DSC 180B Project Proposal

Keith Kwong
khkwong@ucsd.edu

Jun Zhang
juz018@ucsd.edu

Jack Kai Lim
jklim@ucsd.edu

Duncan Watson Paris
dwatsonparris@ucsd.edu

1 Broad Project Statement

Climate change is an urgent issue that relates to everyone's life. The consequences of climate change, rising temperatures, extreme weather, and rising sea levels, all influence the environment, ecosystems, human societies, and the economy. Due to the urgency of the climate change issue, most of the countries agreed to limit global warming to rise above two Celsius degrees. To address the effect of climate change, climate scientists are able to emulate many of the emission pathways that can reach the goal of the Paris Agreement. However, it's difficult to make the public and policymakers to evaluate which emission pathway. Thus, the machine learning models can help to weigh the pathways by predicting the future climate variables from given emission pathways data.

2 Narrow Project Statement

In response to the escalating challenges posed by climate change, our project aims to refine and enhance the ClimateBench Deep Learning Emulator, originally developed by Duncan. The current models in ClimateBench, as identified in Duncan's seminal work, are foundational but offer significant scope for advancement. Our proposal focuses on exploring and optimizing the architecture of these Machine Learning (ML) models to better address climate-related predictions and analyses.

Key areas of our proposed research include:

Enhancement of Tree Models: We plan to experiment with advanced tree-based models like XGBoost, evaluating their performance in comparison to the existing Random Forest models used in ClimateBench. This involves tuning hyperparameters and possibly integrating novel tree-based techniques to improve predictive accuracy.

Gaussian Process (GP) Optimization: We intend to explore various kernel functions and optimization strategies for Gaussian Processes. This includes testing different kernels, utilizing advanced optimizers, and potentially employing ensemble methods to enhance the GP's predictive capability.

Neural Network Architecture Refinement: Given the complex nature of climate data, we propose to modify and optimize neural network architectures. This involves a deeper understanding of the data to tailor the network layers for more precise modeling, which could include experimenting with different activation functions, layer configurations, and training methodologies.

Exploration of Additional ML Models: Beyond the existing frameworks, we aim to investigate other ML models, such as linear regression, to determine their applicability and effectiveness in this context. This exploration will allow us to assess a broader range of techniques and identify the most suitable approaches for climate-related data analysis.

Our approach is designed to systematically explore and enhance the ClimateBench models, contributing to the broader scientific effort to understand and mitigate the impacts of

climate change using advanced machine learning techniques.

3 Statement of outputs

For our primary deliverable, we will create an interactive website in which people can see and compare the results of the different machine learning models in addition to a report containing all the findings in more technical detail. The main analysis of the data generated by the machine learning models we train will be done by comparing the NRMSE scores obtained for each of them and will be communicated to the broader audience via color maps of the Earth. The website will also contain the ability to compare the model's predictions to that of the NorESM2 and potentially other Earth system models as a point of comparison. In these elements, the predictions produced by any of our machine learning models can be easily analyzed against the performance and predictions of other machine learning models as well as the assumed ground truth predictions produced by the currently existing Earth system models. Should time permit, we also would like to build and host the most successful model which we could then create an API for, allowing others to use the model in their own efforts for climate emulation.

4 Why the project will be successful

As our project is an extension of the ClimateBench research done by Professor Duncan, we will be utilizing much of the same data that we have already gathered and analyzed in Quarter 1. We are also already familiar with the entirety of the data pre-processing performed through our work in Quarter 1, as well as where to find the original, unprocessed data (https://esgf-node.llnl.gov/search/input4mips/), so in the unlike case where we need extra data, it can be obtained without much issue. With our current plans, however, there will be no need to look for other data sets besides the ones we've already worked with in Quarter 1, so all of our focus will be going toward training other models and making improvements to the baseline models we've already created in Quarter 1. Taking all this into consideration, we are confident that we will be able to successfully execute this project within the time period provided.