Empirical Orthogonal Function Analysis of Northern Hemisphere Sub-Seasonal Temperature Variability

Emma Allen

Supervisor: Prof. Gang Chen

July 2024

Abstract

Accurate prediction of sub-seasonal temperature variability in the Northern Hemisphere is critical for improving weather forecasts and mitigating societal impacts from extreme events. This study applies Empirical Orthogonal Function (EOF) analysis to wintertime surface air temperature data from ECMWF Reanalysis v5 (ERA5) to identify the dominant modes of variability over North America and Eurasia. Temporal autocorrelation and cross-correlation analyses of the Principal Components (PCs) reveal the persistence and lagged relationships between these modes, providing insight into the underlying dynamics. The leading modes capture a substantial fraction of sub-seasonal temperature variability, suggesting that their accurate prediction could enhance on a subseasonal timeframe.

Contents

1 Introduction

Accurate weather forecasting is critically important, both for our daily lives and its potential impact on a range of industries including infrastructure and energy [1]. Consequently, reliable predictions of such events are essential. While short-range numerical weather forecasts and seasonal climate predictions are well-established and highly developed, subseasonal forecasts, those covering periods of roughly two weeks to one month, remain relatively new and introduce additional challenges for accurate prediction [2, 3, 4].

To Improve these forecasts it is essential to identify the dominant patterns of Northern Hemisphere temperature variability on subseasonal timescales. By identifying these dominant patterns, we can not only improve forecasts but also reduce the computational burden, since analysing the leading modes of variability is more efficient than working with the full dataset. It is also interesting to probe the sources of predictability and to represent them effectively in numerical models. The dominant mode of intra-seasonal variability in the tropics, the Madden-Julian Oscillation (MJO) [5], has been identified as a major source of predictability on the subseasonal time scale [6]. However, whether and how improved prediction of the MJO translates to enhanced extratropical temperature forecasts remains an open question.

Motivated by these considerations, this work primarily aims to determine the dominant patterns of Northern Hemisphere temperature variability on the subseasonal timescale using Empirical Orthogonal Function (EOF) analysis. A secondary aim is to explore the sources of predictability through statistical characterisation of the data. The structure of the paper is as follows: we first present the mathematical background underlying our methods, followed by a discussion of the data and methodology, including initial statistical analyses, EOF analysis, and auto and cross correlation analyses, before presenting the results and conclusions.

2 Mathematical Background

2.1 Empirical Orthogonal Function (EOF) Analysis

EOF analysis is a statistical method for reducing the dimensionality of a dataset with many interrelated variables, while retaining as much of the total variance as possible [7]. This method is closely related to Principal Component Analysis (PCA) [8] to the point where some literature uses these terms interchangably however in climate science these have distinct meanings. EOFs refer to the spatial patterns describing how variability is distributed across a domain, while the corresponding Principal Components (PCs) represent the temporal evolution and amplitude of each spatial mode.

Together, these EOF-PC pairs (or modes) provide an efficient representation of the dataset, with the leading modes capturing the largest fraction of the total variance. This separation of space and time is particularly valuable for forecasting, as it allows models to predict the temporal behaviour of a small number of dominant patterns rather than the full high-dimensional field. To formalise this, we mathematically define and present a method calculate the EOFs.

To begin, suppose $\mathbf{x}$ is a vector of p random variables, and that we are interested in their variances and the structure of their covariances or correlations. For large p, inspecting all p variances and the $\frac{1}{2}p(p-1)$ covariances directly is often uninformative. EOF analysis addresses this by identifying a small number ($m \ll p$) of derived variables that preserve most of the information in the variance–covariance structure.

The first step is to find a linear combination $\alpha_1'\mathbf{x}$ of the elements of $\mathbf{x}$ with maximum variance

where $'$ denotes transpose

$$\alpha_1'\mathbf{x} = \alpha_{11}x_1 + \alpha_{12}x_2 + \cdots + \alpha_{1p}x_p = \sum_{j=1}^{p} \alpha_{1j}x_j,$$

where $\alpha_1 = (\alpha_{11}, \alpha_{12}, \ldots, \alpha_{1p})^\top$ is a vector of constants.

Next, we look for a second linear combination $\alpha_2'\mathbf{x}$, uncorrelated with $\alpha_1'\mathbf{x}$, that has maximum variance. This process continues so that at the kth stage, $\alpha_k'\mathbf{x}$ is found to maximise variance while remaining uncorrelated with all previous $\alpha_1'\mathbf{x}, \ldots, \alpha_{k-1}'\mathbf{x}$. The kth derived variable $\alpha_k'\mathbf{x}$ corresponds to the kth EOF mode. Theoretically up to p EOFs can be computed, but typically most of the variation in $\mathbf{x}$ is captured by m EOFs where $m \ll p$.

Having defined EOFs we now explain how to calculate them. For this consider the case where $\mathbf{x}$ has a covariance matrix $\mathbf{\Sigma} \in \mathbb{R}^{p\times p}$. The (i,j)th element of $\mathbf{\Sigma}$ is the covariance between x_i and x_j $(i \neq j)$ or the variance of x_j if $i = j$. In practice, $\mathbf{\Sigma}$ is usually unknown and estimated from the data by the sample covariance matrix $\mathbf{S}$, however here we assume we know $\mathbf{\Sigma}$. To begin consider $\boldsymbol{\alpha}_1'\mathbf{x}$, the vector $\boldsymbol{\alpha}_1'$ maximises $\text{var}[\boldsymbol{\alpha}_1'\mathbf{x}] = \boldsymbol{\alpha}_1'\mathbf{\Sigma}\boldsymbol{\alpha}_1$. In order for this to be achieved we must add a normalisation constraint

$$\boldsymbol{\alpha}_1'\boldsymbol{\alpha}_1 = 1.$$

Thus all together we have a constrained optimisation problem: To maximise $\text{var}[\boldsymbol{\alpha}_1'\mathbf{x}] = \boldsymbol{\alpha}_1'\mathbf{\Sigma}\boldsymbol{\alpha}_1$ subject to our constraint above. This can be solved using Lagrange multipliers by maximising

$$\mathcal{L}(\boldsymbol{\alpha}_1, \lambda) = \boldsymbol{\alpha}_1'\mathbf{\Sigma}\boldsymbol{\alpha}_1 - \lambda(\boldsymbol{\alpha}_1'\boldsymbol{\alpha}_1 - 1),$$

where λ is the Lagrange multiplier. Differentiating with respect to $\boldsymbol{\alpha}_1$ and setting the derivative to zero gives

$$\mathbf{\Sigma}\boldsymbol{\alpha}_1 - \lambda\boldsymbol{\alpha}_1 = \mathbf{0},$$

or equivalently

$$(\mathbf{\Sigma} - \lambda\mathbf{I_p})\boldsymbol{\alpha}_1 = \mathbf{0},$$

where $\mathbf{I_p}$ is the $p \times p$ identity matrix.

Thus, λ is an eigenvalue of $\mathbf{\Sigma}$ and $\boldsymbol{\alpha}_1$ the corresponding eigenvector. The variance of $\boldsymbol{\alpha}_1'\mathbf{x}$ is

$$\boldsymbol{\alpha}_1'\mathbf{\Sigma}\boldsymbol{\alpha}_1 = \lambda\,(\boldsymbol{\alpha}_1'\boldsymbol{\alpha}_1) = \lambda,$$

so to maximise variance, λ must be as large as possible. Therefore, $\boldsymbol{\alpha}_1$ is the eigenvector corresponding to the largest eigenvalue λ_1 of Σ. In general, the kth EOF satisfies

$$\mathbf{\Sigma}\boldsymbol{\alpha}_k = \lambda_k\boldsymbol{\alpha}_k, \quad \boldsymbol{\alpha}_k'\boldsymbol{\alpha}_k = 1,$$

The projection of the field (in our case, this will always be a temperature field) X^0 onto the k-th EOF a_k, i.e., $c_k = X^0 a_k$ is the k-th principal component (PC)

$$c_k(t) = \sum_{s=1}^{p} x^0(t,s)\, a_k(s)$$

In summary, we have shown how to take a dataset and reduce the dimensions to k leading modes. These EOFs contain the majority of the variance present in the original dataset. The key advantage of this approach is simple: by predicting the PCs associated with the leading EOFs, we can capture the majority of the system's variance and therefore forecast the dominant behaviour of the climate system with relatively few variables.

Having defined the EOFs and PCs and their role in capturing the dominant modes of variability, we now turn to the mathematical tools used to quantify their temporal relationships, namely auto-correlation and cross-correlation, which are essential for understanding persistence and lead-lag connections in subseasonal prediction.

2.2 Auto and Cross Correlation

Autocorrelation measures the similarity between a time series and a shifted, or lagged, version of itself over successive time intervals [9]. In essence, it quantifies the relationship between a variable's current value and its past values. Although there are several ways to calculate autocorrelation, we use the most common metric, the Pearson Product Moment Correlation coefficient, defined for a lag k as

$$\rho_k = \frac{\sum_{t=1}^{N-k}(x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^{N}(x_t - \bar{x})^2},$$

where x_t is the value of the time series at time t, $\bar{x}$ is its mean, N is the total number of observations, and k is the lag under consideration.

Calculating ρ_k for various lags produces the autocorrelation function which is often displayed graphically with autocorrelation plotted against lag time. By definition, $\rho_0 = 1$ since any unshifted series is perfectly correlated with itself. As with ordinary correlation, $\rho_k > 0$ indicates a positive correlation, $\rho_k = 0$ indicates no correlation, and $\rho_k < 0$ indicates a negative correlation. The autocorrelation function is bounded between -1 and $+1$.

Cross-correlation is similar in definition to autocorrelation, except that it measures the correlation between two distinct time series, shifted relative to each other. Given two time series $\{x_t\}$ and $\{y_t\}$, the Pearson Product Moment definition of the cross-correlation at lag k is

$$\rho_{xy}(k) = \frac{\sum_{t=1}^{N-k}(x_t - \bar{x})(y_{t+k} - \bar{y})}{\sqrt{\sum_{t=1}^{N}(x_t - \bar{x})^2}\,\sqrt{\sum_{t=1}^{N}(y_t - \bar{y})^2}},$$

where $\bar{x}$ and $\bar{y}$ are the means of the respective series. Cross-correlation functions are often used to identify time-shifted relationships between two signals, such as leading or lagging effects, and are similarly bounded between -1 and $+1$.

By combining autocorrelation and cross-correlation analysis with the EOF framework, we can quantify the temporal persistence of dominant modes and the lead-lag relationships between them. These tools form the mathematical foundation for the methodology and data analysis described in the following section.

3 Data and Methodology

The data utilised in this project was ERA5, the fifth generation ECMWF atmospheric reanalysis [10]. ERA5, produced by the Copernicus Climate Change Service at ECMWF, provides hourly estimates of numerous atmospheric variables, covering the global climate from January 1940 to the present. This analysis uses only the variable t2m (air temperature measured two meters above the surface, measured in Kelvin), covering the wintertime (December, January and February) from 1980 to 2021.

The variable t2m is stored as a three-dimensional array with the dimensions: time, latitude, and longitude. Although the dataset provides a scale factor and an offset for converting the stored values into physical units, these adjustments are automatically applied when using the `variable()` function from the Python `netCDF4` module [11], and therefore do not need to be handled manually.

The initial step in data preprocessing involves detrending the time series at each spatial grid point to remove long-term linear trends. This is achieved by applying an inverse Fast Fourier Transform, which effectively isolates the detrended fluctuations. Detrending ensures that subsequent analyses focus on variability around a stable baseline climate.

Following detrending, an initial statistical characterisation of the temperature field is performed by calculating three key moments: the mean, variance, and skewness. Next, EOF analysis is performed on the temperature dataset to extract the leading modes of variability. This identifies the dominant spatial patterns (EOFs) and their corresponding PCs, providing a compact representation of the system's variability. Following this, autocorrelation functions are computed for each PC to quantify the persistence of these dominant modes over time. Cross-correlation analysis is then applied to examine lagged relationships between the PCs.

Having outlined our extraction method for the dominant modes via EOF we now move on to present our results.

4 Results and Discussion

4.1 Initial Statistical Analysis

We begin by examining the mean winter surface temperature deviation from 1980 to 2020 (Figure 1). The Northern Hemisphere shows a pronounced warming signal, strongest over high-latitude continental interiors such as Alaska and northern Canada. Dark red and brown regions across the Arctic highlight substantial warming, consistent with polar amplification driven by snow–ice and lapse rate feedbacks [12, 13]. In contrast, Antarctica exhibits a more heterogeneous pattern with weaker warming, reflecting the moderating influence of the Southern Ocean and the limited land area at comparable southern latitudes. Localised cooling occurs in parts of North America, Europe, and Siberia, influenced by regional circulation anomalies, snow-cover feedbacks, and jet stream variability. Near the equator, anomalies are generally close to zero (light green shading), consistent with the high heat capacity of tropical oceans and smaller seasonal variations in solar insolation [13].

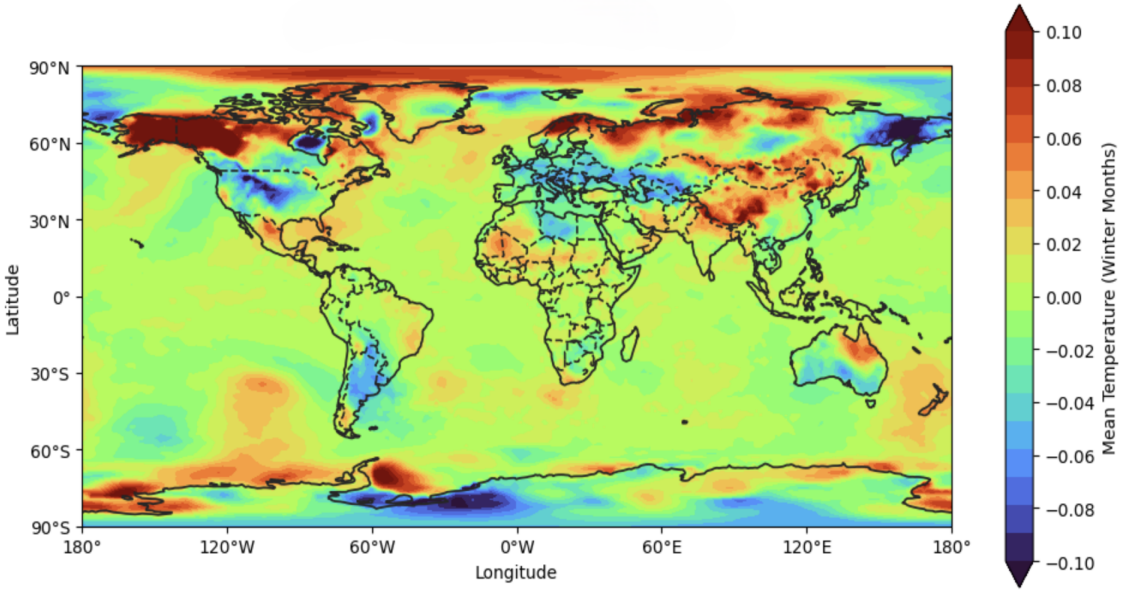

Figure 1: Mean surface temperature deviation for winter (1980–2020).

Now we move on to look at the spatial distribution of winter temperature variance, presented in Figure 2. Here we see three dominant patterns. The largest values ($> 36\,^{\circ}\mathrm{C}^2$) occur in the Arctic and sub-Arctic, including northern Canada, Siberia, and Alaska. This is consistent with polar amplification, where feedbacks such as the sea ice–albedo effect and lapse rate feedback increase both the magnitude and variability of temperature anomalies [14, 15]. Reduced sea ice cover enhances autumn heat release from open water, and coupled with altered circulation patterns such as a weakened polar vortex, leads to large interannual variability [16]. Elevated variance (26 - $33\,^{\circ}\mathrm{C}^2$) is also found over continental interiors like Central Asia, North America, and the Sahel, due to the low heat capacity of land, variability in snow cover, and the influence of blocking high-pressure systems [17, 18]. In the Southern Hemisphere, isolated high-variance regions ($\sim$ 22 - $29\,^{\circ}\mathrm{C}^2$) appear in mid-latitudes such as southern Australia, South Africa, and Patagonia, driven by shifts in storm tracks, ENSO teleconnections, and rapid alternation between maritime and continental air masses [19, 20].

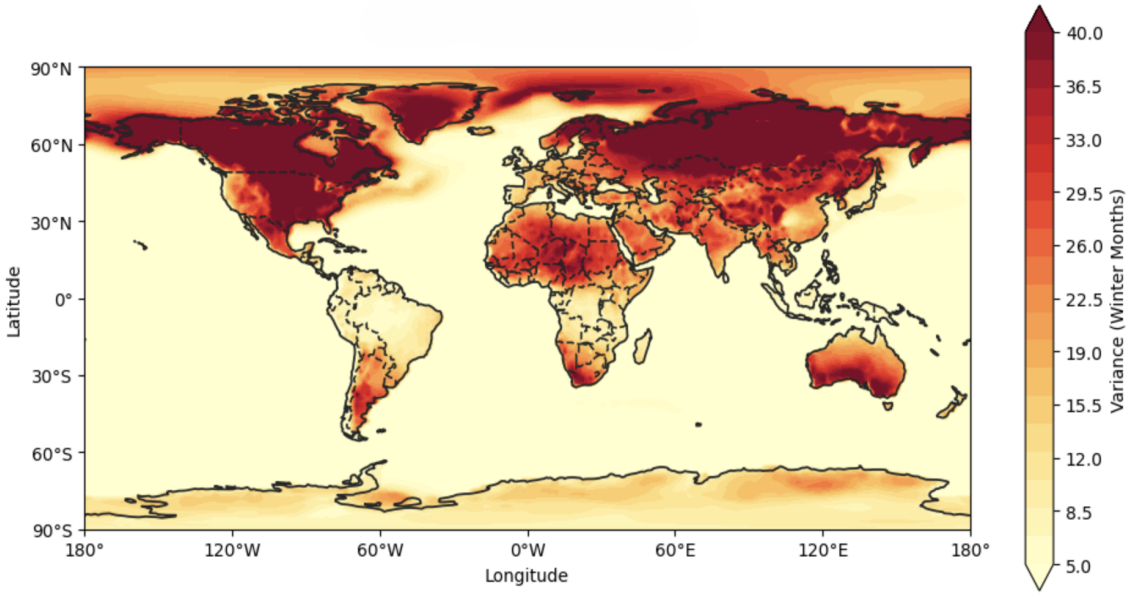

Figure 2: Variance of surface temperature for winter (1980–2020).

Finally we look at the skewness in winter temperature anomalies (Figure 3). This metric provides additional insight into the asymmetry of temperature distributions. Positive values indicate that warm anomalies are more frequent or intense than cold anomalies, while negative values indicate the reverse. Strong positive skewness dominates high-latitude continental interiors such as Siberia, northern Canada, and the Arctic, reflecting conditions that favour strong warm anomalies, likely driven by sea ice loss, enhanced autumn heat release, and circulation changes and suppressing prolonged cold spells. These patterns are further shaped by shifts in the Arctic Oscillation and a weakened polar vortex, which influence cold air persistence.

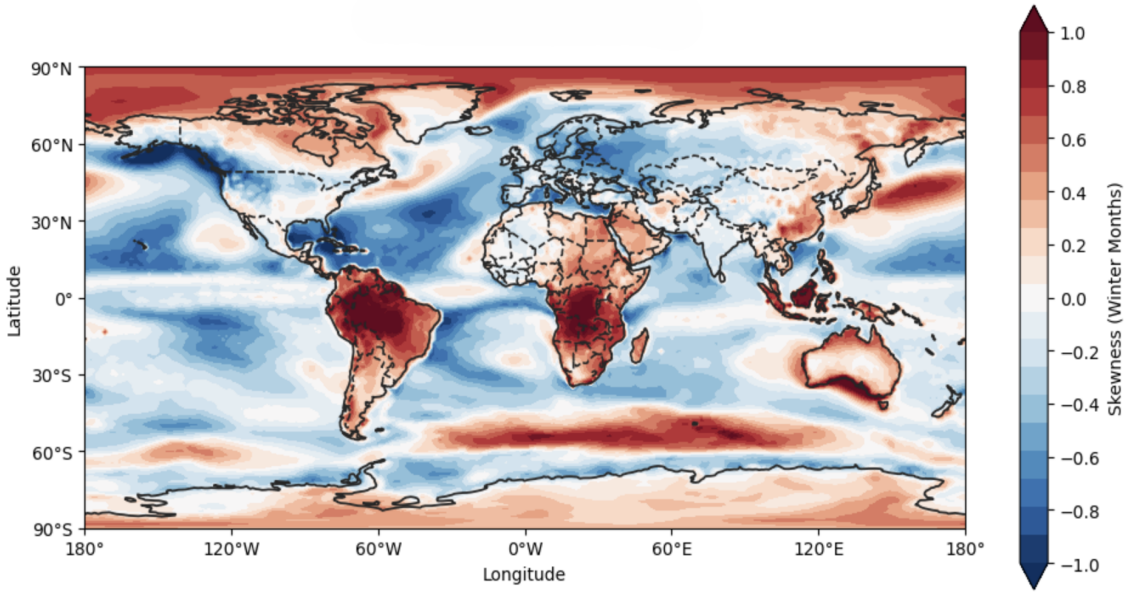

Figure 3: Skewness of surface temperature for winter (1980–2020).

All together, the mean, variance, and skewness fields establish the spatial and statistical structure of winter surface temperature deviations. These metrics provide the physical and statistical

context for the EOF analysis presented in the next section, which identifies and quantifies the dominant modes of variability underlying these observed patterns.

4.2 EOF Analysis

To determine the dominant modes of variability of wintertime temperature we now turn to EOF analysis. Figures 4 and 5 display the corresponding spatial patterns for EOF1 and EOF2. These emerge as the leading modes of wintertime Northern Hemisphere surface air temperature variability on sub-seasonal time scales. EOF1 exhibits its largest centre of variability over central North America, while EOF2 is dominated by a broad centre of opposite sign across Eurasia, accompanied by smaller, opposite-signed regions over North America and northwestern Russia. This pattern is similar to that presented in [21].

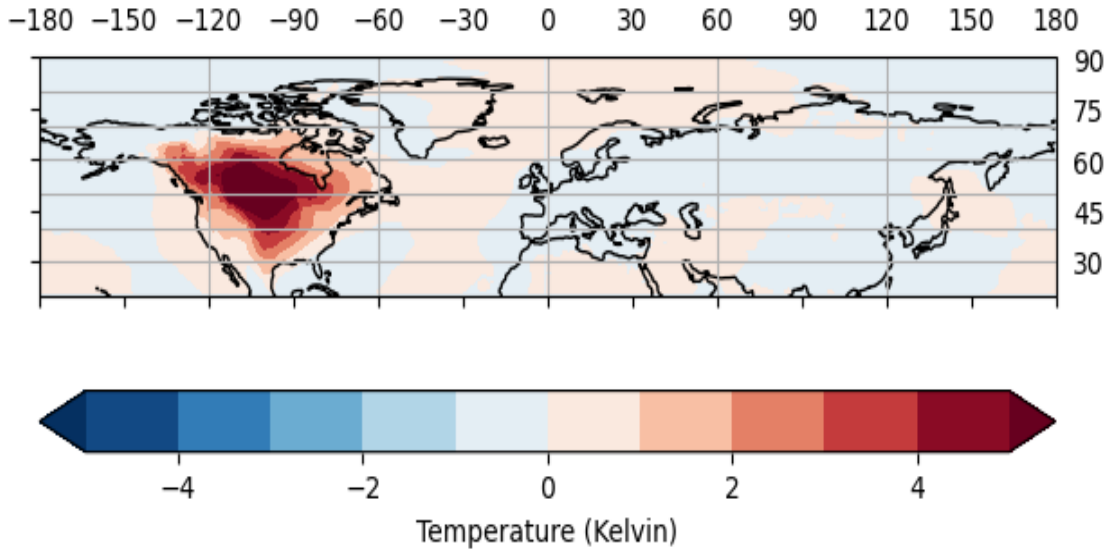

Figure 4: EOF1 spatial pattern (wintertime, 1980-2020).

In EOF1 (Figure 4), the strongest pattern occurs over central North America, indicating that this region contributes most strongly to the leading mode of subseasonal temperature variability. This pattern likely reflects exchanges of air masses between the polar regions and mid-latitude continental interiors. It is also consistent with the Pacific–North American (PNA) teleconnection pattern, where warm anomalies near Alaska are often accompanied by cold anomalies over central North America [22]

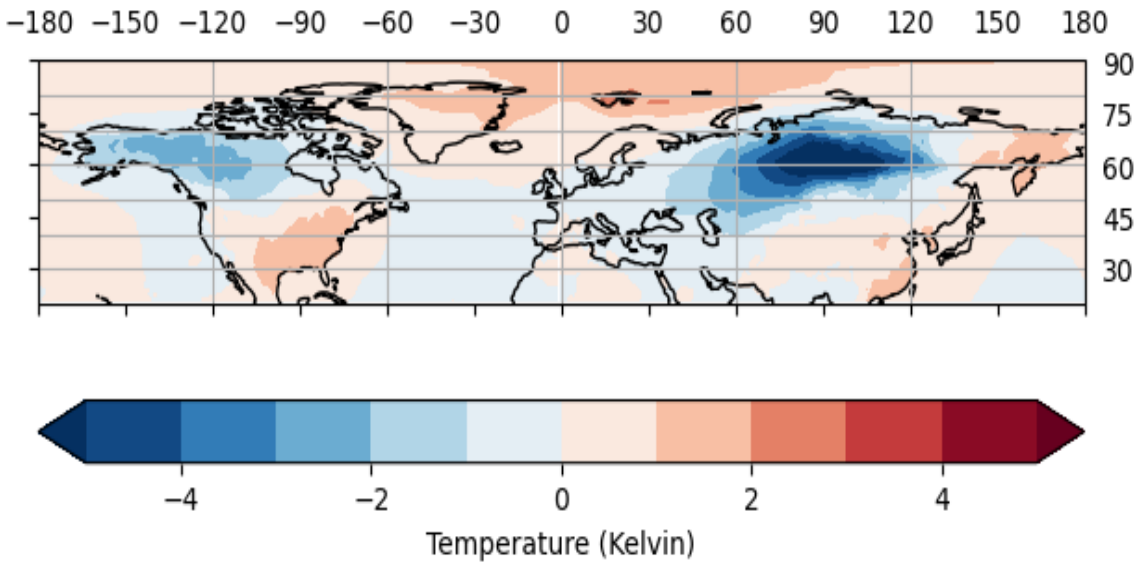

Figure 5: EOF2 spatial pattern (wintertime, 1980-2020).

In EOF2 (Figure 5) we see a dominant negative mode over Russia, a weaker negative mode over the arctic coast of Alaska and northern Canada and positive modes over east America and another around the Arctic. This could suggest a weak dipole temperature structure with one centre over Siberia / Russia and the other with an opposite sign near the Arctic coast, however a true dipole typically the positive and negative modes are the same magnitude which is clearly not the case here. Further studies would be needed to confirm this.

Both EOFs show maximum variability near 50 degrees over continental interiors, with opposite-signed strong anomalies. This could suggest a large-scale shift of air masses between polar / high-latitude regions and mid-latitude continents. This implies that cold spells in North America (EOF1) or Siberia (EOF2) are often accompanied by warm anomalies in adjacent high-latitude regions. This supports the intercontinental connection of subseasonal temperature anomalies [22].

To complement these spatial patterns, Figures 6 and 7 present the time series of Principal Components 1 and 2 (PC1 and PC2), respectively. As expected for standardised anomalies, both series have mean values close to zero, and the maximum amplitude does not exceed four kelvin. These temporal evolutions represent the dominant modes of variability extracted from the dataset, with fluctuations in each component reflecting large-scale dynamical changes in the surface temperature field.

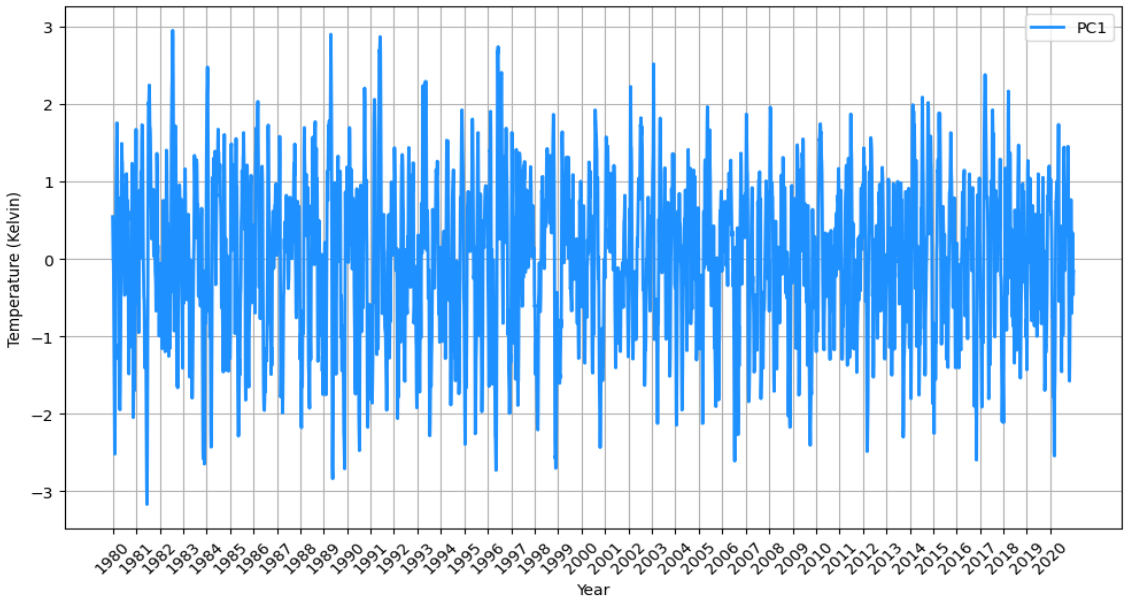

Figure 6: Principal Component One (PC1) (wintertime, 1980 to 2020).

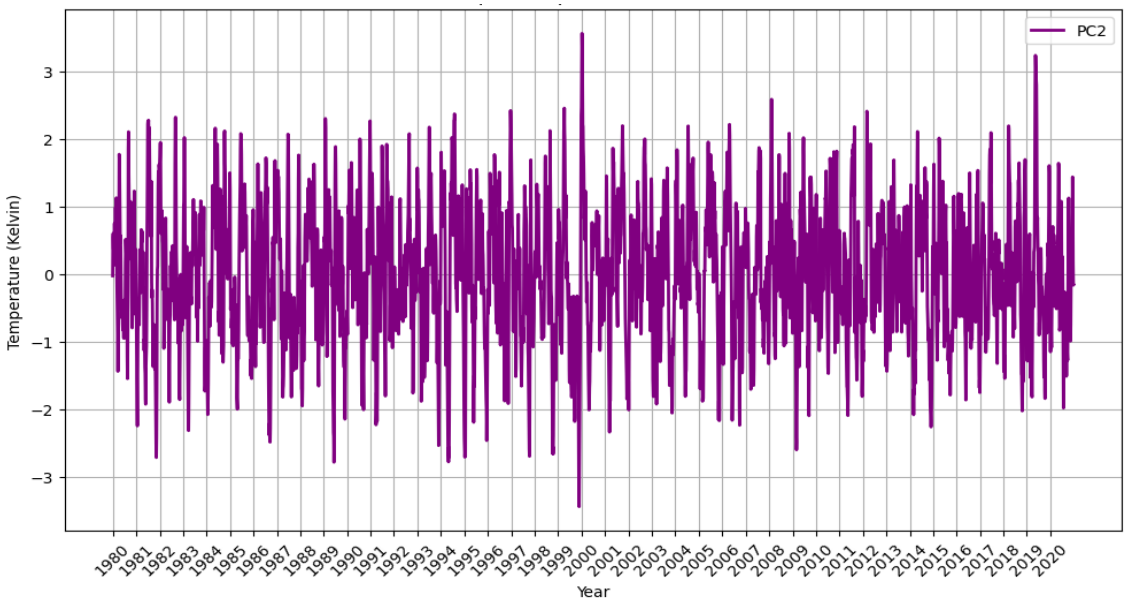

Figure 7: Principal Component Two (PC2) (wintertime, 1980 to 2020).

Having established the dominant patterns, we need to quantify their temporal persistence and interactions to understand how fluctuations in one region may influence others over time; therefore, we turn to auto and cross correlation analyses to examine the lagged relationships within and between the two leading PCs.

4.3 Auto and Cross Correlation

Autocorrelations of the PCs (Figure 8) show that EOF2 is more persistent and changes more slowly, while EOF1 is less persistent and more variable.

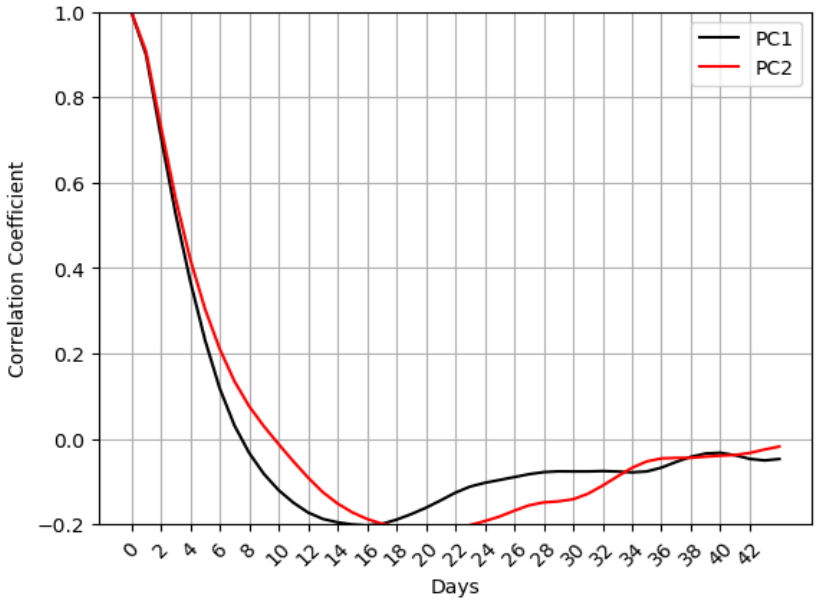

Figure 8: Auto-correlation functions of PC1 and PC2

The cross-correlation (9) between the two EOFs shows a positive peak of 0.5 at a lag of +8 and a negative peak of 0.5 at a lag of +16. This indicates that EOF1 leads the EOF2 by 8 days, with a positive correlation meaning that when the EOF1 is high, EOF2 tends to be high eight days later. The negative peak at lag +16 shows that EOF1 also leads EOF2 by 16 days, but with an inverse relationship, so when EOF1 is high, EOF2 tends to be low 16 days later. The alternating positive and negative peaks suggest an oscillatory or wave-like pattern, with the EOF2 roughly following EOF1 in a repeating cycle of about 16 days. The strength of these correlations indicates a moderate relationship however statistical testing is necessary to assess whether these findings are statistically significant.

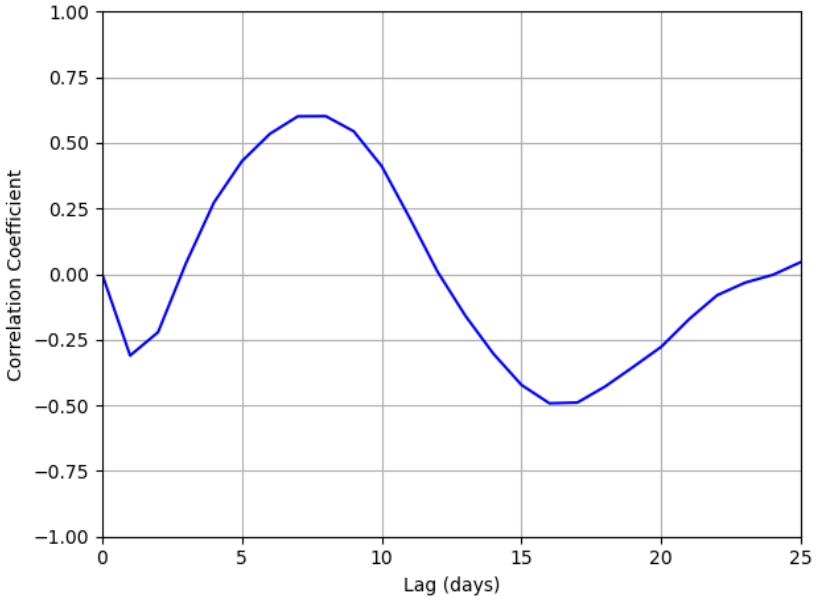

Figure 9: Cross-correlation between PC1 and PC2

Together, the auto-correlation and cross-correlation analyses imply a sub-seasonal temporal interplay between the PCs. The lagged cross-correlation suggests an intercontinental linkage of

temperature anomalies, while the limited auto-correlation persistence points to transient dynamics.

All together, these results provide a quantitative picture of the temporal behaviour and interconnections of the dominant temperature patterns, setting the stage for broader interpretation. In the next section, we bring these findings to assess their implications for subseasonal variability and to frame the conclusions regarding the spatial and temporal coherence of surface temperature anomalies.

5 Conclusions

This study has identified the two dominant modes of sub-seasonal surface air temperature variability in the Northern Hemisphere winter, geographically corresponding to North America and Eurasia. These modes capture a substantial fraction of local temperature variability on subseasonal timescales, consistent with previously documented patterns of atmospheric variability [2]. Analysis of their temporal characteristics and lagged relationships highlights the complex spatiotemporal dynamics underlying temperature anomalies at sub-seasonal scales.

By combining EOF analysis with autocorrelation and cross-correlation analyses, this work provides detailed insight into the persistence and interactions of the leading temperature variability modes. The observed lagged correlations between the PCs reflect physical processes linking temperature anomalies across continents over days to weeks. Our results indicate that EOF1 and EOF2 effectively capture the dominant patterns of sub-seasonal temperature variability over North America and Eurasia, suggesting that accurate prediction of these two modes could account for the majority of the large-scale variability on subseasonal timescales.

In future, statistical testing, such as Student's t-tests, could be applied to verify the significance of lag times between EOF1 and EOF2, while linear regressions of 5-day anomalies could help elucidate the mechanisms driving these dominant temperature patterns.

These modes could be leveraged as forecast targets in Subseasonal-to-Seasonal (S2S) prediction systems, such as the atmosphere–ocean coupled models of ECMWF and NCEP or the atmospheric-only system of ECCC, to assess whether coupling improves prediction skill. The ECMWF and NCEP S2S systems use atmosphere-ocean coupled models [23], whereas the ECCC system uses an only atmospheric model with specified anomalies of sea surface temperature and sea ice [24]. This raises an interesting point about whether the air-sea coupling contributes to the forecast skill of the dominant temperature patterns.

Additionally, given the influence of the MJO on midlatitude weather, further studies could evaluate whether models with enhanced MJO forecast skill also produce more accurate predictions of EOF1 and EOF2, and examine how the initial MJO phase affects their predictability. Although currently the ability of forecasting the MJO varies greatly among forecasting systems. Within the 11 models in the S2S project [25], the ECMWF model was best.

Finally integrating machine learning methods offers promising avenues for capturing the nonlinear and evolving nature of sub-seasonal temperature variability. Approaches such as linear regression, transformer-based models, and deep learning frameworks, including variational autoencoders (VAEs)—could improve low-dimensional representations of spatiotemporal climate

data and enhance forecasting of EOF-based modes [26]. By combining traditional statistical analyses with these advanced machine learning techniques, future work can advance sub-seasonal climate prediction, supporting more effective early warning systems and improved climate risk management.

References

[1] Alberto Troccoli. *Weather & climate services for the energy industry*. Springer Nature, 2018.

[2] E. A. Barnes and J. M. Barnes. "Predicting the Dominant Patterns of Subseasonal Variability of Wintertime Surface Air Temperature in the Extratropical Northern Hemisphere". In: *Geophysical Research Letters* (2018), pp. 12, 983–12, 991. DOI: 10.1029/2018GL077509. URL: https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2018GL077509.

[3] Gilbert Brunet et al. "Collaboration of the weather and climate communities to advance subseasonal-to-seasonal prediction". In: *Bulletin of the American Meteorological Society* (2010).

[4] National Academies of Sciences et al. *Next generation earth system prediction: Strategies for subseasonal to seasonal forecasts*. National Academies Press, 2016.

[5] Roland A Madden and Paul R Julian. "Observations of the 40–50-day tropical oscillation—A review". In: *Monthly weather review* (1994).

[6] DE Waliser et al. "Potential predictability of the Madden–Julian oscillation". In: *Bulletin of the American Meteorological Society* (2003).

[7] Ian T. Jolliffe. *Principal Component Analysis*. Springer, 2002. DOI: 10.1007/b98835. URL: https://link.springer.com/book/10.1007/b98835.

[8] Timothy DelSole and Michael K. Tippett. "Principal Component Analysis". In: *Statistical Methods for Climate Scientists*. Cambridge University Press, 2022. DOI: 10.1017/9781108659055.013. URL: https://doi.org/10.1017/9781108659055.013.

[9] George E. P. Box et al. *Time Series Analysis: Forecasting and Control*. Hoboken, NJ: Wiley, 2016. ISBN: 978-1118675021.

[10] Copernicus Climate Change Service. *ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate*. Copernicus Climate Change Service (C3S).

[11] Unidata. *netCDF4-python Documentation*. Accessed: March 2024. 2024. URL: https://unidata.github.io/netcdf4-python/.

[12] J. P. Peixoto and A. H. Oort. *Physics of Climate*. American Institute of Physics, 1992.

[13] D. L. Hartmann. *Global Physical Climatology*. Elsevier Academic Press, 2016.

[14] Mark C. Serreze and Roger G. Barry. *The Arctic Climate System*. Cambridge University Press, 2009.

[15] Felix Pithan and Thorsten Mauritsen. "Arctic amplification dominated by temperature feedbacks in contemporary climate models". In: *Nature Geoscience* (2014).

[16] James A. Screen, Ian Simmonds, and Clara Deser. "The central role of diminishing sea ice in recent Arctic temperature amplification". In: *Nature* (2010).

[17] Roger G. Barry. *Mountain Weather and Climate*. Cambridge University Press, 2008.

[18] Judah Cohen et al. "Recent Arctic amplification and extreme mid-latitude weather". In: *Nature Geoscience* (2014).

[19] Kevin E. Trenberth. "The definition of El Niño". In: *Bulletin of the American Meteorological Society* (1997).

[20] Wenju Cai, Arnold Sullivan, and Tim Cowan. "Impact of El Niño-Southern Oscillation on the Southern Annular Mode". In: *Journal of Climate* (2011).

[21] Hai Lin. "Predicting the Dominant Patterns of Subseasonal Variability of Wintertime Surface Air Temperature in Extratropical Northern Hemisphere". In: *Geophysical Research Letters* (2018). DOI: https://doi.org/10.1029/2018GL077509. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018GL077509.

[22] C. A. Lin. "The leading intraseasonal variability mode of wintertime surface air temperature over the North American sector". In: *Journal of Climate* (2015). DOI: 10.1175/JCLI-D-20-0096.1.

[23] Suranjana Saha et al. “The NCEP climate forecast system version 2”. In: *Journal of climate* (2014).

[24] Hai Lin et al. “GEPS-based monthly prediction at the Canadian Meteorological Centre”. In: *Monthly Weather Review* (2016).

[25] Frédéric Vitart. “Madden—Julian Oscillation prediction and teleconnections in the S2S database”. In: *Quarterly Journal of the Royal Meteorological Society* (2017).

[26] Yi-Chang Chen et al. “Exploiting a variational auto-encoder to represent the evolution of sudden stratospheric warmings”. In: *Environmental Research: Climate* (2024). DOI: 10.1088/2752-5295/ad3a0d. URL: https://iopscience.iop.org/article/10.1088/2752-5295/ad3a0d.