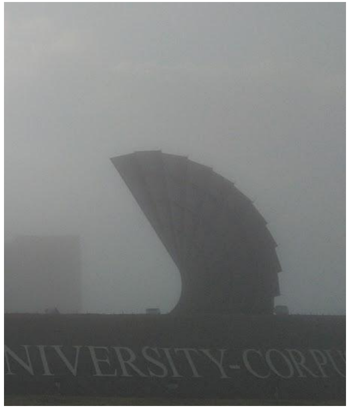
SCOTT: The Influence of Feature Group Schemes on Explainable AI for Geoscience AI Models

August 18, 2023
Evan Krell

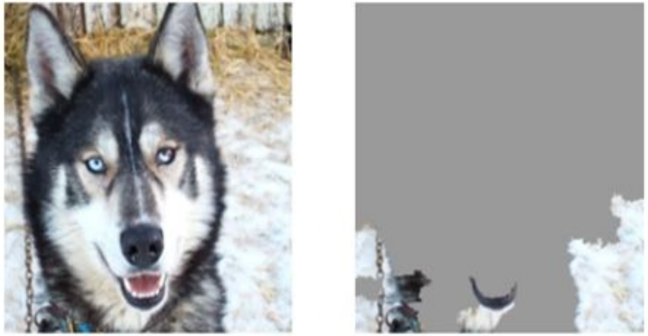


Synopsis: Gridded spatial data can be used to develop high performance machine learning models, but their complexity makes it hard to verify that the model learned realistic strategies. Explainable AI (XAI) techniques can be used to investigate models, but they struggle with correlated features. A proposed solution is to group correlated features for XAI. We use FogNet, a deep learning model for coastal fog prediction, to explore XAI grouping schemes. We demonstrate that using a hierarchy of feature groups can be used to gain insights into the scale of the learned features.

Bio: Evan Krell is a Ph.D. student in the GSCS program and a member of the Innovation in COmputing REsearch lab (iCORE) as well as the NSF AI Institute for Research on Trustworthy AI in Weather, Climate, and Coastal Oceanography (AI2ES). He is broadly interested in XAI, geoscience models, data visualization, marine robotics, fishing, and boating. His current project is to learn Chinese cooking and his three-cup chicken (三杯鸡) is way better than the dish at Dao.

Explainable Artificial Intelligence (XAI)

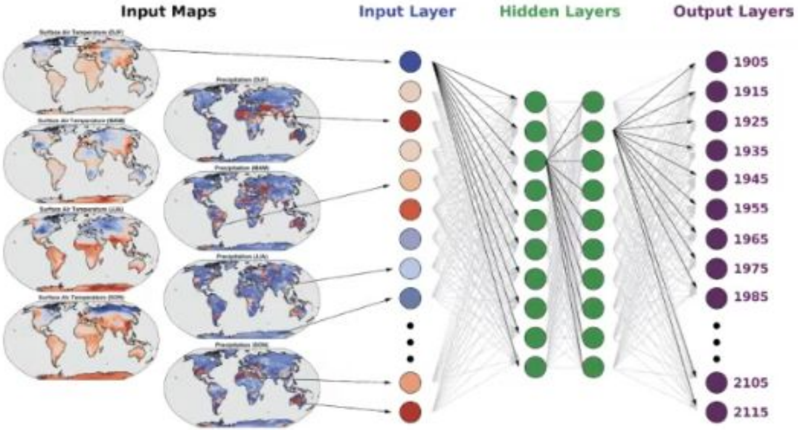
Model verification

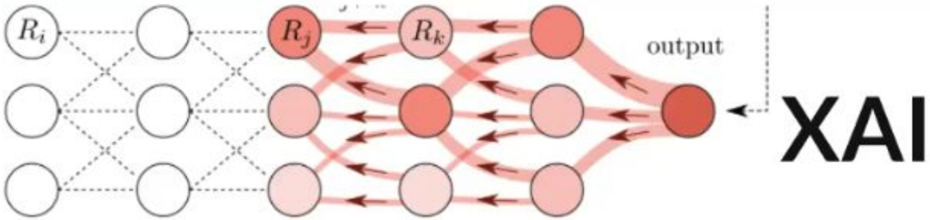


(a) Husky classified as wolf (b) Explanation

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). " Why should i trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1135-1144).

Scientific insights



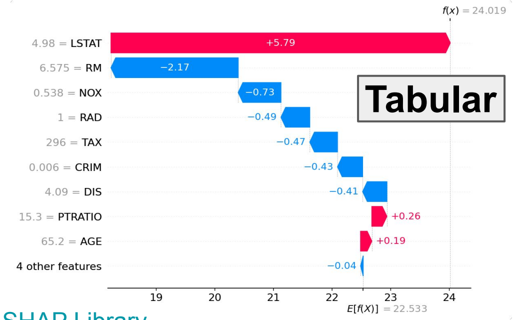


Which regions are ***relevant*** *for correctly predicting the year?*

Presentation: Explainable AI (XAI) for Climate Science: Detection, Prediction and Discovery. Elizabeth Barnes. 2022. https://www.imsi.institute/videos/explainable-ai-xai-for-climate-science-detection-prediction-and-discovery/

Explainable Artificial Intelligence (XAI)

Local Explanation: instance explanation based on a single sample



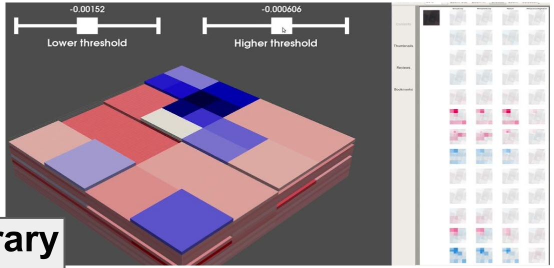
Tabular

[SHAP Library]

RGB

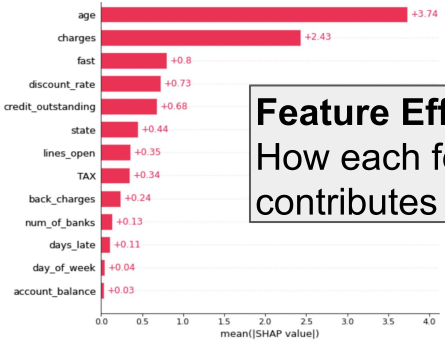
[Gradient-weighted Class Activation Mapping - Grad-CAM- | by Mohamed Chetoui | Medium]

Arbitrary raster



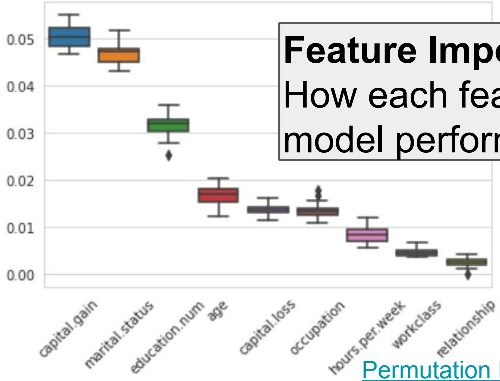
[PartitionShap: viewing multi-channel explanations in 3D]

Global Explanation: summary explanation over a set of samples



Feature Effect:
How each feature contributes to output

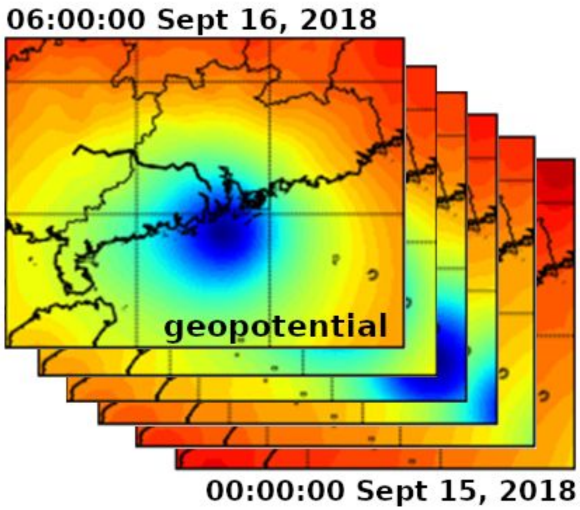
[Feature Importance - Arize AI]

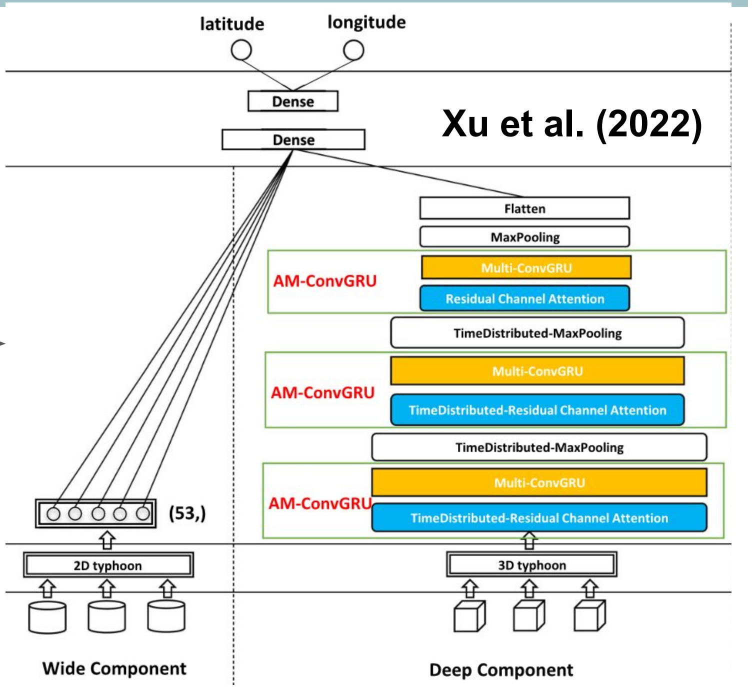


Feature Importance:
How each feature influences model performance

[Permutation Feature Importance]

Geoscience AI Models

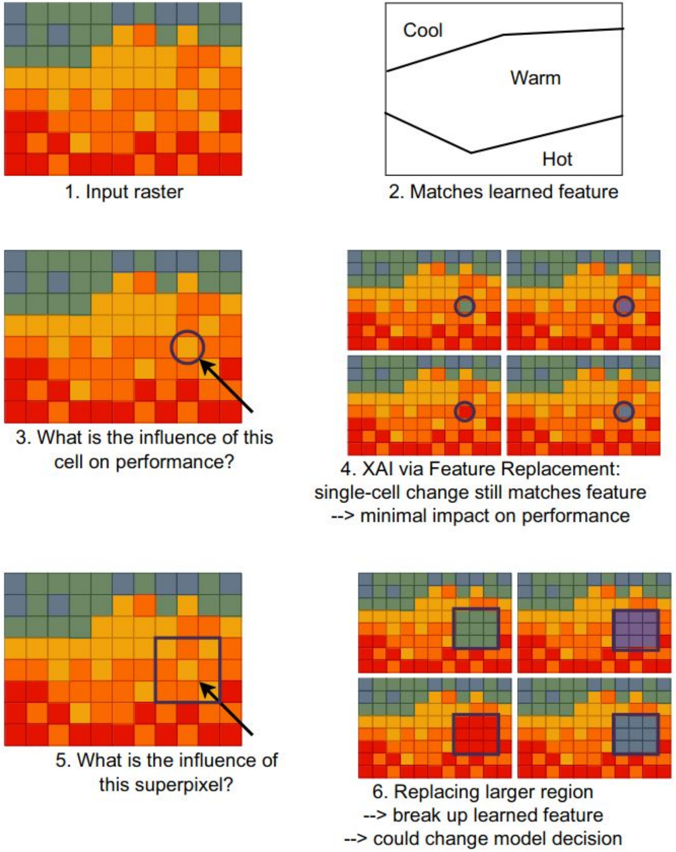




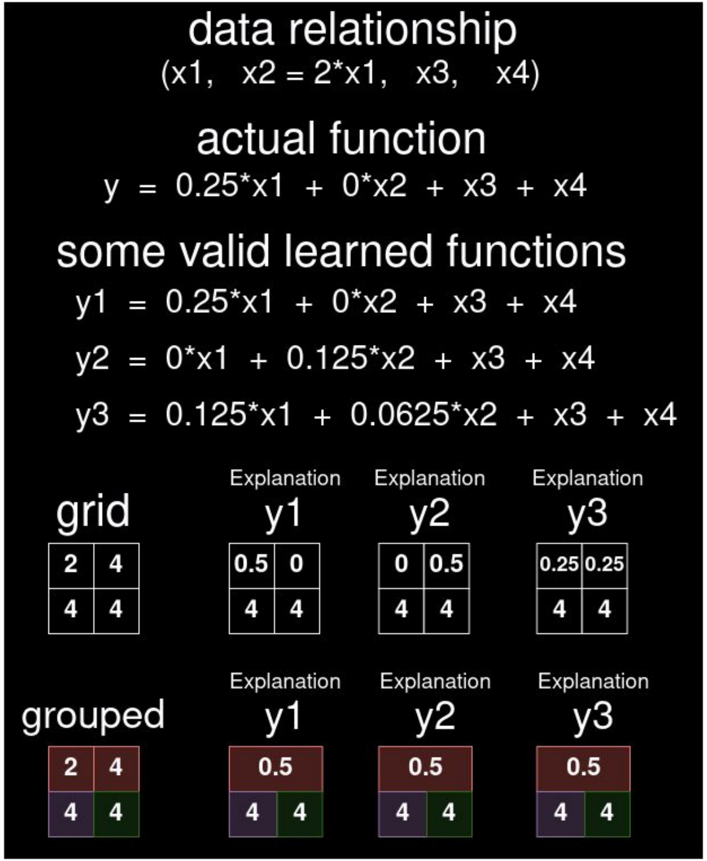
- High-dimensional geospatial raster (gridded) data is used to train complex machine learning models.
- Often complex models (e.g. Deep Neural Net) greatly outperform simpler alternatives (e.g. Random Forest).
- These models are hard to interpret: what are the model's decision-making strategies?

XAI Challenge: Correlated Features

For attribution dilution



For model variance



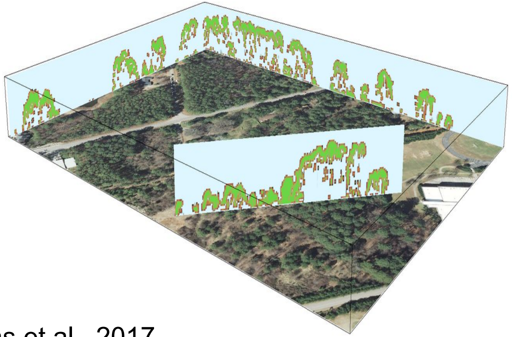
Spatial & Temporal Autocorrelation

Harmful algal bloom

NASA Earth Observatory

2D spatial

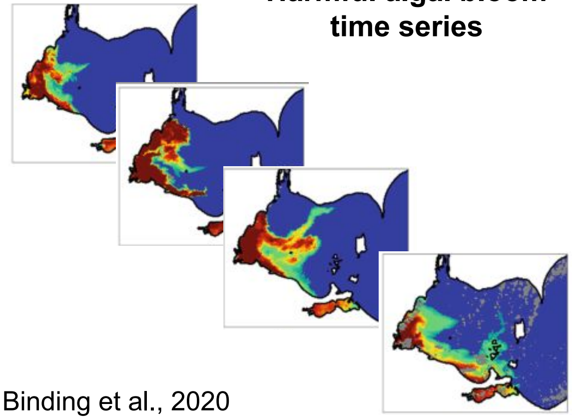
3D vegetation structure derived from lidar point clouds



Petras et al., 2017
https://opengeospatialdata.springeropen.com/articles/10.1186/s40965-017-0021-8

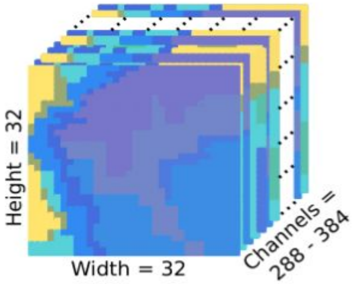
3D spatial

Harmful algal bloom time series



Binding et al., 2020
https://link.springer.com/chapter/10.1007/698_2020_589

3D temporal



FogNet: 4D data (spatio-temporal) packaged as 3D

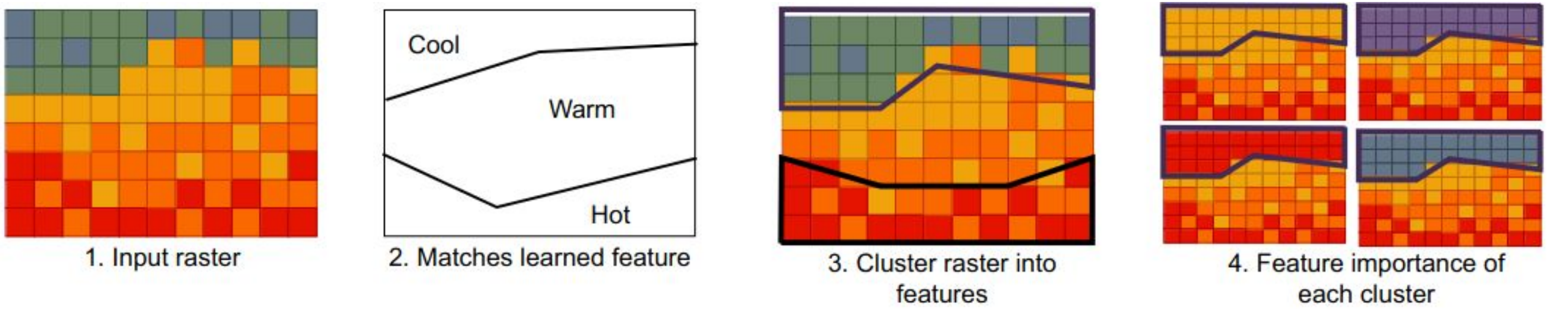
VVel 850mb t0	VVel 850mb t1	VVel 850mb t2	VVel 850mb t3	VVel 875mb t0	…

4 adjacent bands → time sequence

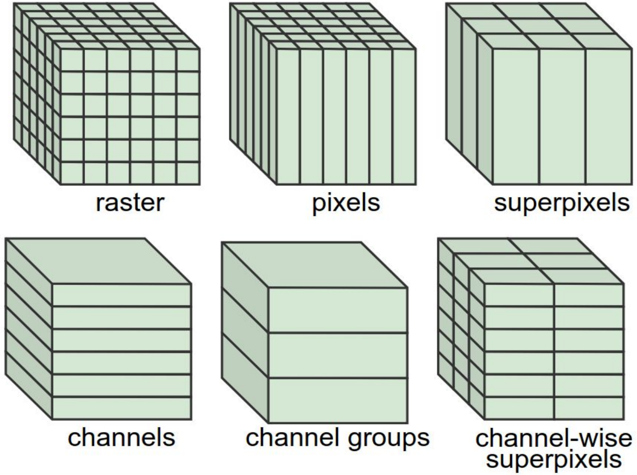
followed by next altitude

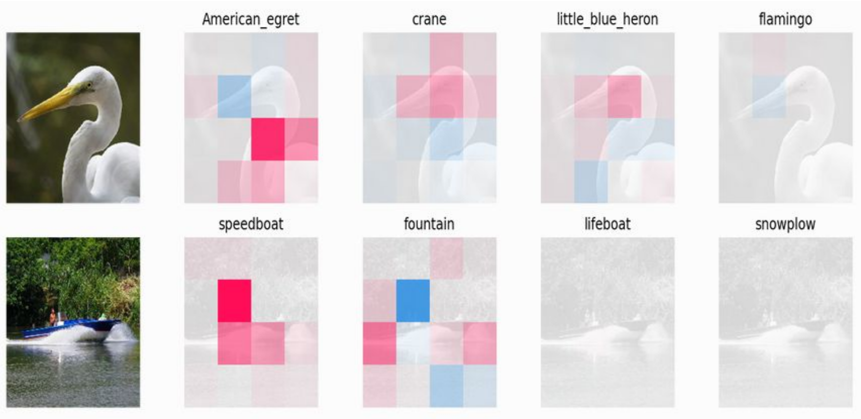
Combining Grid Cells into Feature Groups

Clustering-based



Geometry-based



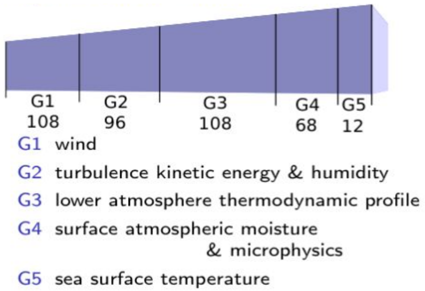


[shap.explainers.Partition — SHAP documentation]

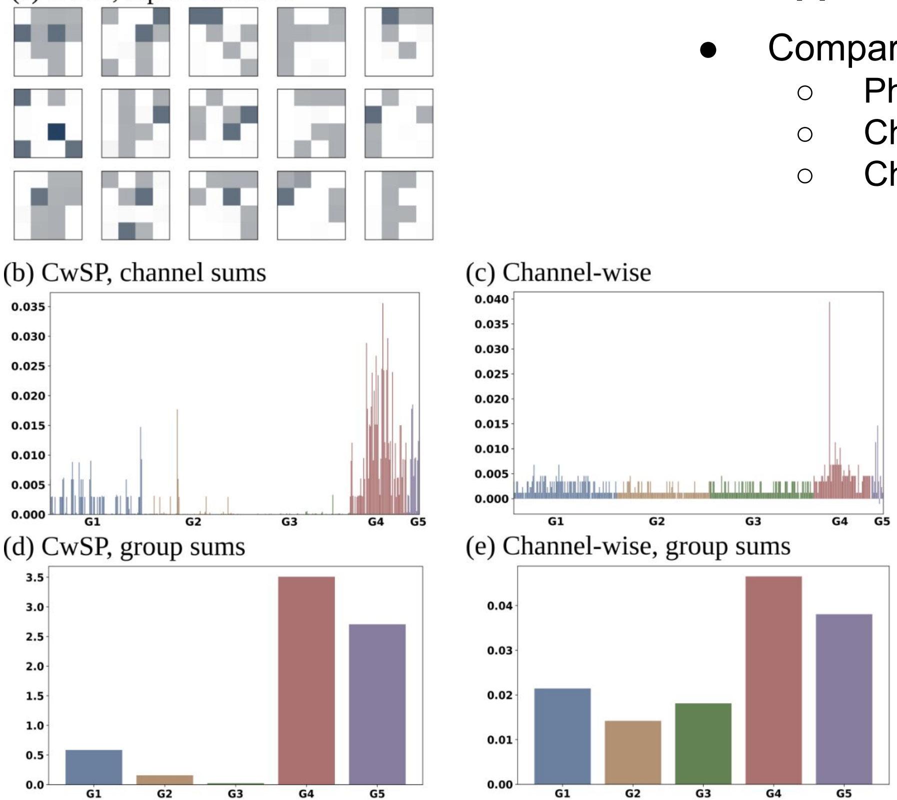
Case Study: FogNet XAI Results

Permutation Feature Importance

- 3D CNN with double-branch dense block & attention mechanism
- **Applied geometric rather than data-driven groupings for XAI**
- Compared 3 grouping schemes:
  - Physics-based channel groups
  - Channel-wise
  - Channel-wise SuperPixels (CwSP)

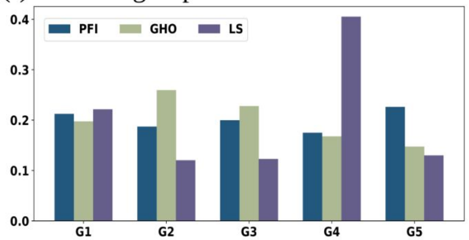


(a) CwSP, top 15 channels
(b) CwSP, channel sums
(c) Channel-wise
(d) CwSP, group sums
(e) Channel-wise, group sums



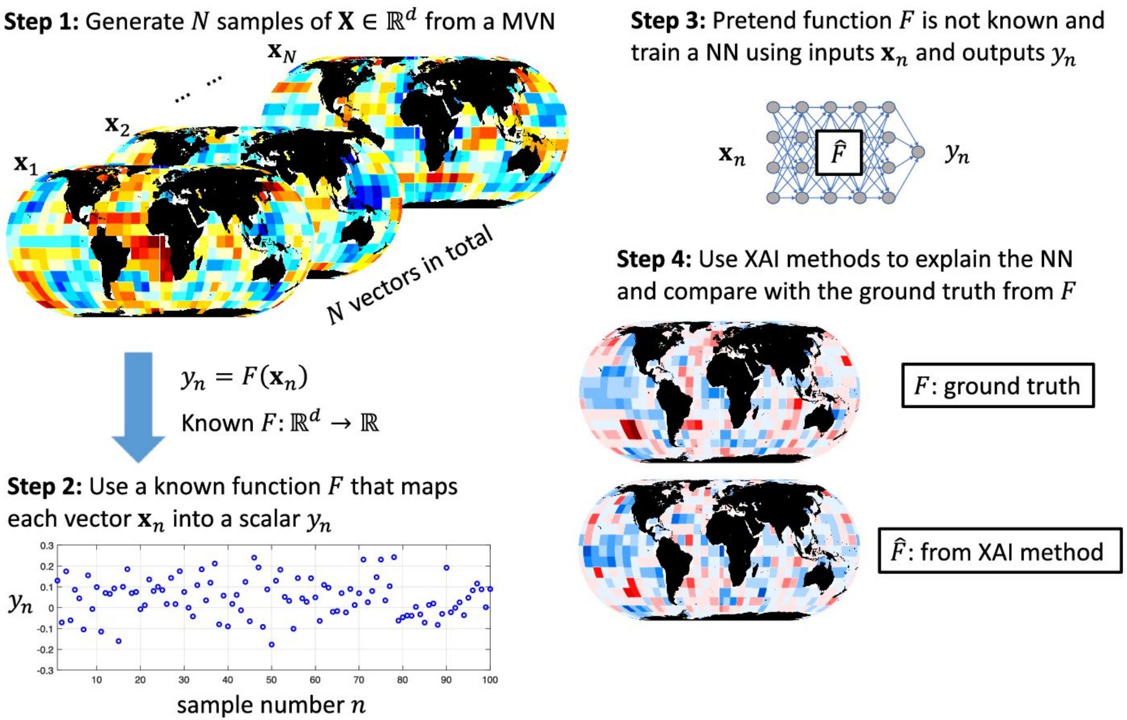
- Groups 1-3 dilute as we increase granularity
- Groups 1-3 contain vertical profiles where small-scale features have little predictive power
- Suggests that FogNet learns 3D features

(f) Channel groups



XAI Verification Benchmarks

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



I am currently building on the XAI verification benchmarks research by Mamalakis et al.

The goal is to build a suite of benchmarks based on various types & strengths of correlation.

We will then use benchmarks to assess methods for data-driven feature groups.

Can clustering strategies be used to improve XAI results?

[Antonios Mamalakis](), [Imme Ebert-Uphoff](), [Elizabeth A. Barnes]()
https://www.cambridge.org/core/journals/environmental-data-science/article/neural-network-attribution-methods-for-problems-in-geoscience-a-novel-synthetic-benchmark-dataset/DDA562FC7B9A2B30710582861920860E