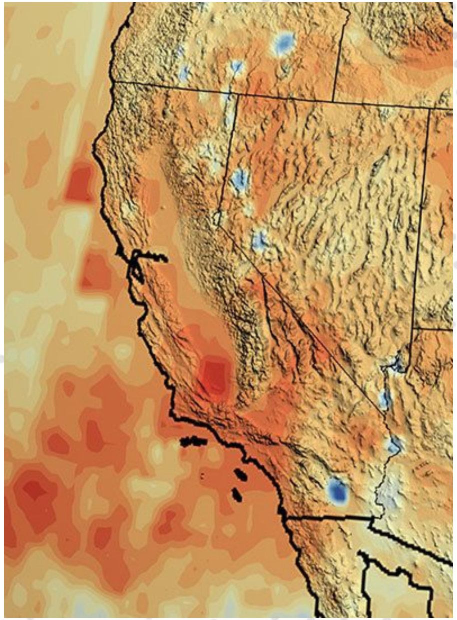
UVA DATA SCIENCE

Predicting Winter California Precipitation with Convolutional Neural Networks



Anthony Chiado, Kristian Olsson, Luke Rohlwing, Michael Vaden

Sponsor and Faculty Mentor: Antonios Mamalakis

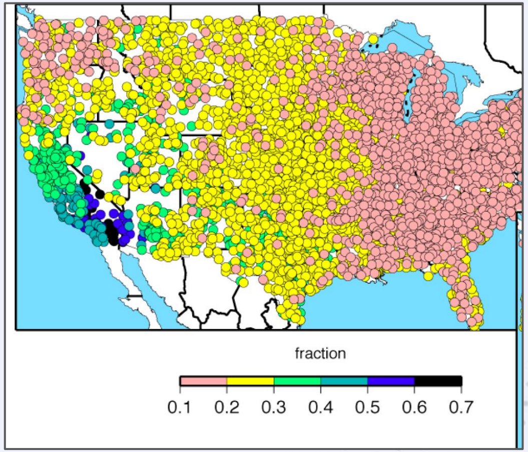
May 1, 2024

Table of Contents

Introduction

Motivation & Goal

- Motivation:
  - Difficult to predict California's hydroclimate
  - Previous literature has used simpler models (linear regression, trees, etc.)
  - Benefit of prediction for policymakers
- Goal:
  - Use deep learning to predict precipitation during winter months using data from summer of the same year

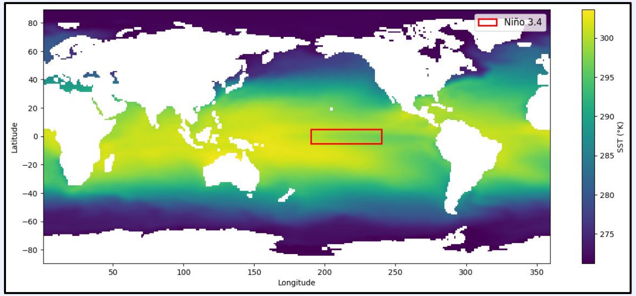


Coefficient of Variation For US Precipitation (Dettinger, et al. 2011)

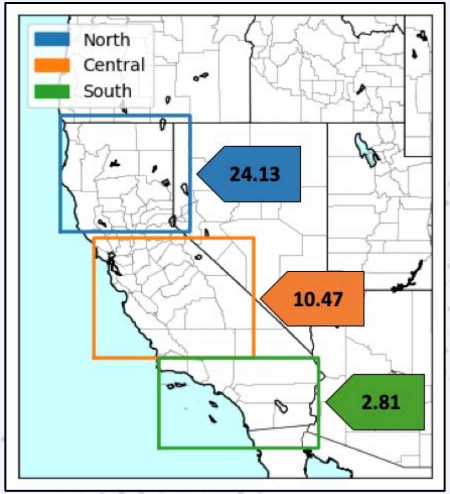
Data

Data Features

- 2D maps of sea surface temperatures
  - Monthly averages from July-October in Kelvin
  - Detrended across time for every lat/lon
  - El Niño subset used for linear model



- Precipitation for each region (North, Central, South)
  - Monthly averages from November-March in mm/day
  - Detrended across time for each region



Data Sources

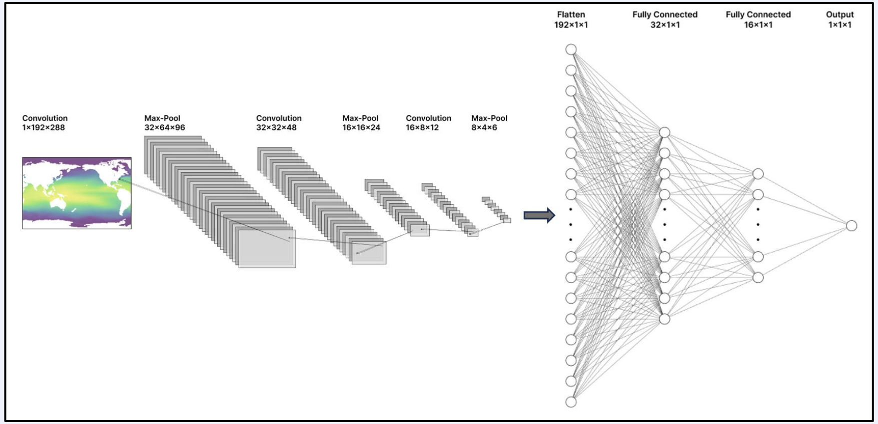
- Simulated data from the Community Earth System Model 2 (CESM2)
    - Assume CESM2 accurately portrays the real climate system
    - 7400 simulated data points from 1940 to 2013 (100 simulations of 74 years)
    - 80 simulations for training, 10 for validating, 10 for testing

- Real world data from
    - 72 data points from 1950 to 2021
    - 40 years for model fine-tuning, 32 years for testing

Methods

Methods

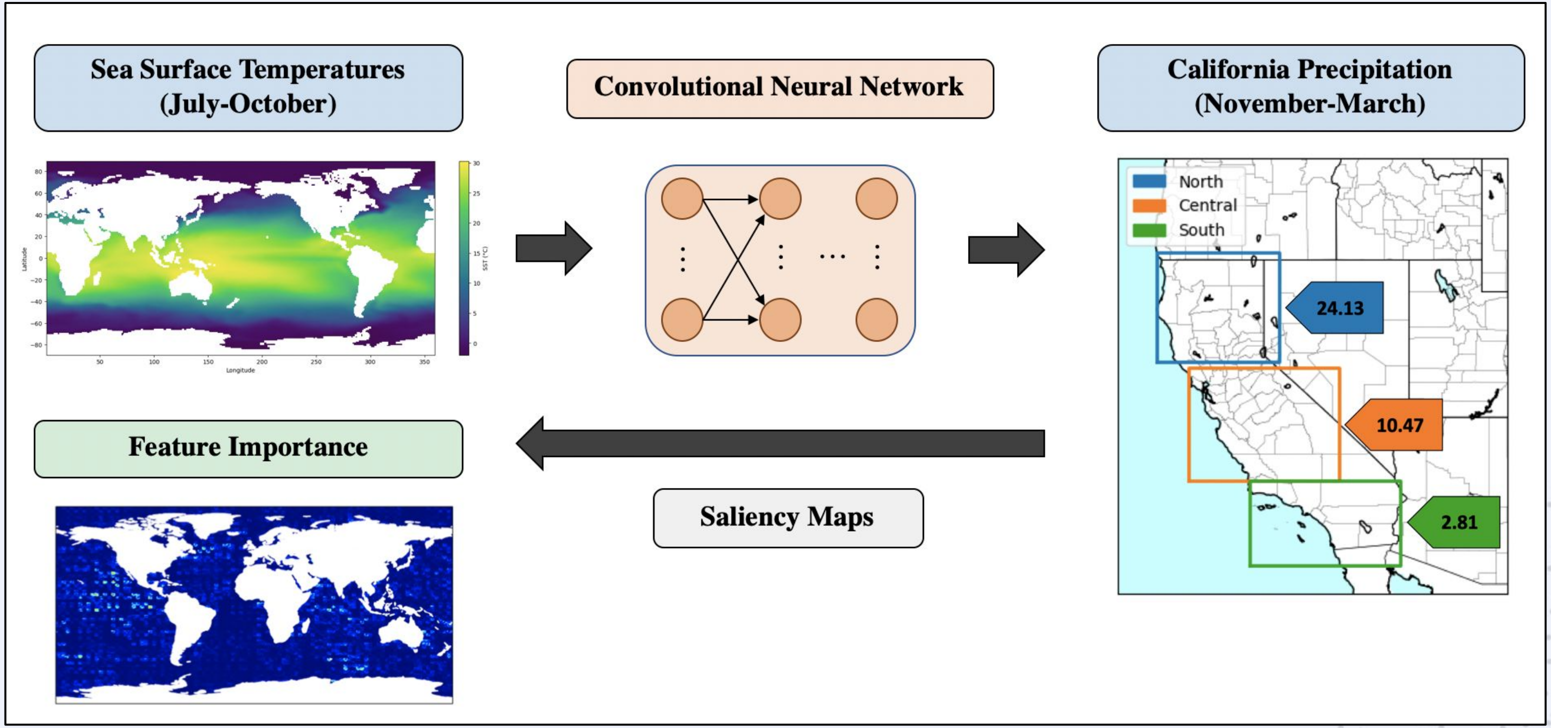
- El Niño Linear Model
  - Most common baseline model



- CNN (separate for each region)
  - Input: Matrix of summer averages of Sea Surface Temperatures
  - Output: Precipitation prediction
  - Pretrained on simulated data and fine-tuned final layer on real observations

- Saliency Maps

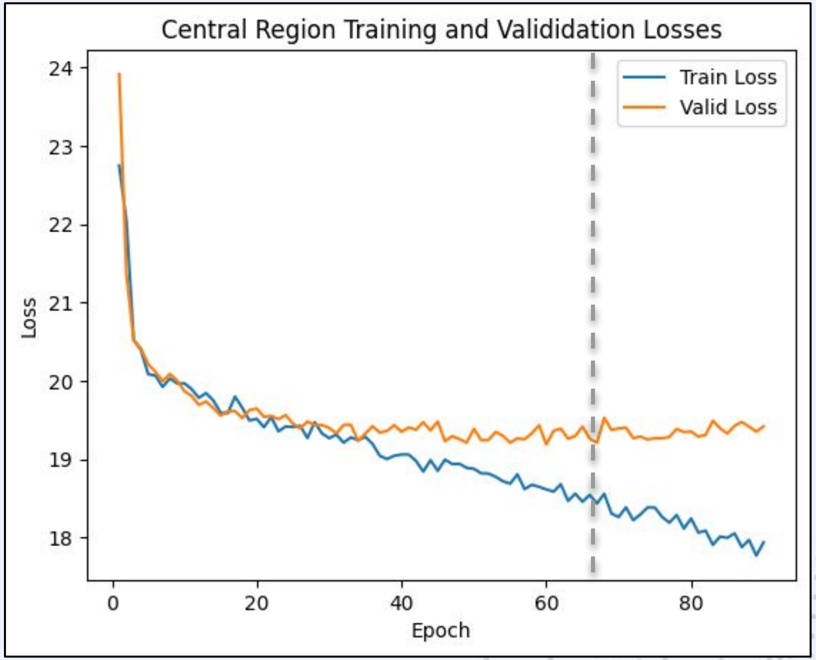
Methods (cont'd)



Results

Results - Model Training

- Model succeeded in learning trends on simulated data
- Employed early stopping to prevent model overfitting on training data
- Model training obtained similar results for North, Central, and South regions



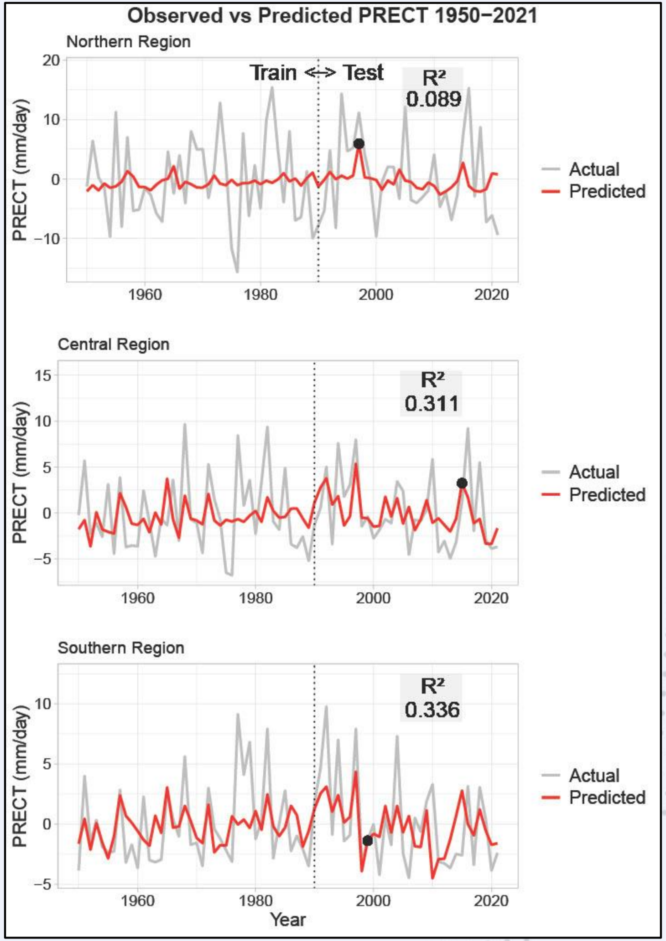
Results - Simulated Data

- Used 80 simulations for training, 10 for validating, 10 for testing
- CNN model outperforms baseline El Niño Model in each region
- Predictive performance is strongest in South, then Central, then North, as expected

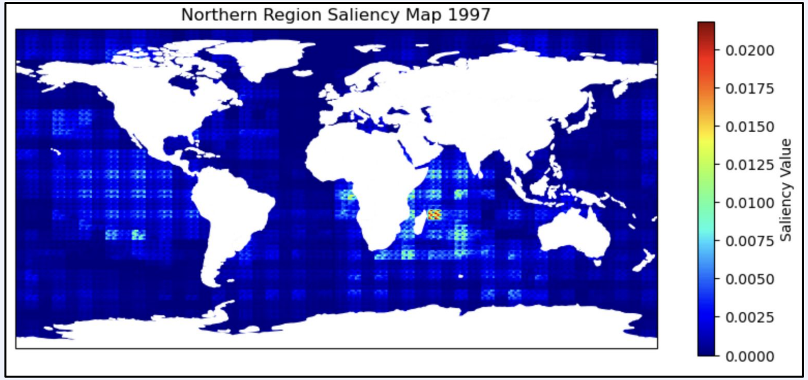
R^2 Values		
	CESM2 Data	
California Region	***El Niño Linear Model***	***CNN Model***
Northern	0.000	0.037
Central	0.084	0.132
Southern	0.155	0.212

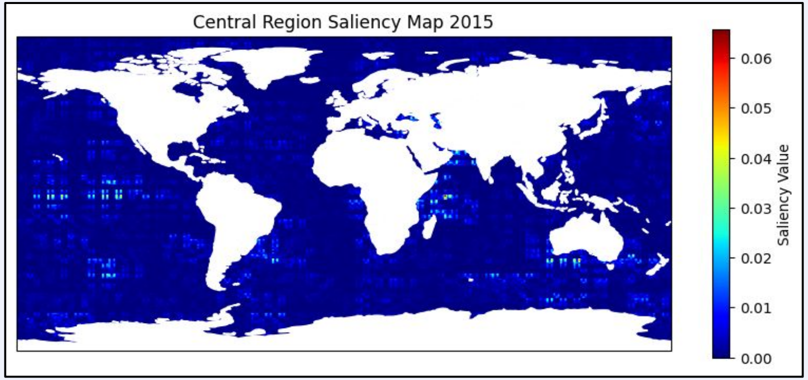
Results - Real World Data

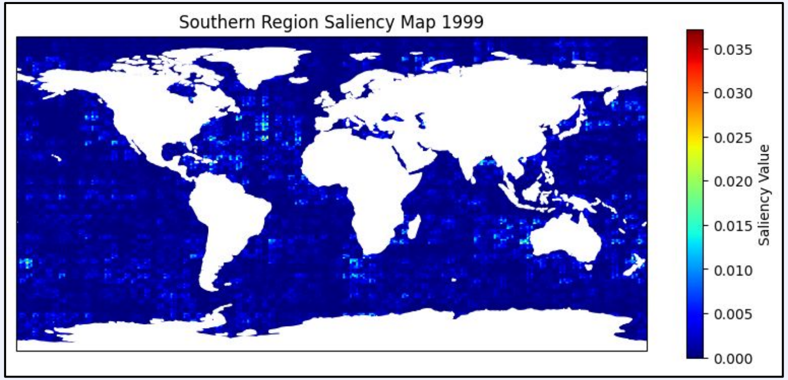
R^2 Values				
	CESM2 Data		*Real World Data (Test Years 1990-2021)*	
California Region	*El Niño Linear Model*	*CNN Model*	*El Niño Linear Model*	*CNN (Fine Tuned)*
Northern	0.000	0.037	0.005	0.089
Central	0.084	0.132	0.029	0.311
Southern	0.155	0.212	0.108	0.336



Results - Saliency Maps







UVA DATA SCIENCE

Conclusion

Conclusion

- CNNs outperformed linear regression across northern, central, and southern California, showing promise for applying Deep Learning to precipitation prediction
- Future steps:
    - Augment simulation data with other climate models (UKESM and CanESM5)
    - Further tune hyperparameters and model structure
    - Include more global weather predictors like air pressure or wind
    - Introduce different time lags to analyze different precipitation patterns
    - Enhance Explainable AI techniques for deeper insights into model predictions

Acknowledgements

Antonios Mamalakis
Sponsor and Faculty Mentor

Assistant Professor of Data Science
School of Data Science

Assistant Professor of Environmental Sciences
Department of Environmental Sciences

Capstone Team

Anthony Chiado

Kristian Olsson

Luke Rohlwing

Michael Vaden