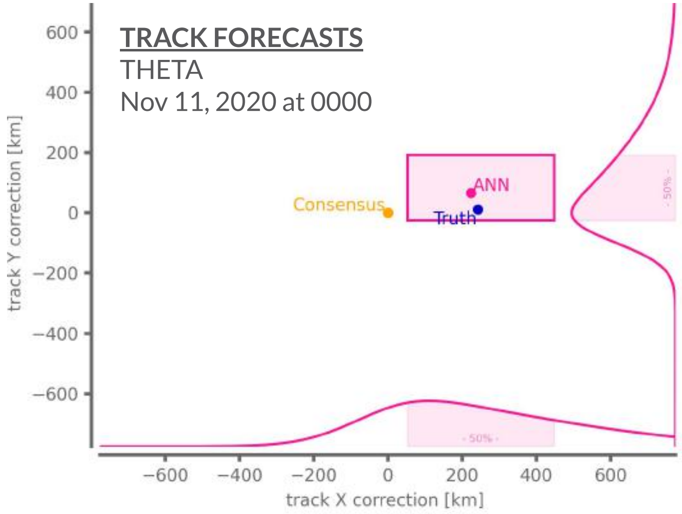
$$\mathcal{L}(x_i) = -\log p_i$$

trained with -log(p), based on maximum likelihood estimation
Barnes, Barnes and Gordillo (2021)

Operational hurricane forecasts

- **Network Task:** Train neural networks to predict error of the “Consensus” forecast of physics-based models (used by the National Hurricane Center)
- **Predicted PDF:** Allows us to update the forecast as well as understand the uncertainty of the ANN



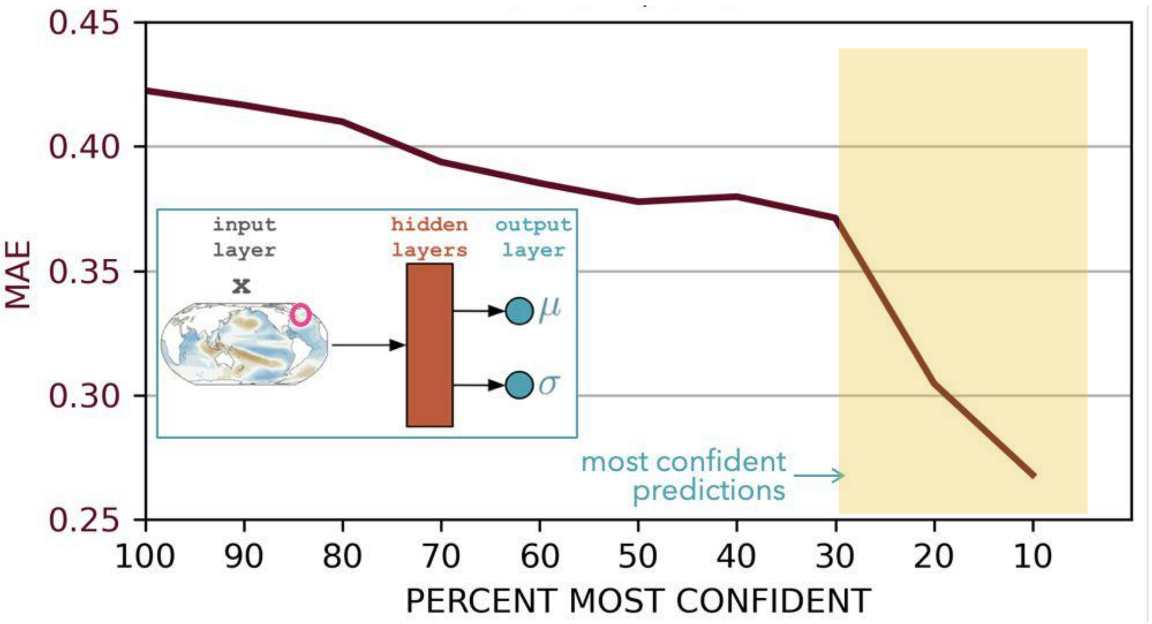


Barnes, Barnes and DeMaria (in prep)

Decadal Prediction

- **Network Task:** Train a neural network to ingest maps of Ocean Heat Content and predict future sea-surface temperatures averaged over the next 1-5 years
- **State-Dependent Predictability:** Confident predictions lead to more smaller errors
- **Explainable AI Approach:** Learn decadal patterns of variability that lead to predictability

Emily Gordon



Gordon and Barnes (in prep)

Controlled Abstention Networks (CAN)

The abstention loss works by incorporating uncertainty in the network's prediction to identify the more confident samples and abstain (say "I don't know") on the less confident samples.

...the abstention loss is applied *during training* to preferentially learn from the more confident samples.

Our work has heavily informed by the dissertation of Dr. Sunil Thulasidasan (2020):

- https://arxiv.org/abs/1905.10964
- https://github.com/thulas/dac-label-noise/blob/master/dac_loss.py
- https://digital.lib.washington.edu/researchworks/handle/1773/45781

General Idea of CANs

1. Estimate uncertainty of each prediction during training
    - **Classification:** simple - just use the softmax output
    - **Regression:** we need a way to predict uncertainty [more later]

2. Implement a loss function that learns to identify more confident predictions and learn them better
    - **Classification:** we introduce the *NotWrong Loss*
    - **Regression:** we introduce a modified *negative log likelihood*

3. Compare to baseline methods that filter out samples post training
    - While the baseline methods perform very well, we find that the **abstention method outperforms the baseline** for a variety of tasks

Our work has heavily informed by the dissertation of Dr. Sunil Thulasidasan (2020):

- https://arxiv.org/abs/1905.10964
- https://github.com/thulas/dac-label-noise/blob/master/dac_loss.py
- https://digital.lib.washington.edu/researchworks/handle/1773/45781

General Idea of CANs

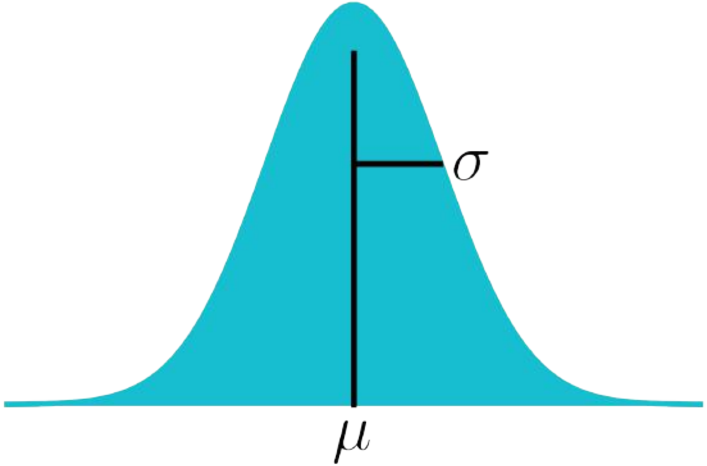
1. Estimate uncertainty of each prediction during training
    - **Classification:** simple - just use the softmax output
    - **Regression:** we need a way to predict uncertainty [more later]

2. Implement a loss function that learns to identify more confident predictions and learn them better
    - **Classification:** we introduce the *NotWrong Loss*
    - **Regression:** we introduce a modified *negative log likelihood*

3. Compare to baseline methods that filter out samples post training
    - While the baseline methods perform very well, we find that the **abstention method outperforms the baseline** for a variety of tasks

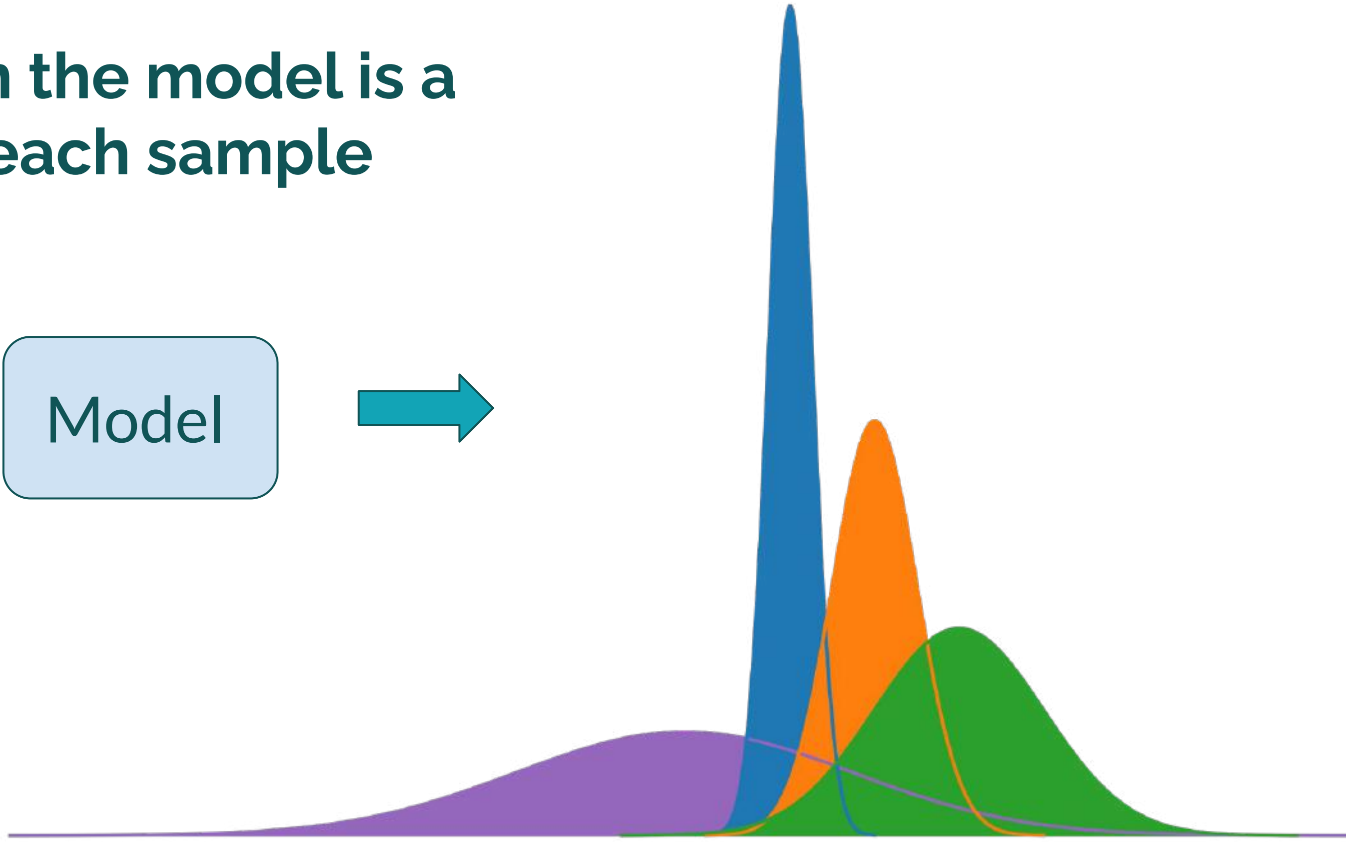
The output from the model is a distribution for each sample

Sample → Model →

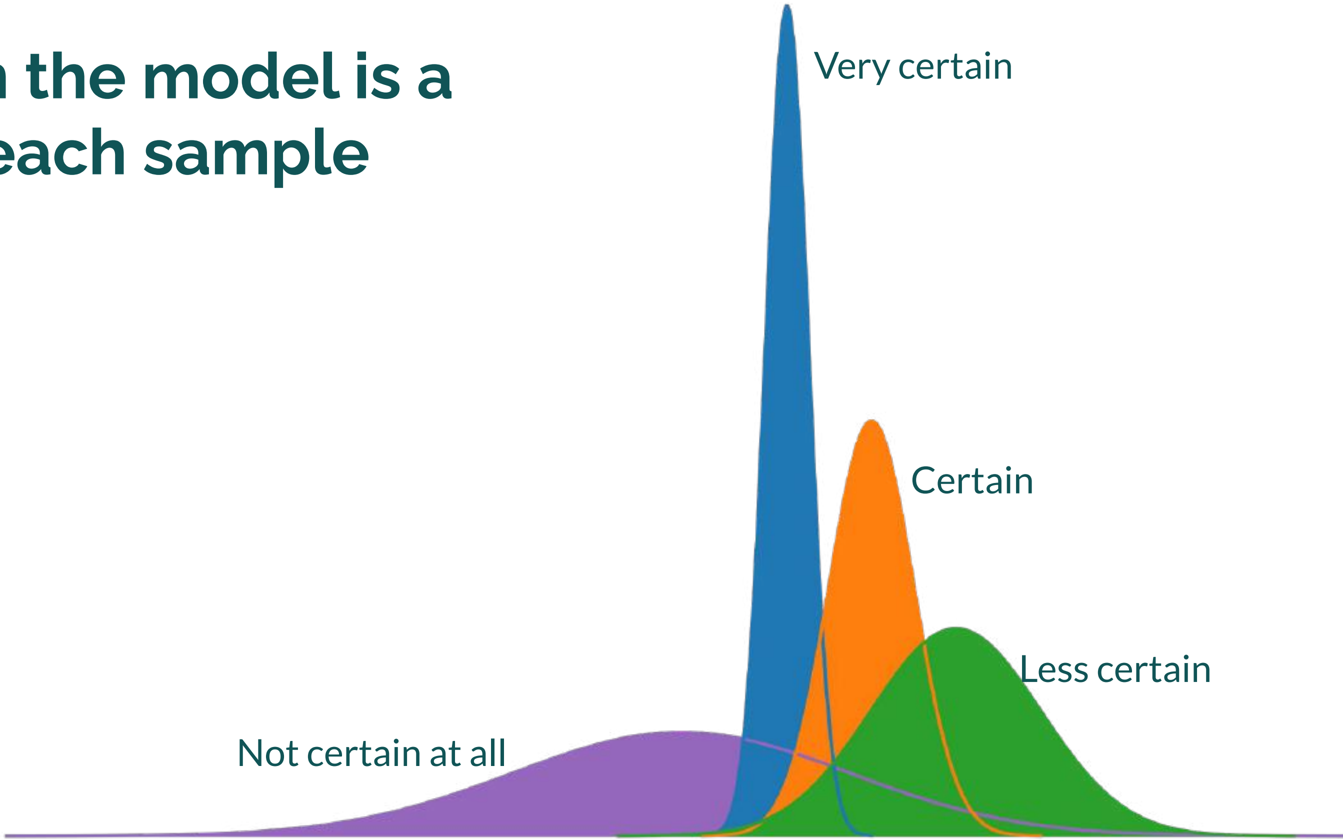


The output from the model is a distribution for each sample

Many Samples → Model →



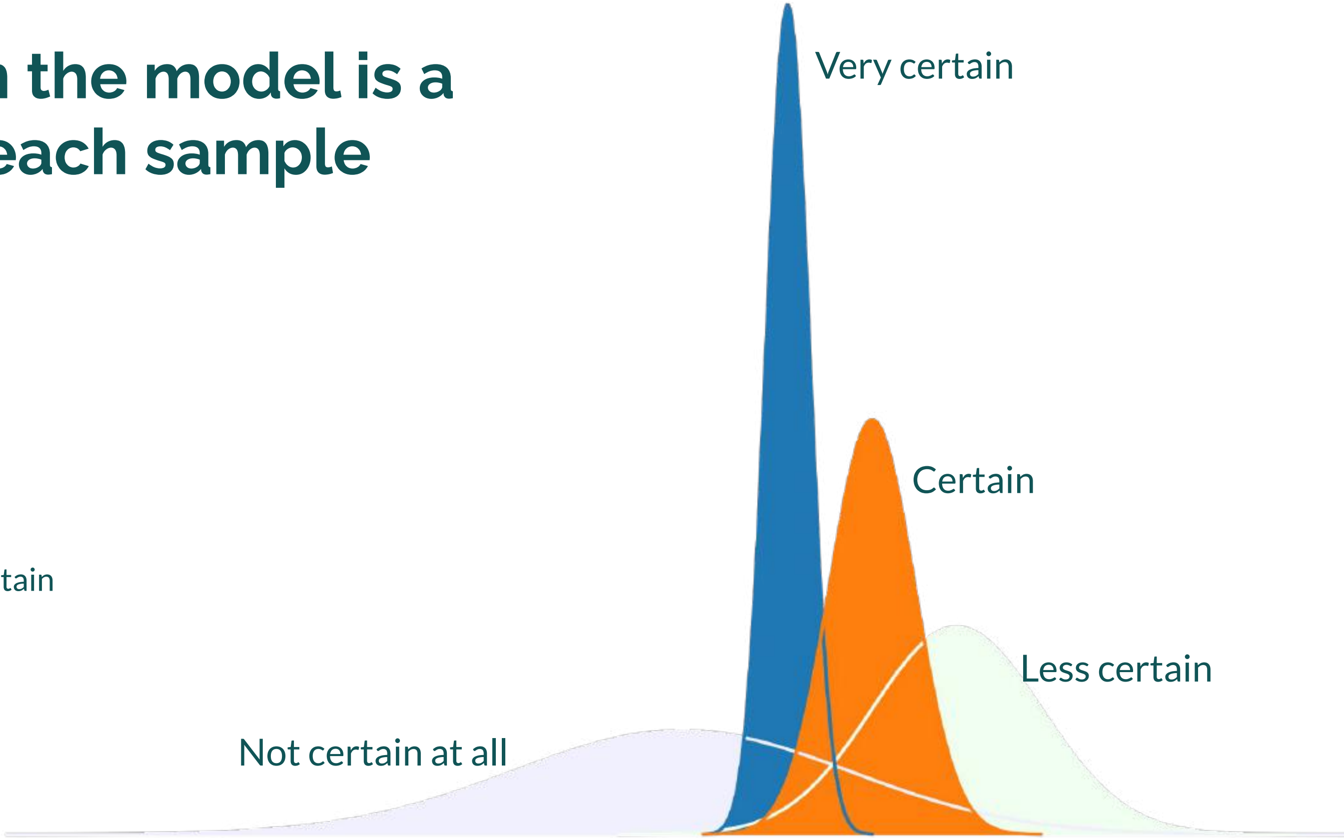
The output from the model is a distribution for each sample



The output from the model is a distribution for each sample

Baseline:

- Train the model
- Sort the results by sigma
- Keep only the x% most certain distributions.



Now for abstention during neural network training

Abstention During Training

- Abstention loss is very similar for both classification and regression
- The abstention regression loss is a modified log loss, weighted by the "prediction weight" determined by the uncertainty sigma
- An additional term penalizes abstention

3.2.1 *Abstention loss*

Unlike the baseline ANN, the CAN loss is designed to identify the less confident predictions so as to preferentially learn from the more confident predictions. The CAN loss for sample x_i is defined as

$$\mathcal{L}(x_i) = -q_i \log p_i - \alpha \log q_i. \tag{4}$$

where α controls the amount of abstention (see next subsection) and q_i represents the prediction weight defined as

$$q_i = \min\left(1.0, \left[\frac{\kappa}{\sigma_i}\right]^2\right). \tag{5}$$

Barnes and Barnes (2021)

Abstention During Training

- Abstention loss is very similar for both classification and regression
- The abstention regression loss is a modified log loss, weighted by the "prediction weight" determined by the uncertainty sigma
- An additional term penalizes abstention

3.2.1 Abstention loss

Unlike the baseline ANN, the CAN loss is designed to identify the less confident predictions so as to preferentially learn from the more confident predictions. The CAN loss for sample x_i is defined as

baseline -log(p)

controls amount of abstention

$$\mathcal{L}(x_i) = -q_i \log p_i - \alpha \log q_i. \tag{4}$$

where α controls the amount of abstention (see next subsection) and q_i represents the prediction weight defined as

$$q_i = \min\left(1.0, \left[\frac{\kappa}{\sigma_i}\right]^2\right). \tag{5}$$

prediction weight

data-specific scale

Barnes and Barnes (2021)

Abstention During Training

- Abstention loss is very similar for both classification and regression
- The abstention regression loss is a modified log loss, weighted by the “prediction weight” determined by the uncertainty sigma
- An additional term penalizes abstention
- **alpha:** abstention fraction can be set by a PID controller or user can have network predict the best abstention fraction

3.2.1 Abstention loss

Unlike the baseline ANN, the CAN loss is designed to identify the less confident predictions so as to preferentially learn from the more confident predictions. The CAN loss for sample x_i is defined as

baseline -log(p)

controls amount of abstention

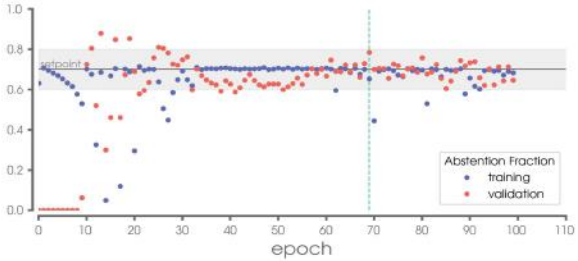
$$\mathcal{L}(x_i) = -q_i \log p_i - \alpha \log q_i. \tag{4}$$

where α controls the amount of abstention (see next subsection) and q_i represents the prediction weight defined as

$$q_i = \min\left(1.0, \left[\frac{\kappa}{\sigma_i}\right]^2\right). \tag{5}$$

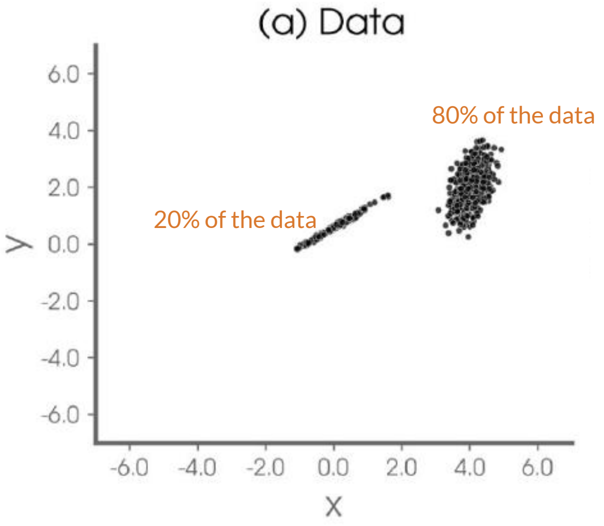
prediction weight

data-specific scale



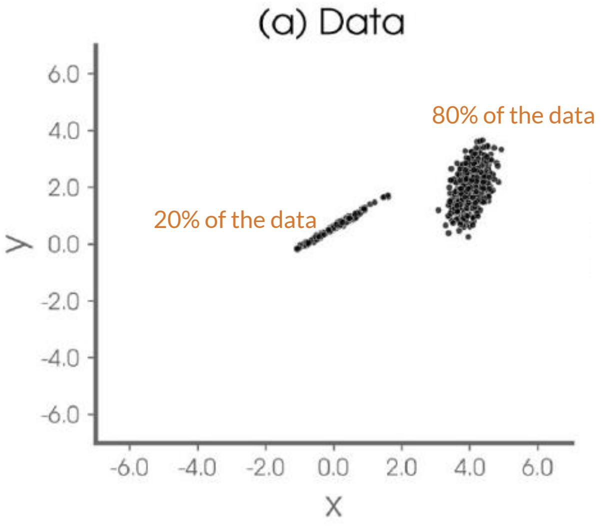
Barnes and Barnes (2021)

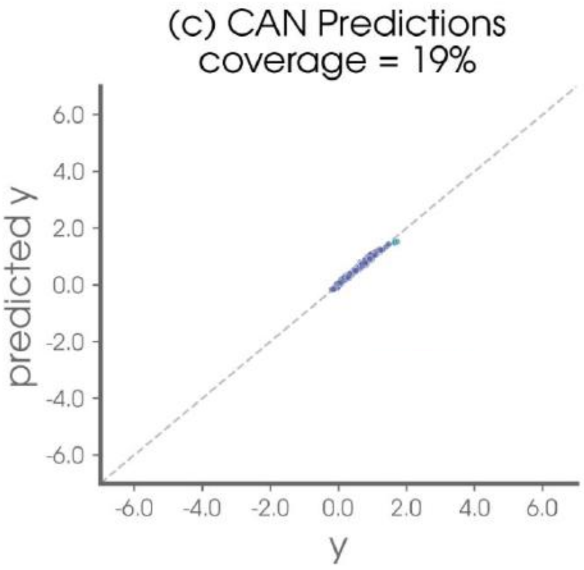
A simple 1D example



Barnes and Barnes (2021)

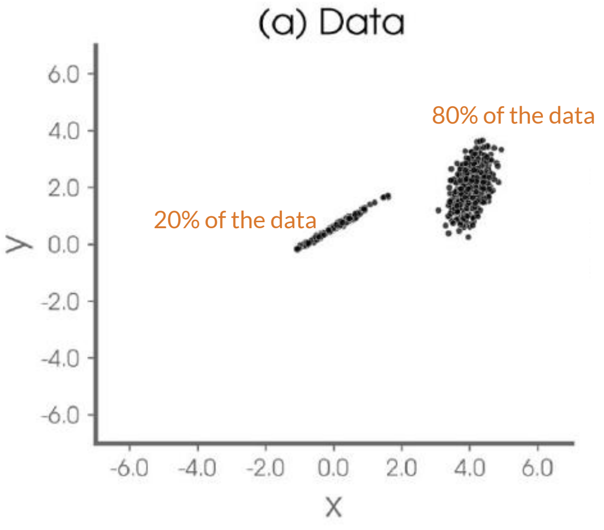
A simple 1D example

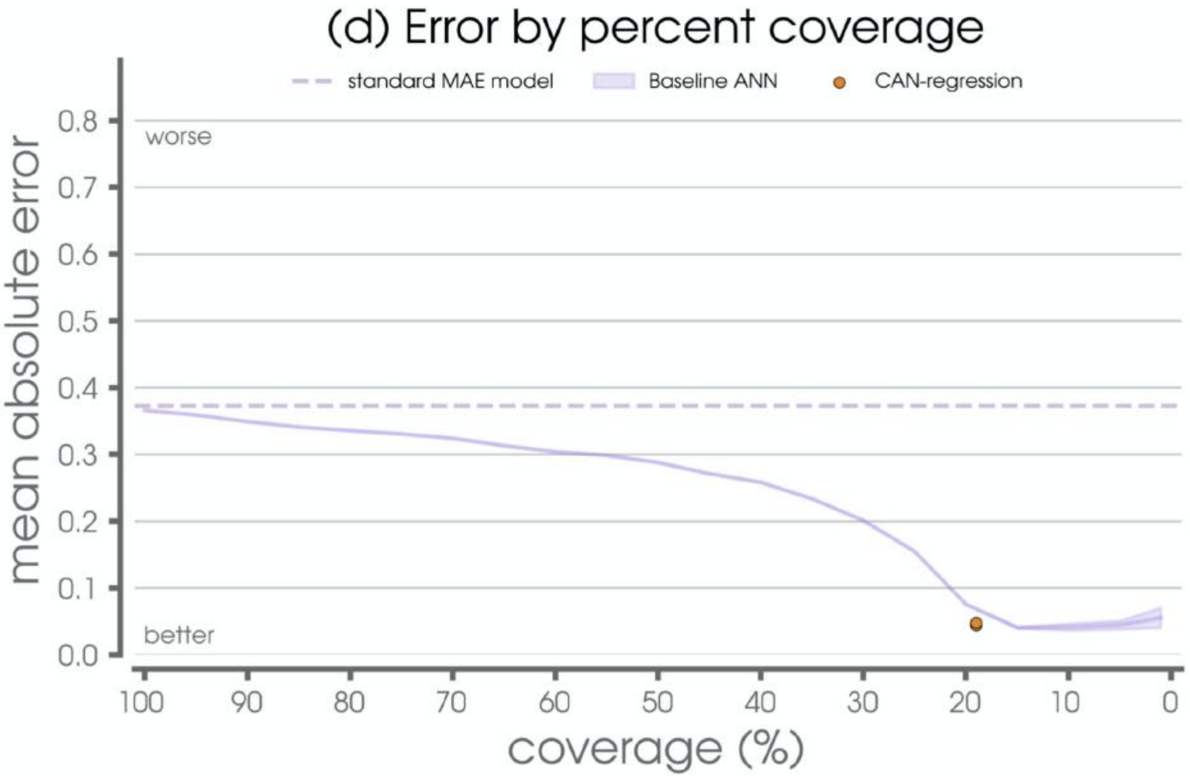




Barnes and Barnes (2021)

A simple 1D example

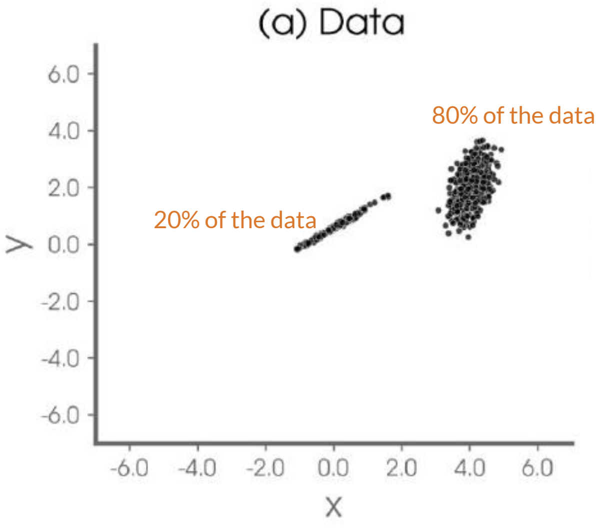


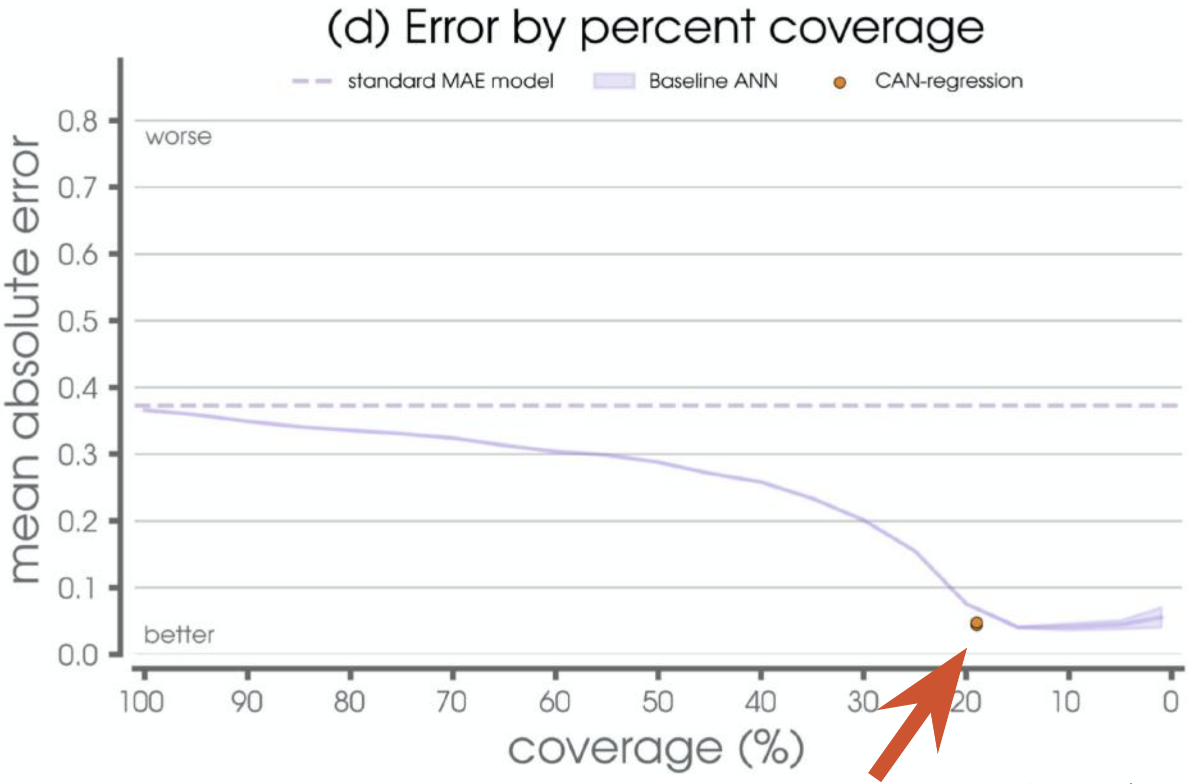


COLORADO STATE UNIVERSITY

Barnes and Barnes (2021)

A simple 1D example





Barnes and Barnes (2021)

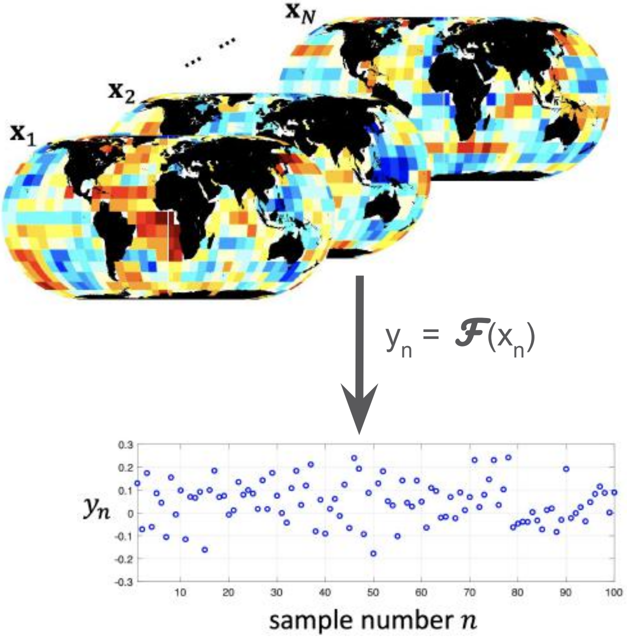
A more complex example

—

Synthetic Climate Data

Dr. Antonios Mamalakis

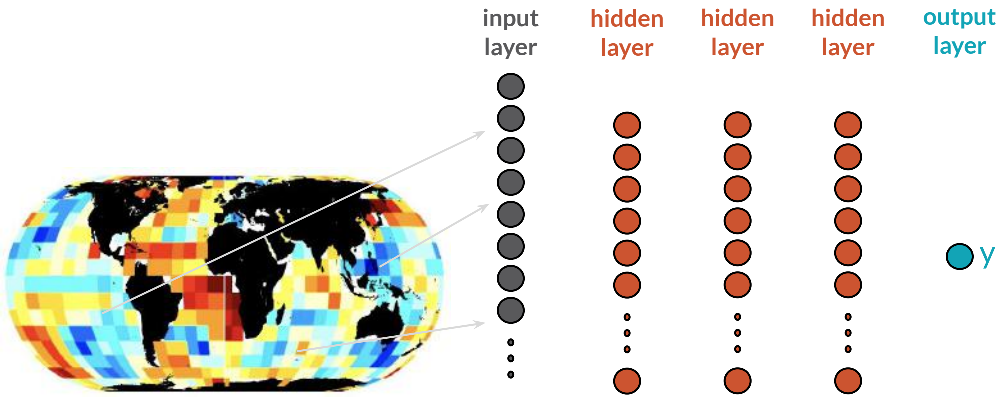
- Created by CSU postdoc Dr. Antonios Mamalakis
- Each sample is one global map of "SSTs" computed from real-world spatial covariances
- Use a known nonlinear function $\mathcal{F}$ to map each map x_n to a scalar y_n



Mamalakis, Ebert-Uphoff and Barnes (2021)

Synthetic Climate Data

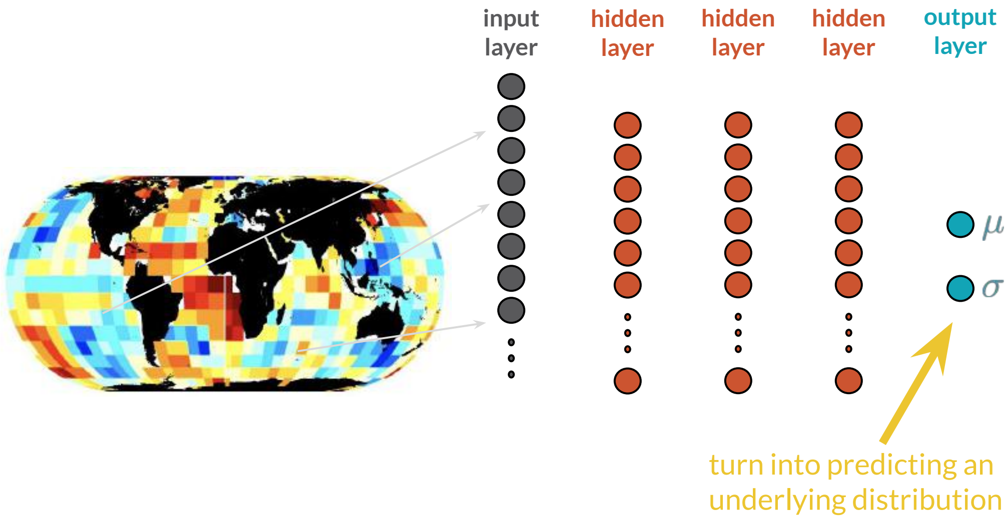
- Created by CSU postdoc Dr. Antonios Mamalakis
- Each sample is one global map of “SSTs” computed from real-world spatial covariances
- Use a known nonlinear function $\mathcal{F}$ to map each map x_n to a scalar y_n
- **Network Task: predict the value “y” for each input map**



Barnes and Barnes (2021)

Synthetic Climate Data

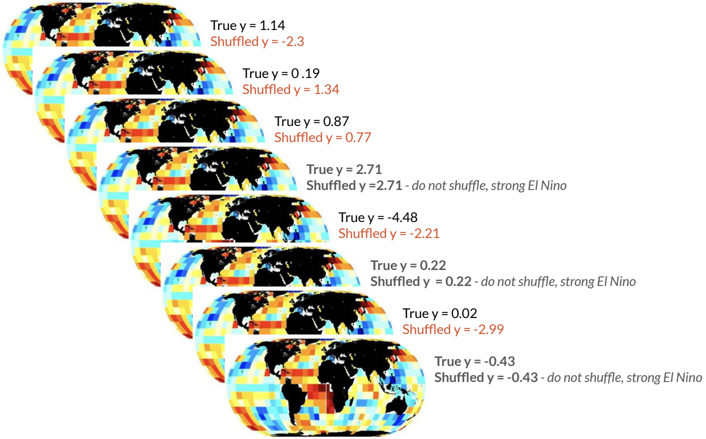
- Created by CSU postdoc Dr. Antonios Mamalakis
- Each sample is one global map of "SSTs" computed from real-world spatial covariances
- Use a known nonlinear function $\mathcal{F}$ to map each map x_n to a scalar y_n
- **Network Task: predict the value "y" for each input map**



Barnes and Barnes (2021)

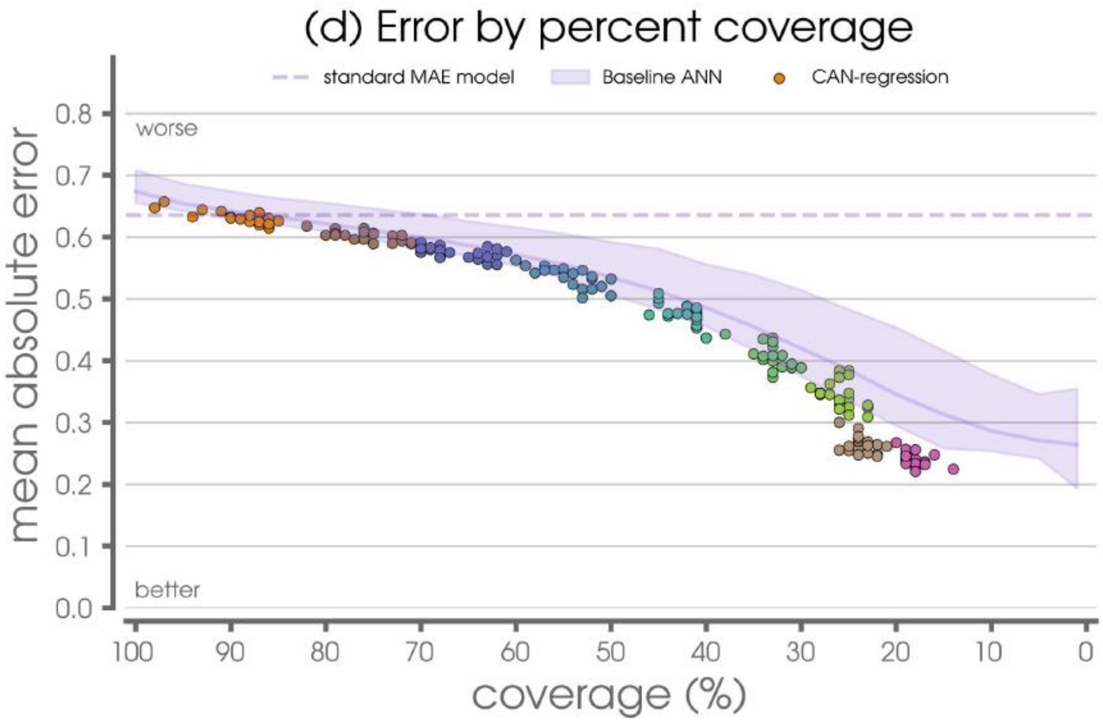
Forecasts of Opportunity Experiment

- All ENSO+ samples (average ENSO region > 0.5) are untouched
- 100% of the other samples are corrupted (shuffled)
- 29% untouched
- 71% corrupt
- **Only samples with strong El Nino signals have a learnable relationship with their labels**



Abstention outperforms baseline

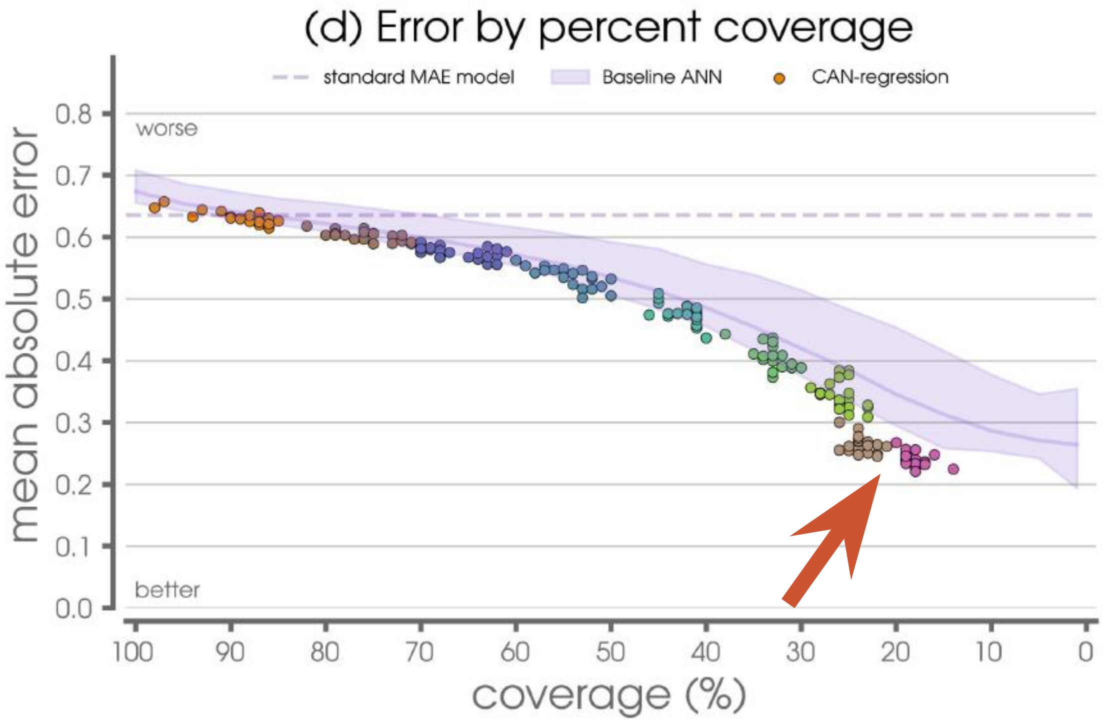
- Train abstention network for different abstention setpoints
- The best CAN models are always better (lower error) than the best baseline ANN



Barnes and Barnes (2021)

Abstention outperforms baseline

- Train abstention network for different abstention setpoints
- The best CAN models are always better (lower error) than the best baseline ANN



COLORADO STATE UNIVERSITY

Barnes and Barnes (2021)

Abstention outperforms baseline

- Train abstention network for different abstention setpoints
- The best CAN models are always better (lower error) than the best baseline ANN

CAN outperforms baseline networks

Corrupted inputs
(input data cleaner)

Shuffled sample labels
(arbitrary label noise)

Forecasts of opportunity
(skillful predictions)

Specific labels corrupted
(structured noise)

Regression tasks: Barnes and Barnes (2021)
Classification tasks: Barnes and Barnes (2021b)

Forecasts of Opportunity

*aka "State-Dependent Predictability"

Leveraging "Forecasts of Opportunity"

1. This is more than just uncertainty quantification and more than just a post-processing application.
   it is worthwhile for anyone working on AI for climate science to consider taking this mindset
2. Impossible predictions may be hampering learning of predictable behaviour
   e.g. predicting climate variables, predicting dynamical forecast errors, etc.
   could this be helpful in filtering out the "harder" predictions to train them separately?
3. May support hybrid approach to climate model parameterizations
   e.g. use uncertainty measures or abstention to kick predictions to the ML or physics-based parameterizations in real-time; could this be helpful for out-of-sample climate change?
4. Utility of this concept revolves around the fact that we have a "small" amount of data to train on
   if we had lots and lots of data, presumably the ML could figure out what to ignore and what to use?

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

A few reference links

- Mamalakis, Antonios, Imme Ebert-Uphoff and Elizabeth A. Barnes: Neural Network Attribution Methods for Problems in Geoscience: A Novel Synthetic Benchmark Dataset, submitted to IEEE Transactions on Neural Networks and Learning Systems, 03/2021, preprint available https://arxiv.org/abs/2103.10005.
- Barnes, Elizabeth A. and Randal J. Barnes: Controlled abstention neural networks for identifying skillful predictions for regression problems, submitted to JAMES, 04/2021, preprint available at https://arxiv.org/abs/2104.08236
  - https://github.com/eabarnes1010/controlled_abstention_networks
- Barnes, Elizabeth A. and Randal J. Barnes: Controlled abstention neural networks for identifying skillful predictions for classification problems, submitted to JAMES, 04/2021, preprint available at https://arxiv.org/abs/2104.08281
  - https://github.com/eabarnes1010/controlled_abstention_networks
- Barnes, Elizabeth A., Randal J. Barnes and Nicolas Gordillo: Adding Uncertainty to Neural Network Regression Tasks in the Geosciences, 2021: https://arxiv.org/abs/2109.07250
- Thulasidasan, Sunil. 2020. "Deep Learning with Abstention: Algorithms for Robust Training and Predictive Uncertainty." https://digital.lib.washington.edu/researchworks/handle/1773/45781.
- Thulasidasan, Sunil, Tanmoy Bhattacharya, Jeff Bilmes, Gopinath Chennupati, and Jamal Mohd-Yusof. 2019. "Combating Label Noise in Deep Learning Using Abstention." arXiv [stat.ML]. arXiv. http://arxiv.org/abs/1905.10964.
  - https://github.com/thulas/dac-label-noise/blob/master/dac_loss.py

Extra Slides