Homework 1: Final

ATS780: Weather-to-Climate Data Driven Forecasting
Kelsey Ennis
September 20th, 2024

Estimate of Time to Completion: 25 hours
Maximum Allotted Time to Completion: 32 hours
Actual Time to Completion: 27 hours

Introduction

Recently, Deep Learning forecast prediction models (emulators) have been becoming more popular in atmospheric science research and operational weather forecasting. Emulators are proving to be skillful forecast tools; however, forecasts at longer lead times can still result in large errors ("forecast busts"). Common causes of forecast busts include weather and large-scale circulation extremes such as heat waves, atmospheric blocks, and strong frontal passages.

In this project, I focus on the Neural General Circulation Model (NeuralGCM), which includes three different deterministic models with varying resolution and one stochastic model. The deterministic model is trained for short range weather forecasting but is also capable of subseasonal forecast prediction (Kochov et al. 2024). In this project I use the NeuralGCM deterministic 1.4-degree resolution model to investigate how NeuralGCM hindcasts compare to the ERA5 reanalysis for US heat waves at varying lead times. To investigate whether emulators can produce more reliable forecasts for extreme heat and associated circulation regimes, I will generate NeuralGCM hindcasts for 1000-hPa temperature and 500-hPa geopotential height and investigate errors in the model hindcasts during each heat wave event. This work will elucidate whether emulators are a key technological advancement toward improving medium-range weather forecasts of heat extremes.

Data

The NeuralGCM was trained on a super set of ERA5 data that was regridded to the respective Gaussian grids of each of the pretrained deterministic and stochastic models. The predictors/predictands (inputs/outputs) of the model include the vertical structure of wind vectors, temperature, geopotential, specific humidity, specific cloud ice water content, and specific cloud liquid water content on 37 pressure levels (Kochov et al. 2024). ERA5 reanalysis is the 5th generation ECMWF reanalysis (Hersbach et al. 2020). It covers atmospheric, land, and ocean variables and is an hourly dataset with global coverage at approximately 28-km resolution (~0.25° x 0.25°). The data ranges from 1940 to present.

I did not perform data pre-processing for this project. Each of the deterministic and stochastic NeuralGCM models are pre-trained and all the training data were pre-processed by model developers. The preprocessing of ERA5 reanalysis data to regrid it to NeuralGCM's native model resolution for comparison was embedded in the publicly available NeuralGCM python script.

The NeuralGCM model I am using for this project was trained on ERA5 data from the years 1979 to 2019. The testing data used was year 2020 for the 1.4-degree deterministic model and was used to evaluate accuracy of NeuralGCM hindcasts. To evaluate the model hindcasts for the year 2020, the results were compared to ERA5 data and outputs from the X-SHiELD global cloud-resolving model (Kochov et al. 2024).

Methods

I have chosen to investigate 1000-hPa temperature and 500-hPa geopotential height errors in NeuralGCM hindcasts during heat wave events. To do this objectively, I had to create a method to properly identify heat wave events since there is no such archive of these events that currently exists. There is no universal definition of heat waves (Perkins 2015). However, many studies have stipulated a percentile-based threshold approach with a minimum duration of three days (Perkins 2015). To identify a relatively large number of heat wave events across the US, I used ERA5 data (Hersbach et al. 2020) for climatological summers from 2000–2023. I calculated and compared 95th, 98th, and 99th percentile 2-meter temperature events lasting >= three days. The 95th percentile was chosen because during the study period, the 98th and 99th percentiles did not have at least five heat wave events in each US National Climate Assessment (NCA) region (USGCRP 2023). I required each heat wave to have at least 100 continuous grid points, such that it was an event that affected a wide area. There needs to be a minimum size requirement, which must be somewhat arbitrary. For example, in June 2021 the Pacific Northwest was struck with a record-shattering heat wave. Looking at Fig. 1 below, only a few grid points exceeded the 95th percentile threshold on 27 June. In Fig. 2 (28 June), there is continuous red shading over most of the Pacific Northwest, exceeding the 100 continuous grid point criterion; as such, I defined the start day of the heat wave as 28 June. After doing this evaluation throughout the study period, 100 continuous grid points (a 2.5° x 2.5° box if a square) provided a reasonable number of heat wave events during my study period. In addition, an event was allowed to exist in multiple NCA regions, simultaneously and/or over the course of the event. For example, a large heat wave event in 2012 that lasted from 28-June to 1-July affected the Southern Great Plains, Midwest, Northeast, and Southeast regions.

Figure 1: Investigating a Pacific Northwest heat wave event on 27 June 2021. The red shaded area is 2-meter temps > 95th percentile. Note that in Washington and Oregon, there are only a few grid points that are shaded.

Figure 2: As in Fig. 1, but for 28 June 2021. The red area in Washington & Oregon is now a much larger than on 27 June, exceeding the 100 grid point criterion and signifying the start of a heat wave event.

There are seven regions in the US that are defined based on the US National Climate Assessment. For the sake of time, I will evaluate four US regions: the Northwest, Southeast, Midwest, and Northeast. A total of eight heat waves will be chosen for analysis, two from each of the four regions. I will then generate NeuralGCM hindcasts, extract ERA5 reanalysis data for the same dates as the hindcasts and evaluate NeuralGCM's prediction skill at lead times of 10, 15, and 20 days. To evaluate the model prediction skill, both temperature and geopotential height forecast errors will be calculated by taking the model hindcast predictions and subtracting them from the ERA5 reanalysis data.

For each heat wave event, I will decide on 2–3 grid points within the heat wave region to evaluate 1000-hPa temperature error. I am choosing to examine temperature errors at 1000-hPa because NeuralGCM does not have surface variables (i.e., 2-meter temperature) since their model outputs variables on pressure levels only. In this model, the 1000-hPa pressure level is the most representative of the surface, but this will be kept in mind when examining regions at higher elevation. The goal is to choose grid points with varying geographical differences (i.e., coastal proximity) to examine if those differences can have an impact on NeuralGCM's 1000-hPa temperature prediction skill at longer lead times. After zooming in and examining temperature at grid points, I will zoom out and evaluate 500-hPa geopotential height errors for the same events/hindcast dates used in the temperature evaluation. 500-hPa geopotential height provides a useful visualization of large-scale synoptic regimes such as troughs, ridges, atmospheric blocks, etc. During heat waves, geopotential height anomalies are persistently positive, and the anomalies grow as the associated ridge intensifies. This next goal is to assess 500-hPa geopotential height errors in the NeuralGCM hindcasts to determine if the model can understand and predict varying ridge strength during a heat wave at longer lead times.

Historically, dynamical models have struggled with forecasting extremes due to large uncertainty in the model regime. This project will investigate the NeuralGCM prediction skill for 1000-hPa temperature and 500-hPa geopotential height on subseasonal timescales.

References

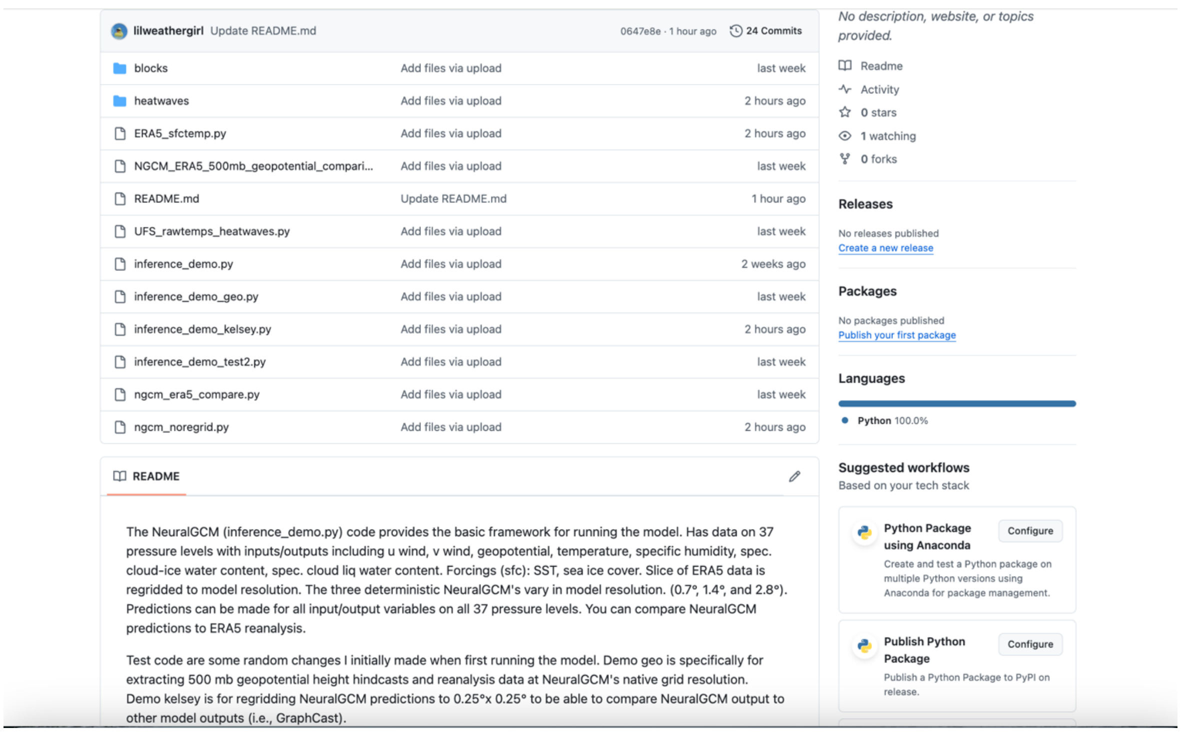
Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc*., 146, 1999–2049, https://doi.org/10.1002/qj.3803.

Kochov, D., and Coauthors, 2024: Neural general circulation models for weather and climate. *Nature*, 632, 1060–1066, https://doi.org/10.1038/s41586-024-07744-y.

Perkins, S. E., 2015: A review on the scientific understanding of heat waves—–Their measurement, driving mechanisms, and changes at the global scale. *Atmos. Res.*, 164, 242–267, https://doi.org/10.1016/j.atmosres.2015.05.014.

USGCRP, 2023: Fifth National Climate Assessment. U.S. Global Change Research Program, 1834 pp., https://doi.org/10.7930/NCA5.2023.

Github repository/commit history



Commits on Sep 18, 2024

Update README.md
lilweathergirl committed 1 hour ago — Verified 0647e8e

Add files via upload
lilweathergirl committed 2 hours ago — Verified 9c3c10d

Add files via upload
lilweathergirl committed 2 hours ago — Verified 0c45033

Add files via upload
lilweathergirl committed 2 hours ago — Verified 9ce1f01

Commits on Sep 11, 2024

Add files via upload
lilweathergirl committed last week — Verified f3262ae

Add files via upload
lilweathergirl committed last week — Verified d53d5dd

Add files via upload
lilweathergirl committed last week — Verified c6e7f8e

Update README.md
lilweathergirl committed last week — Verified f8066ba

Update README.md
lilweathergirl committed last week — Verified 5dd59d3

Add files via upload
lilweathergirl committed last week — Verified d546697

Create docs
lilweathergirl committed last week — Verified f278ebf

Add files via upload
lilweathergirl committed last week — Verified f2fcd27

Create docs
lilweathergirl committed last week — Verified 2df3e05

Add files via upload
lilweathergirl committed last week — Verified 907bbcd

Create docs
lilweathergirl committed last week — Verified 2df3e05

Add files via upload
lilweathergirl committed last week — Verified 907bbcd

Delete docs/heatwaves
lilweathergirl committed last week — Verified 5daf4e7

Create heatwaves
lilweathergirl committed last week — Verified 9777aa7

Delete docs
lilweathergirl committed last week — Verified ae2501c

Create docs
lilweathergirl committed last week — Verified e880f7d

Add files via upload
lilweathergirl committed last week — Verified 1c5865a

Commits on Sep 9, 2024

Update README.md
lilweathergirl committed last week — Verified 5a4a629

Update README.md
lilweathergirl committed last week — Verified 059f145

Add files via upload
lilweathergirl committed last week — Verified 57d8dfb

Commits on Sep 4, 2024

Create README.md
lilweathergirl committed 2 weeks ago — Verified 61485df

Add files via upload
lilweathergirl committed 2 weeks ago — Verified 0d5cedf