199 Directed Research Spring

Temporal and Spatial Analysis of Northern Hemisphere Temperature Variability

Emma Allen
June 2024

Contents

1 Introduction

This project represents a continuation of the analysis conducted during winter quarter. Previously, the variance, skewness, and mean temperature were calculated and found to be consistent with existing literature. Plots from this previous analysis can be found in Appendix A.

This document focuses on progress made this quarter which includes Empirical Orthogonal Function (EOF) analysis encompassing both temporal and spatial dimensions, correlation plots and starting aspects of linear regression for temperature forecasting. The overarching goal is to analyse patterns of sub-seasonal (time range between two weeks and three months) variability of wintertime surface air temperature in the northern hemisphere.

2 Background

2.1 Data

The data utilised in this project was ERA5, the fifth generation ECMWF atmospheric reanalysis [1]. ERA5 was produced by the Copernicus Climate Change Service at ECMWF and provides hourly estimates of atmospheric variables covering global climate from January 1950 to 2021. For this analysis, the only variable used was t2m, temperature two metres above sea level, from 1980 to 2021.

The variable t2m represents the 2-metre air temperature measured in Kelvin. It is stored as a 3D array with dimensions (time, latitude, longitude). The scale factor and add offset are provided for conversion to physical units, however these are automatically accounted for by using the variable() function from netcdf4 module in python [6] therefore not directly utilised.

2.2 Empirical Orthogonal Functions (EOF) Analysis

EOF analysis, also known as Principal Component Analysis is a is a linear dimensionality reduction technique [7]. The high level idea is to break the data down into multiple basis vectors, discard basis vectors with the least variation and keep those with the largest variance, these vectors that are left are the Empirical Orthogonal Functions (EOFs), effectively they are the linear combinations of the original variables that account for the variance in the data. This dimension reduction is highly useful as it increases computational efficiency without huge information loss.

Typically they are calculated through finding the eigenvectors of the covaraince matrix however, in practice calculating the covariance matrix is computationally costly. Singular value decomposition (SVD) is a generalised way of calculating EOFs. At a high level SVD is a mathematical technique M into three matrices: an orthogonal matrix U, a diagonal matrix Σ, and the conjugate transpose of an orthogonal matrix V [2]. This allows us to find the most important features in the data.

This method can be used to find EOFs as the columns of U are the eigenvectors of the EOF analysis, The columns of V are the Principal Components (PC), or EOFs (normalised to unit variance) and the diagonal elements of Σ are the amplitudes corresponding to the PC pairs, this is analogous to eigenvalues, but with a different scaling.

2.3 Auto-Correlation and Cross-Correlation

Auto-correlation is a method that looks at correlation in one variable with itself at different points in the time series [5]. Effectively this analysis measures how a value is correlated with itself at different points in time. The correlation is calculated using the variable at the current time and at a time lag, the value of the variable at a previous point in time.

Cross-Correlation is a measure that looks at the correlation between two or more variables relative to each other [3]. Typically it is used to compare multiple time series and determine how strong the relationship between them is and when this correlation is strongest.

3 Results and Discussion

3.1 EOF Analysis

Figures 1 and 2 depict the time series of principal components one and two respectively. These plots exhibit characteristics aligned with our expectations: the average values hover around zero for each component, and the magnitude of maximum variance does not exceed four kelvin.

These principal component time series captures the temporal evolution of the dominant modes of variability in the dataset. The fluctuations in the principal components reflect the underlying dynamics of the system.

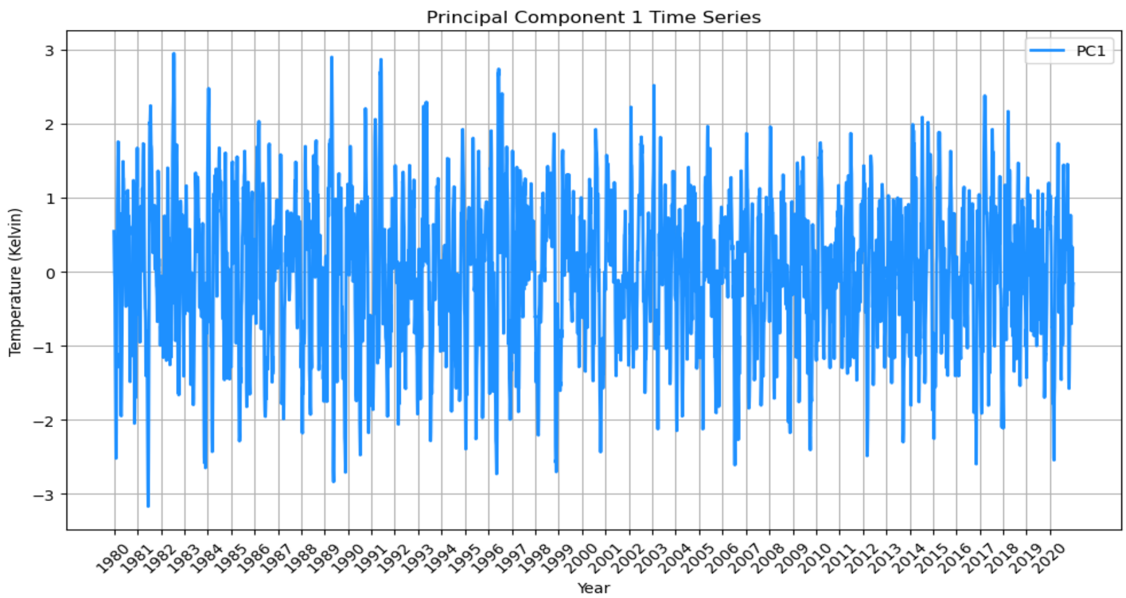

Figure 1: Principal Component One from 1980 to 2021

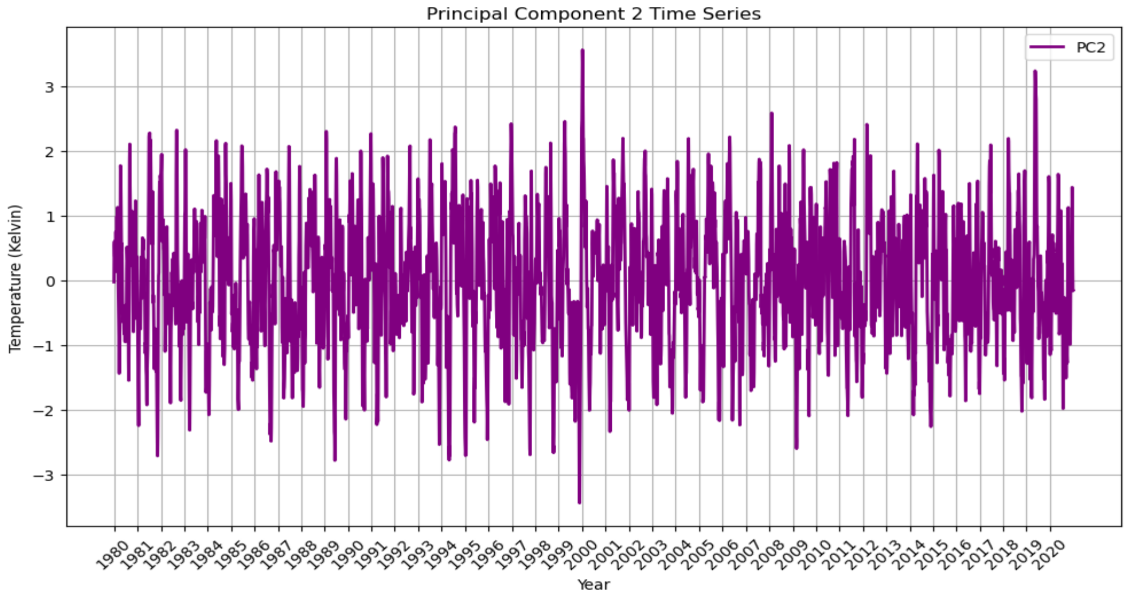

Figure 2: Principal Component Two from 1980 to 2021

Figures 3 and 4 show the spatial plot of principal components one and two respectively. Comparing these plots to the literature [4], one observes good agreement.

EOF1 and EOF2 emerge as the primary modes of wintertime Northern Hemisphere surface air temperature variability on the sub-seasonal time scale. These patterns exhibit distinct spatial characteristics, with EOF1 predominating in North America and EOF2 prominent in Eurasia.

Collectively, these EOFs capture a significant portion of the local variance. EOF1 is characterised by one large centre of variance centred in North America, while EOF2 is exhibits a large centre of opposite sign across Eurasia, with smaller pools of variance across North America and Northwestern Russia.

In both EOF1 and EOF2, a prominent feature is the maximum centre of T2m variability is situated over the central continent near 50°N. This pattern likely reflects a significant movement of air mass on a large scale between the high-latitude region and the central continent [4]. Consequently, on the sub-seasonal time scale, a prolonged cold period in North America tends to be associated with warm anomalies near Alaska, while a lasting cold spell in Siberia is often linked to warm anomalies over the Arctic coastal region of northwest Eurasia.

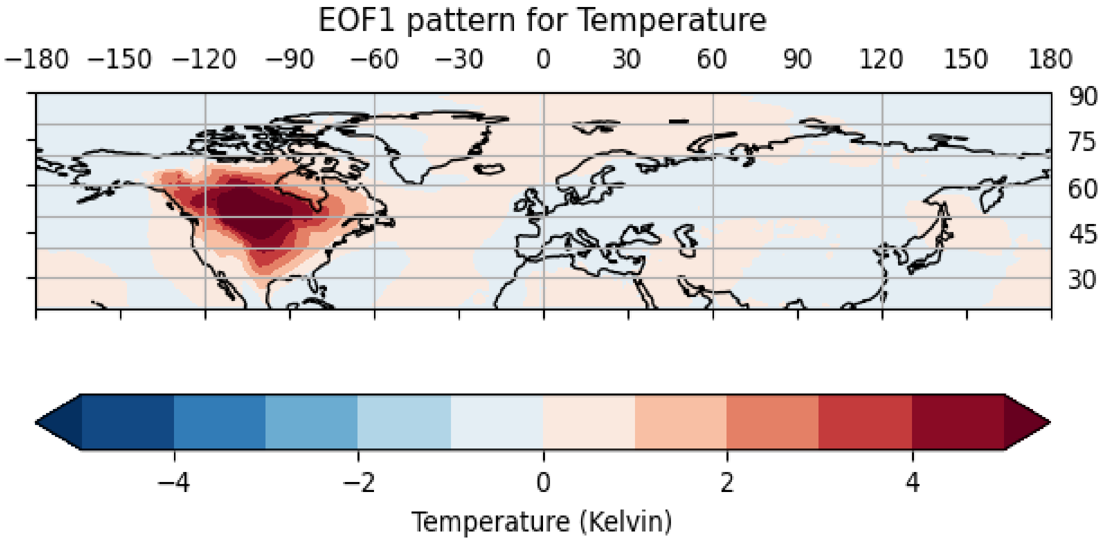

Figure 3: EOF1 from 1980 to 2021

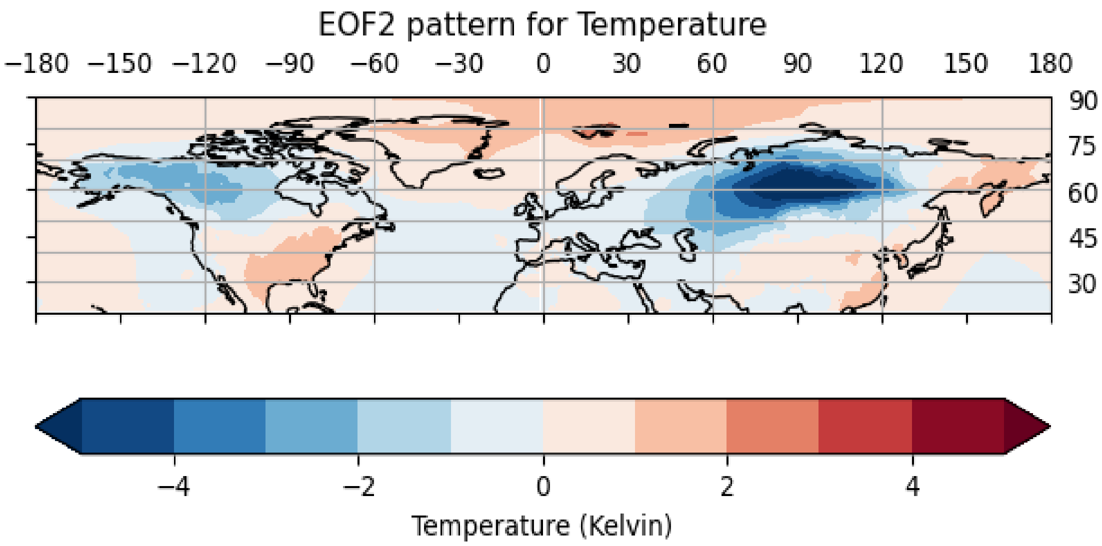

Figure 4: EOF2 from 1980 to 2021

3.2 Auto-Correlation and Cross Correlation

Figures 5 and 6 depict the auto-correlation and cross-correlation between principal components one and two, respectively. We observe a lagged correlation between PC1 and PC2, suggesting that EOF2 lags EOF1 by approximately two days, this agrees with the literature [4]. However if the graph is extended the auto-correlation decreases to -0.2 indication a weak negative correlation until it increases and stabilises around zero indicating very slight weak correlation as the number of days increases to roughly 40.

The temporal relationship between EOF1 and EOF2 indicates an intercontinental connection of temperature anomalies on the sub-seasonal time scale. However, it's important to note that this correlation does not imply causation; rather, it suggests a common source of variability operating at different time lags.

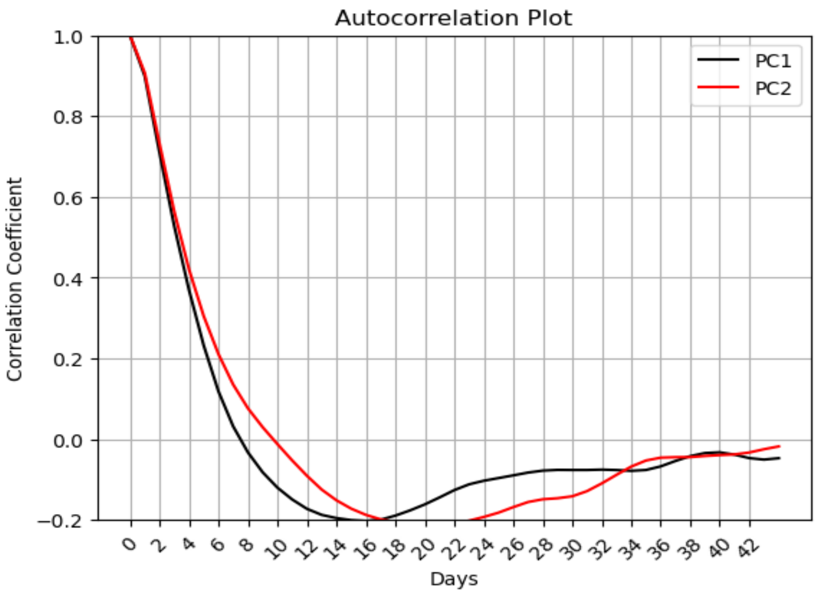

Figure 5: Auto Correlation of PC1 and PC2

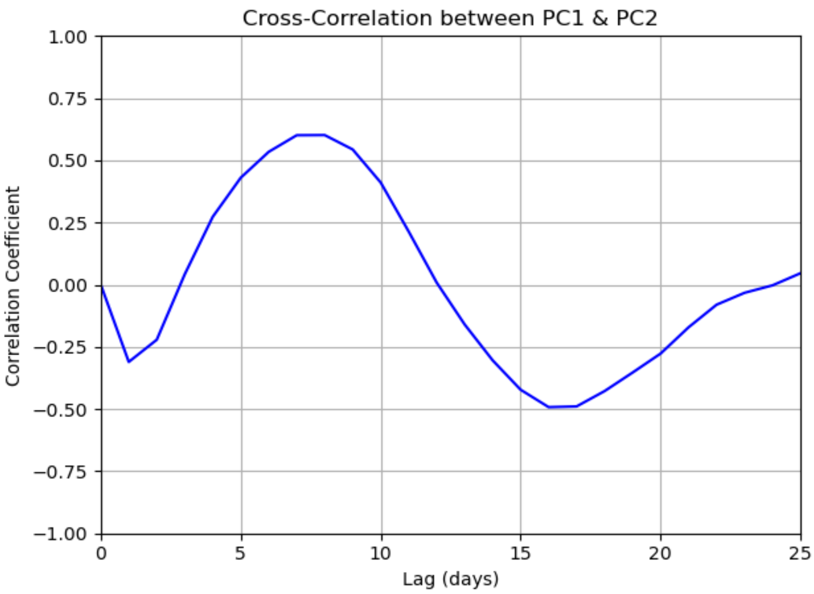

Figure 6: Cross Correlation Between PC1 and PC2

4 Conclusions

The two dominant modes of surface air temperature sub-seasonal variability in wintertime Northern Hemisphere are identified in this study. They are well separated geographically and account for a large part of local variability in North America and Eurasia, respectively.

In the future, this analysis provides an excellent basis to use machine learning techniques such as linear regression and transformers for predicting surface temperature across North America. Additionally, alternative methods of sub-seasonal prediction, such as correlation skill, could be investigated to enhance the accuracy and reliability of these forecasts.

References

[1] Copernicus Climate Change Service. Era5: Fifth generation of ecmwf atmospheric reanalyses of the global climate. Copernicus Climate Change Service (C3S).

[2] Gregory Gundersen. Singular value decomposition. `https://gregorygundersen.com/blog/2018/12/10/svd/`, December 10 2018.

[3] Adam Hayes. Cross-correlation. `https://www.investopedia.com/terms/c/crosscorrelation.asp#:~:text=Cross%2Dcorrelation%20is%20a%20measurement,point%20the%20best%20match%20occurs`, 2023. Accessed: 2024-05-31.

[4] Hai Lin. Predicting the dominant patterns of subseasonal variability of wintertime surface air temperature in extratropical northern hemisphere. *Geophysical Research Letters*, 45(8):3842–3849, 2018.

[5] Joshua Noble and Eda Kavlakoglu. Autocorrelation. `https://www.ibm.com/topics/autocorrelation`, 2024. Accessed: 2024-05-31.

[6] Unidata. *netCDF4-python Documentation*, 2024. Accessed: March 2024.

[7] Zhihua Zhang and John C. Moore. Empirical orthogonal functions. In *Mathematical and Physical Fundamentals of Climate Change*, pages 285–318. Elsevier, 2015.

A Appendix A

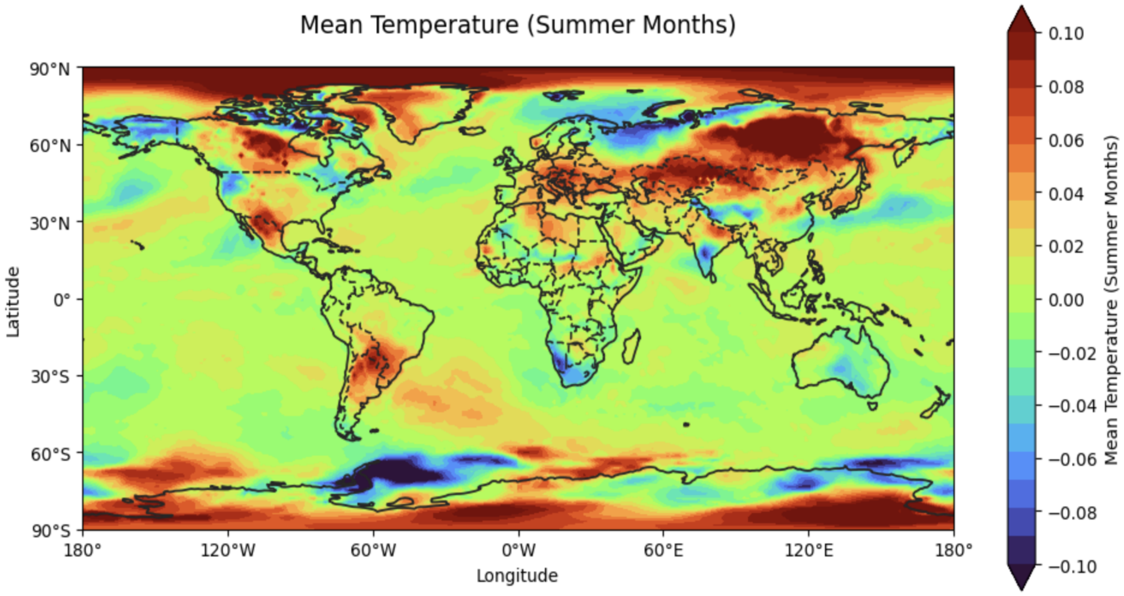

Figure 7: Mean Summer Temperature from 1980 to 2021

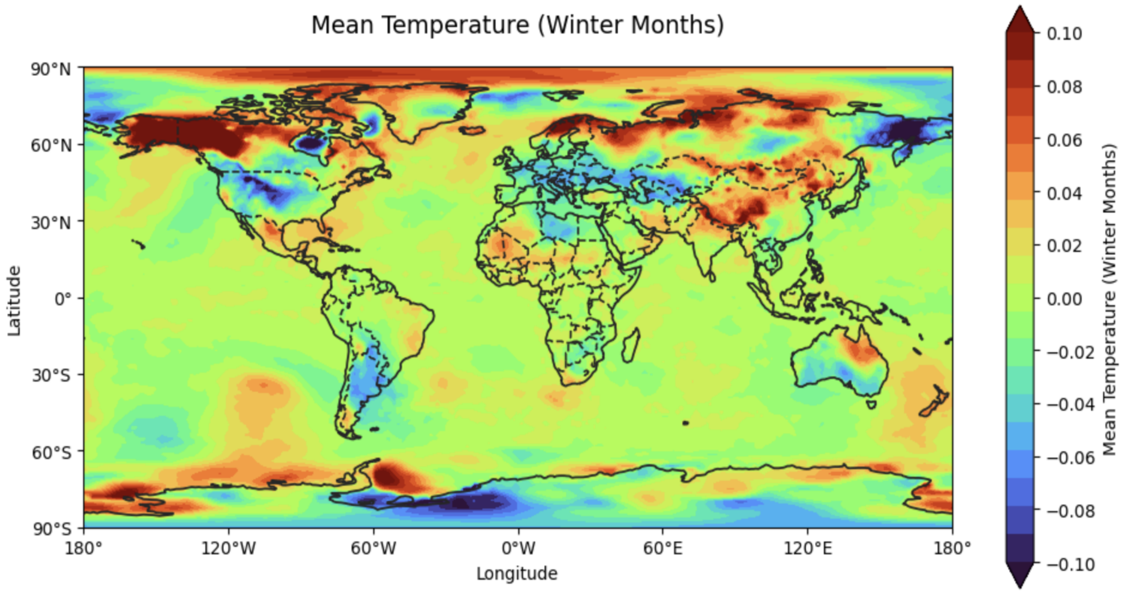

Figure 8: Mean Winter Temperature from 1980 to 2021

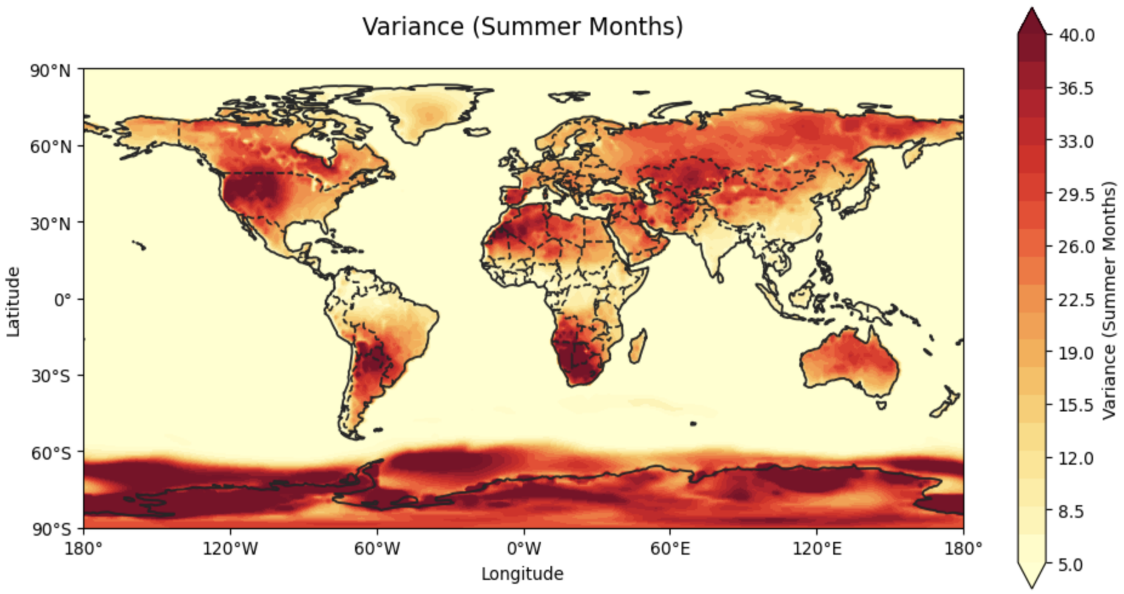

Figure 9: Variance in Summer Temperature from 1980 to 2021

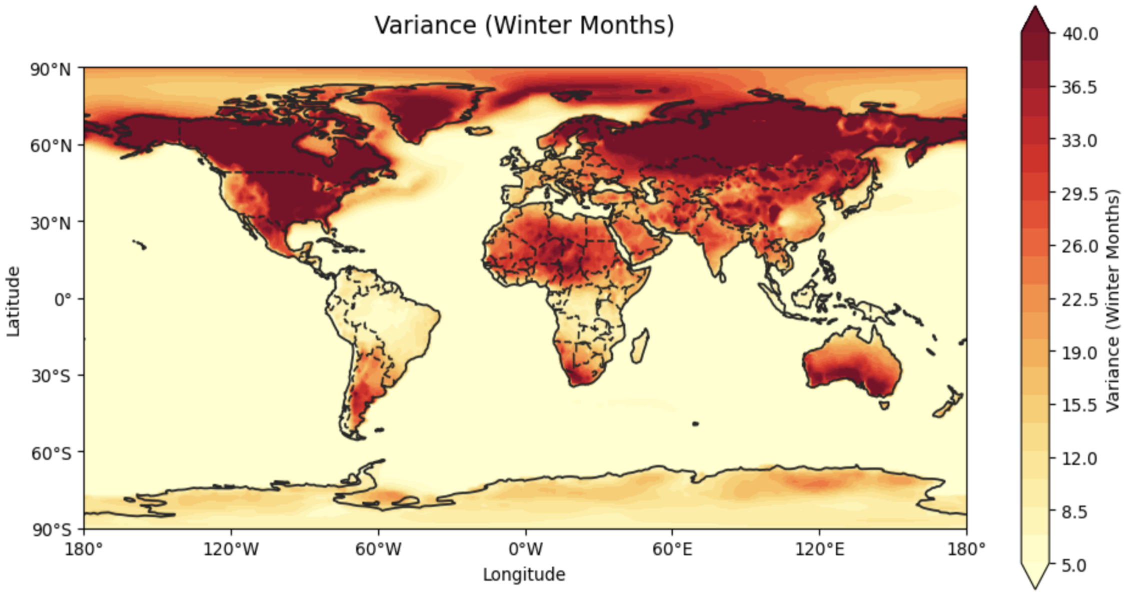

Figure 10: Variance in Winter Temperature from 1980 to 2021

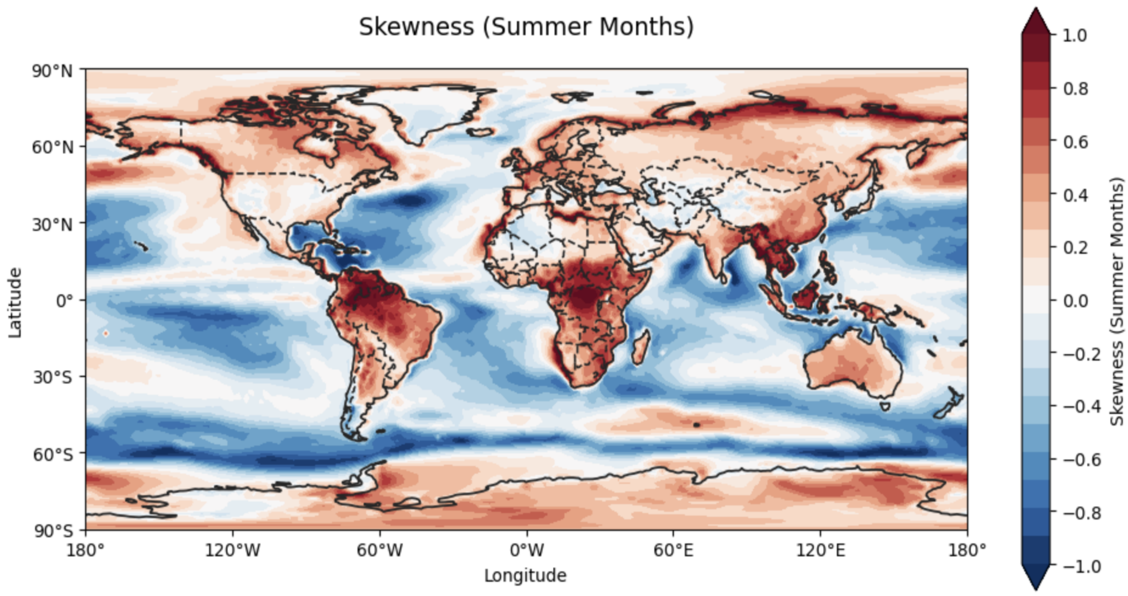

Figure 11: Skewness in Summer Temperature from 1980 to 2021

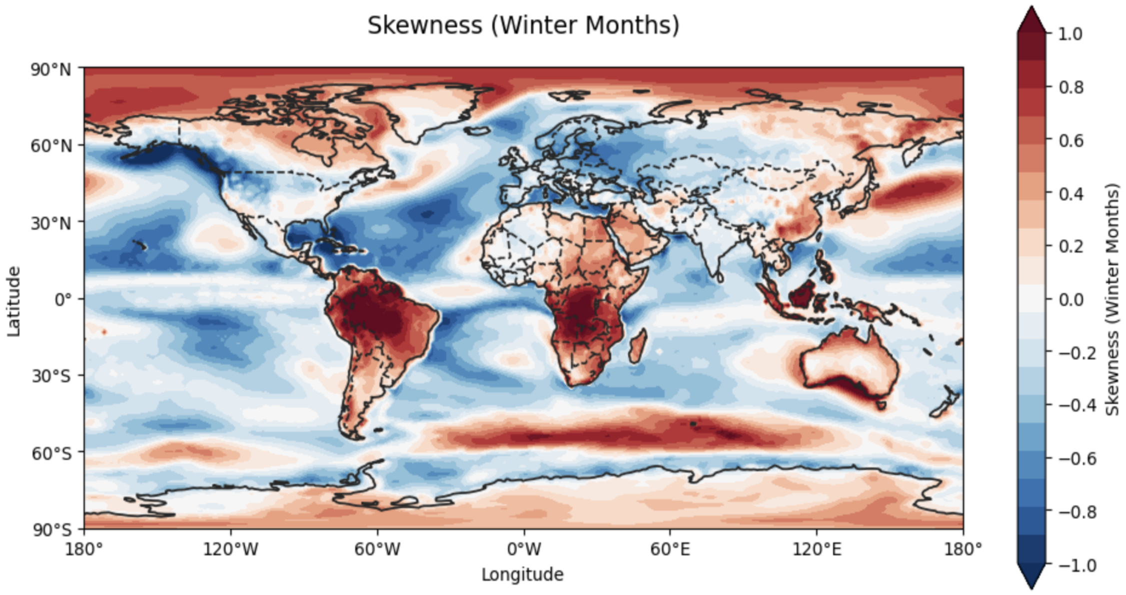

Figure 12: Skewness in Winter Temperature from 1980 to 2021