An analysis of weather systems under certain flow patterns in the atmosphere for our latitudes

BSc Thesis

Written by *Josephine Gondan Kande*

Supervised by

Jens Hesselbjerg Christensen

University of Copenhagen

UNIVERSITY OF COPENHAGEN

NAME OF INSTITUTE:	Niels Bohr Institute
NAME OF DEPARTMENT:	Physics of Ice, Climate and Earth
AUTHOR(S):	Josephine Gondan Kande
EMAIL:	wfx832@alumni.ku.dk
HANDED IN:	20.01.2023

Abstract

The weather in Europe is impacted by the synoptic state of the atmosphere. The atmosphere is a complex system however some stable states can be detected by statistical analysis. This is referred to as modes of variability. NAO and EAO are both modes of the circulation variability. Since the variations in the synoptic scale of the atmospheric circulation impacts the weather it is an increasing scientific filed of interest. To reveal the relationships a study of the connection between the atmospheric circulation variability and the surface weather condition is needed.
This thesis aims to look at the weather system under certain flow patterns in the atmosphere in the North Atlantic region. This project aims to understand how atmospheric flow patterns contributes to precipitation and temperature patterns. The variability in the climate of the North Atlantic region can largely be associated with the NAO (North Atlantic Oscillation). This thesis however will focus on the second leading climate mode in the North Atlantic, the EA (East Atlantic pattern).

Monthly reanalysis data is used to compute the values of the atmospheric variability over the North Atlantic region. This is compared to temperature and precipitation pattern. This paper will examine the meteorological variables and the modes for significant correlation. The approach is an empirical orthogonal function analysis. The technique is used to derive the dominant pattern of variability. This method is good for data reduction, data visualization and feature extraction. I have examined the correlation between atmospheric circulation variability and anomalies in temperature and precipitation with the PCR analysis. The ERA5 data set is used for the time period 1950-2022 with a focus on the extended winter season (NDJFM).

Contents

List of Figures

List of Acronyms

PCA	Principal Component Analysis
EOF	Empirical Orthogonal Functions
SVD	Singular Value Decomposition
NDJFM	November, December, January, February, March
PCR	Principal Component Regression
C3S	Copernicus Climate Change Service
CDO	Climate Data Storage
NAO	North Atlantic Oscillation
EA0	East Atlantic Oscillation
ECMWF	European Centre of Medium-Range Weather Forecasts
msl	Mean sea level pressure
tp	Total precipitation
t2m	Temperature 2 meters above sea or land
gp500	Geopotential height at 500hPa
IL	Icelandic low
AH	Azores High

1 Introduction

Processes in the atmosphere is characterized by variations in both temperature, pressure and density. On a rotating sphere, such as Earth, the variations leads to accelerating strong currents and the friction at the bottom of the atmosphere decelerates them. This relationship is the driver of the variation in the large-scale flow currents and irregular turbulent flows which we experiences as ever changing weather in Earths atmosphere (**?**). Climate however is the changing weather over long periods of time. In a more mathematical sense it is the long-term statistics of weather. *"Climate is what we expect but weather is what we get"* (A. Hannachi, I.T. Jolliffe, D.B. Stephenson, 2007). Naturally climate changes varies in time on different meteorological aspects and influences the Earth system differently. The general weather we experience in Europe is influenced by the varying synoptic scale of the atmosphere for the North Atlantic region. Recent years shows the frequency and intensity of extreme weather events over Europe to be increasingly common (e.g. Heatwaves, floods, storms, droughts, fires) (Buric et al., 2017). Resulting in large consequences for the human world, where we are influenced by the disruption in the climate consequently leading to disruptions in agriculture, water supply, fire risk and so on. Consequently, it becomes more important for the environment and society to be able to predict and plan for high variations in temperature and precipitation over Europe, which leads to environmental events such as floods and droughts.(I, 2007) Variations in the mean sea level pressure and the geopotential height reflects the dynamics happening in the mid and upper level of the atmosphere. The dynamics of the atmosphere are highly interesting since they reveal a possibility of seasonal prediction. The forecast system relies on the ability to correctly simulate the impact of the slowly evolving components of the climate system such as the atmospheric circulation variability. A deeper understanding of the varying synoptic scale weather patterns may reveal the ability to produce seasonal forecasts and thereby become a key feature in the adaption to new extreme weather events.

2 Atmospheric variability

The variability in the climate of the North Atlantic region can largely be associated with the NAO (North Atlantic Oscillation). It is said to be the most important indicator for the climate in the North Atlantic and Western Europe. (G.W.K. Moore, 2012) Shortly explained, the NAO is an oscillation in the atmosphere between a positive and negative phase. It changes the atmospheric air mass with pressure centers at Iceland and at Azores. This climatic pressure state with an Icelandic Low (IL) and Azores High (AH) varies where the positive and negative phases are defined by the gradient between the low and high pressure center. The variation between weak and strong pressure gradient of IL and AH determines the westerly winds across the Atlantic and is the driving factor of the surrounding climate in this region. The positive phase of NAO is characterized by an increasing strength of the IL and AH resulting in stronger westerly winds. As a result, this allows advection of relatively warm and moist ocean water air over Europe resulting in increased temperature and precipitation. However, for the negative phase the pressure gradient between IL and AH is weakened, and thus the westerlies are weakened as a result leading cold air to build up over Europe causing a more dry state of the atmosphere. (for

Environmental Prediction) The Icelandic low at sub-polar latitudes and the Azores high at mid-latitudes are dominant features in the mean North Atlantic pressure pattern. This pattern is semi-permanent pressure systems, which are present at all times during the year. However, the location and intensity changes, effectively affect most of Europe's weather condition. (Barry and Chorley, 2003).

Another less dominant climate mode is known as the East Atlantic oscillation (EAO). Structurally it is similar to the NAO, however it is shifted slightly more towards the south. The EAO pressure pattern is characterized by a single center of action located off the British Island and off the south of Iceland (this definition is based on recent papers definition of the EAO). The monopole in the msl field at approximately (55°N, 20–35°W). This location is along the nodal lines of the NAO and is said to have the potential to influence the location and strength of the NAO. (Harry West, 2022) Just as the NAO the EAO are an important driven factor of the climate over Europe and the general atmospheric circulation. Especially the phase with a strong blocking over the North Atlantic is interesting since it has a large impact on the weather condition. The high pressure blocks the passage of the westerly weather pattern. The positive phase of the EAO pattern is associated with positive anomalies in surface temperatures over Europe annually and in precipitation over northern Europe and Scandinavia. Negative anomalies in precipitation across southern Europe (Team, 2012). This mode of atmospheric variability is less well know compared to the NAO, however the importance of this oscillation during winter season i becoming recognized. The role of the EAO for precipitation and temperature is increasingly being acknowledge. Making the EAO a key feature in the driver for the European climate. (G.W.K. Moore, 2012)

For a deeper understanding of these physical systems and dynamical behaviours we need a tool that can boil down the information to a few dominant modes of variability. The EOF analysis were developed and used in the late 1940s. It provided the researchers with a tool to decompose the large space-time atmospheric data into spatial patterns with associated time indices. (A. Hannachi, I.T. Jolliffe, D.B. Stephenson, 2007)

Using the Empirical Orthogonal Function (EOF) analysis we will derive the modes of variability from monthly mean measurements of the ERA5 data set. This project will relay on msl data to visualise the spatial high and low pressure centers over North Atlantic region.

A lot of studies has relied on the most dominant mode of winter atmospheric variability over north Atlantic region. However, the presented thesis will focus on the second dominant mode, the EAO.

3 Data and methodology

3.1 ERA5 data

This study is based on the ERA5 data set which has been downloaded from the CDS website. ERA5 represents the fifth generation of global climate and weather ECMWF reanalysis from the past decades. (CDS) The data is produced with the data assimilation method which combines model data from the ECMWF with observational data from around the world under the constraints of the laws of physics.

3.1.1 Area and period

The grid is organised as *(time, latitude, longitude)* for each output parameter. All the data is downloaded in subsections of time from 1959 to 1978 and then from 1979 to 2022 and only the extended winter season (November to March). In the NetCDF(experimental) format. The region over which the analysis is conducted covers the North Atlantic region (20-85°N/80°W-40°E). The monthly averaged reanalysis for geopotential at pressure level 500 hPa is specifically downloaded from the **ERA5 monthly averaged data on pressure levels from 1959 to present** and **ERA5 monthly averaged data on pressure levels from 1959 to 1978 (preliminary version)**. This is afterwards converted to geopotential height by dividing with the gravitational acceleration as defined in equation 1. **ERA5 monthly averaged data on single levels from 1959 to present** and **ERA5 monthly averaged data on single levels from 1959 to 1978 (preliminary version)** served as the source of the monthly averaged reanalysis for mean sea level pressure, total precipitation and 2 meter temperature. For the year 2022 only the first three months (January, February and March) was available on the time of the analysis.

3.1.2 Sub-region

The extreme weather anomalies has been found by looking at sub regions with a strong correlation from the PC2 plot. For the two meter temperature a region mainly North and East of Iceland (60-75°North, 30°West-10°East) has been downloaded. For the total precipitation the region over the Southern tip of Greenland (60-75°N, 60-30°W).

3.1.3 Data variables included in the analysis

Geopotential height (gp500) is a definition of the altitude in meters relative to mean sea level of a given pressure surface in the atmosphere. It dictates the variation in the gravitational field over Earth's surface and can change based on the temperature of the air. Warm air is less dense than cold air. Meaning that the pressure surface will be at higher heights for warmer air than for colder air masses. This height measurement accounts for the variations in the gravitational field with latitude and altitude. It is expressed as the following.

$$Z = \frac{\Phi(z)}{g_0} = \frac{1}{g_0}\int_0^z g dz \tag{1}$$

where Z is the geopotential height, g_0 is the globally averaged gravitational acceleration at Earth's surface, z is the geometric elevation height, ϕ is the latitude and g is the acceleration due to gravity and $\Phi(z)$ is the gravitational potential energy per unit mass at a given elevation, called geopotential. The gp500 is commonly used in synoptic meteorology. Charts of gp500 can help identify troughs and ridges in the atmosphere. (John M. Wallace, 2006)

Mean sea level pressure (msl) corresponds to the atmospheric pressure at mean sea level. The unit is pascal. It is the pressure adjusted to the sea level. The atmosphere extends many kilometers above sea level and the weight of all the air masses in a column above a point acts as a force. This force acting on a unit area at a given point is the air pressure. This meteorological variable, as well as the gp500, is a good indicator of the synoptic activity.

Two meter temperature (t2m) measured in kelvin is the temperature two meters above land or sea.

Total precipitation (tp) is the sum of the precipitation that falls to the Earth surface both as rain and snow. It does not include dew, fog or precipitation evaporated in the atmosphere before landing on the surface of Earth. (CDS)

3.2 Data preprocessing

3.2.1 CDO - merging and detrend

For the data analysis I have used the tools of CDO (Climate Data Operators) to merge the netCDF files and obtain a continous data file from 1950 to 2022 for all parameters. Merging the netCDF files with the following command: *cdo -b F64 mergetime *.nc msl_1950-2022.nc*. CDO is also used to detrend the data. By detrending the data we are able to remove the long term trends from the time series. In this particular case we are looking to remove the warming trend from the variables influenced by global warming (e.g. gp500 and t2m). The command is: *cdo detrend ifile ofile* (CDO climate Data Operators GUIDE). All the coding is done in Python.

3.2.2 Five-year-running mean and extended winter seasonal mean

We wish to calculate a mean containing the 5 months of the extended winter season (NDJFM). This is done such that November and December for the previous year is included as a part of the 5 months mean. Resulting in 72 time steps. Five-year-running mean also called the moving average is calculated based on the extended winter seasonal mean. The preceding two variables plus, the two following and the year itself, divided with the size window. This method smoothes out the data based on the variations and looking at changes over a decade resolution allowing the ability to compare with proxy data with no yearly resolution.

3.3 Empirical Orthogonal Function analysis

The Empirical Orthogonal Function analysis, interchangeable with Principal Component Analysis, is a technique used to derive the dominant pattern of variability from a statistical field. In geophysics the term EOF are more often used which corresponds to the weighted PCAs. Data covering large latitudinal ranges needs to be geographical weighting to account for grid sizes covering a much smaller area at the poles than those at the equator. Therefore the weighting factor, equation 2, is applied to compensate for this decreasing area of grid point towards the pole and ensure equal contribution.

$$w = \sqrt{|cos(lat_{radians})|} \tag{2}$$

In this thesis the EOF refers to the patterns and PCs for the time series. The EOF is a map-series method. It changes all the variability in the time evolving field and turns it into standing oscillations, with each standing oscillation a corresponding time series is calculated. Each standing oscillation presents a mode of variability. The time series, PCs, of the mode refers to the change in the oscillation mode change in time. This way of separating the space-time field into a space display and a time display allows us to filter out small-scale noise with a minimal loss of information. This method seek to reduce the data matrix of some size $n \times p$ stored with information and decompose to a few modes that describes the different modes of the field. (A. Hannachi, I.T. Jolliffe, D.B. Stephenson, 2007) (Stephenson and Benestad)

The modes we are presented with from the EOF analysis are primarily statistical modes and not necessarily physical modes. The physical mode is established when a subjective interpretation is made from the EOF pattern. The pattern of the EOFs are dimensionless and the sign is arbitrary. When the EOF and the PC is put together the output is the original data (eg. temperature relative to time). For this study we use EOF analysis to examine variability of the scalar fields.

A step by step approach for the EOF is given in the appendix following closely the outline given in (Bjornsson and Venegas, 1997):

The goal of the EOF method is to find a set of decomposes the information into aligned eigenvectors c_i that maximizes the projection of the row vector of the map matrix x_n on the basis vectors. In terms of modes, the EOF analysis finds the leading EOF patterns that maximises the variance. Solving the eigenvalue problem for covariance matrix, R, will result in the eigenvectors and eigenvalues. Eigenvectors is the EOFs and the eigenvalues is the explained variance. Some of the eigenvalues, λ_i, are more dominant than others. This is why we can use only some of the EOFs and only a few basis vectors is needed to explain the behavior of the data matrix. (Bjornsson and Venegas, 1997)

3.3.1 Eigenvalue spectrum

The field X at time t can be written:

$$X(t) = \sum_{i=1}^{r} \lambda_i a_{ti} \mathbf{u}_i \tag{3}$$

where λ_i is the eigenvalue, a_{ti} represents the PC, $\mathbf{u}_i$ is the EOF, r is the rank of the field matrix and n is the number of eigenvaules. This form is especially useful when decomposing the data into smaller dimensions. By taking the sum for the first n terms, where n is much smaller than the rank of X. The choice of n, meaning the choice of how many EOF is sufficient, is not fixed. It is however good to base the choice of n, such that the n leading EOFs altogether explains about 80 % of the variance. Choosing the n leading EOFs can be done on the background of the eigenvalue spectrum. The eigenvalue spectrum is a particularly important property. Two eigenvalues can be indistinguishable if they are within each other uncertainties. To determine the uncertainty of the eigenvalues, we use North rule of thumb

$$\Delta\lambda_i^2 \sim \lambda_i^2\sqrt{\frac{2}{n^*}} \tag{4}$$

$$\Delta\mathbf{u}_i \sim \frac{\Delta\lambda_i^2}{\lambda_j^2 - \lambda_i^2}\mathbf{u}_j \tag{5}$$

where n^* is the number of independent observations in the sample and λ_j^2 is the closet eigenvalue to λ_i^2. Most of the behavior can be described by the first leading EOFs. All the other eigenvectors besides the first do not contributes as much. (Storch and Zwiers)

3.3.2 Singular Value Decomposition

The field matrix, X, is given by into a matrix with

$$X = \Phi D C^t \tag{6}$$

where Φ is the normalized times series, ϕ_j is the column vector of Φ, normalizing the PCs, $\vec{a}$, by: $\phi\sqrt{\lambda_j}$), D a diagonal matrix, with the square rot of eigenvalues on the diagonal, $\sqrt{\lambda_j}$ and a matrix of eigenvectors, C.

In general the The Singular value decomposition (SVD) decomposes the matrix X of $n \times p$ into the form:

$$X = U\Gamma V^t \tag{7}$$

where U is a $n \times n$ orthonormal matrix, V is a $p \times p$ orthonormal matrix and Γ is a diagonal $n \times p$ with k elements on the diagonal. It follow that $\Gamma_{ij} = \delta_{ij}\gamma_{ij}$ for $i = 1, k$ with $k \leq min(n,p)$. where the numbers on the diagonal, γ_{ij} is the singular values of the matrix and the columns of U and V are the singular vectors of X.

The two method are very similar and the relationship between SVD and EOF is that the eigenvalues and the singular value are related as:

$$\lambda_i = \gamma^2_{(i,i)} \tag{8}$$

In this project the analysis is based on the *eof* package in python which uses the SVD method since it has computational advantages. (Bjornsson and Venegas, 1997)

3.4 Principal Component Regression analysis

This analysis seeks to determine the correlations/connections between surface weather condition and the modes of variability. From fluctuations in the dependent variable the behavior if the independent variable can be described with a least-squares fit of a polynomial of first dimension. We use this linear regression analysis technique to estimate the regression coefficient in a standard linear regression model. This is different from normal linear regression as instead of using the independent variable directly as the regressor, the principal component is used.

Firstly a EOF analysis is done on the data to reduce the dimension of the data as described in the earlier section. From the EOF analysis we obtain a presentation of the variance of the data in a space with less dimensions as originally. Secondly a correlation maps is calculated. Furthermore a regression analysis based on the PC output from the EOF analysis.

3.4.1 Correlation

There will be a limited amount of, $n_{(eofs)}$, independent numbers for the EOF analysis found from the eigenvalue spectrum. The correlation is found by using the $n_{(eofs)}$ PCs, weighted by the EOF patterns and their variance. We use this correlation method to compare how two fields, $Y = [\vec{y_1}, \vec{y_2}, \vec{y_3}]$ and $X = [\vec{x_1}, \vec{x_2}, \vec{x_3}]$, couples. Each Field contains several observations sampled respectively at two locations, p and q over a time period, t, with m measurements at each location.

The correlation matrix allows us to identify the variables with high degree of correlation. Pearson correlation coefficient, r, is used to determine the linear association between two variables and thereby the strength of the correlation. The correlation coefficient, r, ranges from -1 to 1 with stronger correlation closer to 1 and -1 and closer to zero a weaker relation. The correlation refers to the statistical relationship between the two variables of interest.

3.4.2 Regression

The regression is done to estimate and make conclusions about correlation coefficients. We have measured to variables as a function of time $y(t)$ and $x(t)$. For a single predictor, we assume a linear fluctuation in $x(t)$ will predict fluctuations in $y(t)$.

$$\hat{y}(t) = a_1 x(t) + a_0 \tag{9}$$

where $\hat{y}(t)$ is the linear least square fit of $y(t)$ to $x(t)$. The slope, a_1 of the linear least square fit provides the estimate on how $y(t)$ changes as a linear function of *x(t)*. This is what we call the linear regression coefficient, R, the fraction of variance in $y(t)$ explained by the $\hat{y}(t)$ is the linear regression coefficient. The total variability in $y(t)$ can be described by the linear relationship with $x(t)$. The amount of the total variability, that can be described with this method is determined with how the data points is located around the $\hat{y}(t)$ line. If a large fraction of the data points is located closely around the $\hat{y}(t)$ line, then a large fraction of the variability in $y(t)$ is predicted by $\hat{y}(t)$. If the data points are widely spread out only a small fraction of the variability in $y(t)$ is predicted by $\hat{y}(t)$.

The error function:

$$Q = \sum_{i=1}^{N} (\hat{y}(i) - y(i))^2 \tag{10}$$

The error function describes the quantitative difference between the estimates of $\hat{y}(t)$ and the actual values of $y(t)$. N is the time steps in $y(t)$ and $x(t)$. Squaring the difference in the error function ensures it is a linear problem. The error will never be negative and the average error will not be close to zero.

Finding the best fit of $\hat{y}(t)$ for the data is done by this linear least square fit, which as indicated in the name means finding where the error is at a minimum. With some calculations we can find that the slope of $\hat{y}(t)$ is given as

$$a_1 = \frac{\overrightarrow{x'y'}}{\overrightarrow{x'^2}} \tag{11}$$

where the numerator is the covariance of $x(t)$ and $y(t)$ and the denominator as the variance of $x(t)$.

The linear regression is a dimensional measurement of the linear relationship between the dependent variable, x and independent variable, y. How does y change with one unit of x. Each coefficient indicates the direction of the relationship between the predictor variable and the response variable. (Storch and Zwiers) (han)

4 Results

We are interested in analysing the extended winter season going from November to March for the North Atlantic region. In this section the result of the EOF analysis applied to the monthly ERA5 data for the extended winter season (NDJFM) is presented. Moreover, the same analysis is done for the extended winter seasonal mean and a five-year-running mean is calculated. An identification of the teleconnection between the surface conditions (eg. temperature and precipitation) and the governing atmospheric pattern from msl is presented with the PCR analysis. Lastly, the extreme conditions are identified from the sub-region and the EOF and PCR analysis are applied to these warm, cold, wet and dry quartiles of PC2.

4.1 EOF analysis

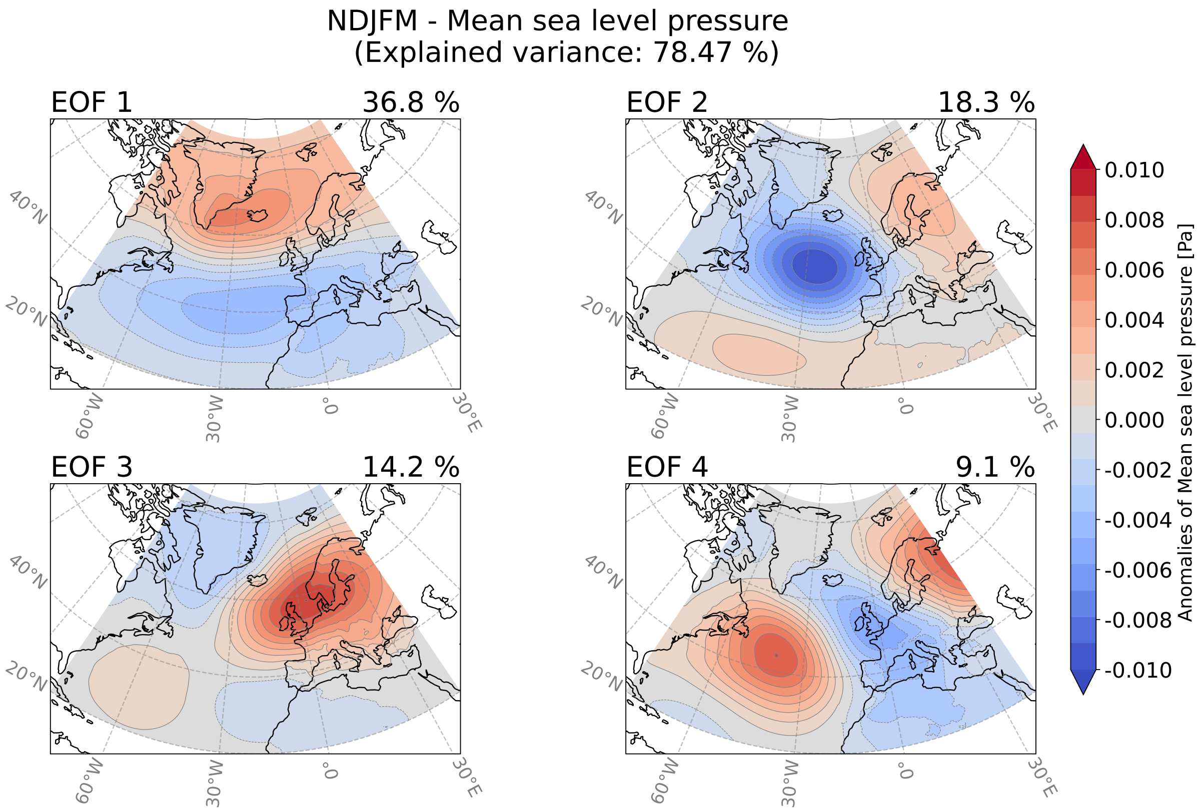
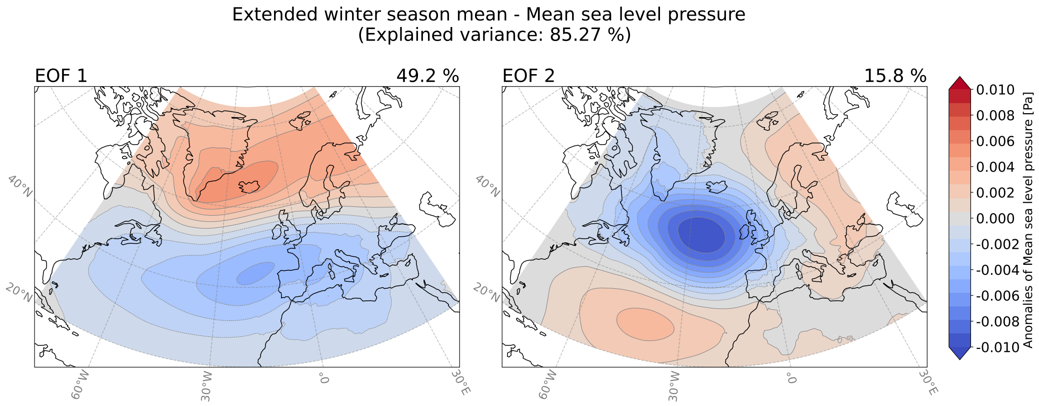


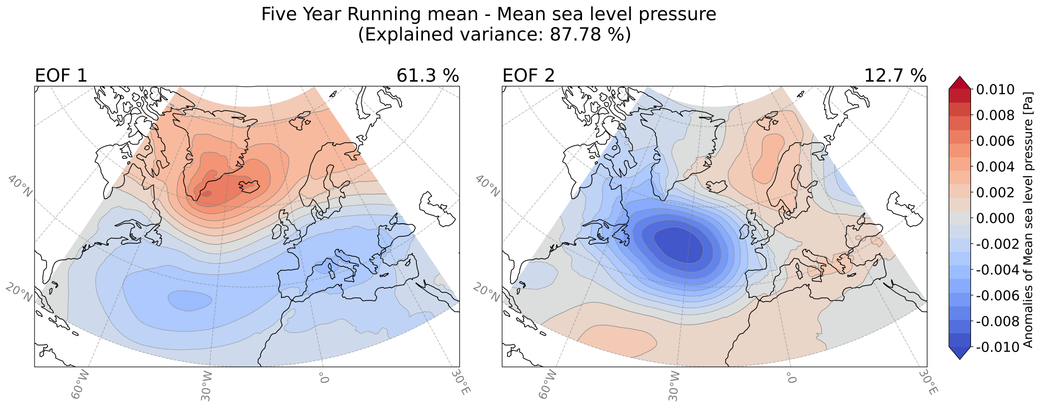
Figure 1: The four leading EOF for msl over the extended winter season (NDJFM) for the time period 1950-2022 covering the North Atlantic region. The total explained variance is specified in the header of the figure. Each individual EOFs explained variance is shown in the upper-right corner of the EOF sub-plot.

Presented on figure 1 is the pattern of the four leading EOFs for the msl over the extended winter season (NDJFM) with a total explained variance of 78.47%. The first EOF pattern obtained from the EOF analysis presented on figure 1 shows two centers of action with opposite signs. One located between Southern Greenland and Iceland and the other closely around Azores. The explained variance for this EOF is 36.8 %. The second leading mode derived from the EOF consists of one center located south of Iceland. This mode accounts for 18.3 % of the total explained variance in the msl data. EOF3 represents a quadruple pole with a strong center over Europe. The fourth EOF shows a center between two opposite sign centers located largely over the North Atlantic Ocean and Europe. The last two EOF patterns accounts for less explained variance than the first two. The further analysis presented in this thesis will include the first two leading EOF patterns with a focus towards the second dominant mode. Figures with all four modes are included in the appendix. A comparison with the explained variance for the EOFs of the gp500 showed that the EOF analysis of the msl data have a higher explained variance, meaning it contains more information. Thus, the further analysis will exclusively be based on the msl data and it will work as the variable for which this project will study the activity on large-scale atmospheric circulation.

4.2 Extended winter seasonal mean and 5 year running mean



(a) Extended Winter Seasonal mean



(b) five-year-running mean

Figure 2: The two leading EOF for the mean sea level pressure comparing the extended winter seasonal mean and the five-year running mean for the time period 1950-2022 for the North Atlantic region.

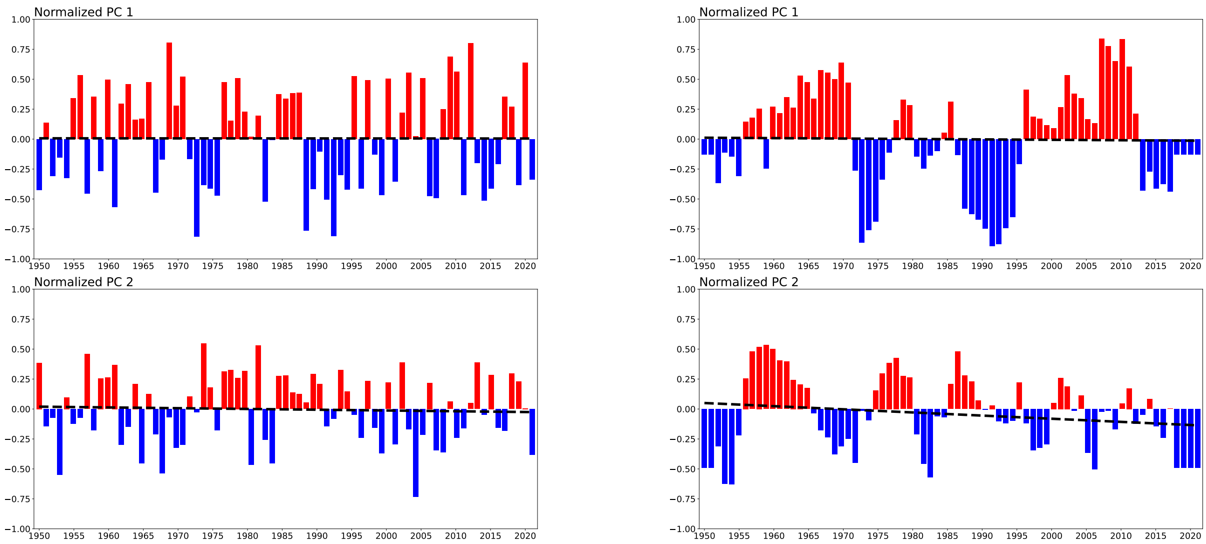
On figure 2 the two leading EOF for (a) the extended winter seasonal mean and (b) the five-year-running mean is shown for the msl data over the time period 1950-2022 covering the North Atlantic. The total explained variance is higher for both the 5 year running mean and the extended winter seasonal mean compared to the raw winter months (NDJFM) on figure 1. The first EOFs on figure 2 (a) and (b) accounts for a significantly higher variance than for the raw extended winter season data (NDJFM) on figure 1. Whereas the second leading EOF has a generally lower explained variance compared to EOF2 on figure 1. EOF2 of the extended winter seasonal mean accounts for more variance than EOF2 of the 5 year running mean and only sligthly less explained variance compared to EOF2 for the raw data on figure 1.

The spatial pattern of the first two leading EOFs are consistent for all three plots. The 5 year running mean however has the center of action shifted a bit westwards in both EOF1 and EOF2. Whereas the extended winter seasonal mean and raw extended winter months (NDJFM) show the same pattern.

4.3 Associated PCs

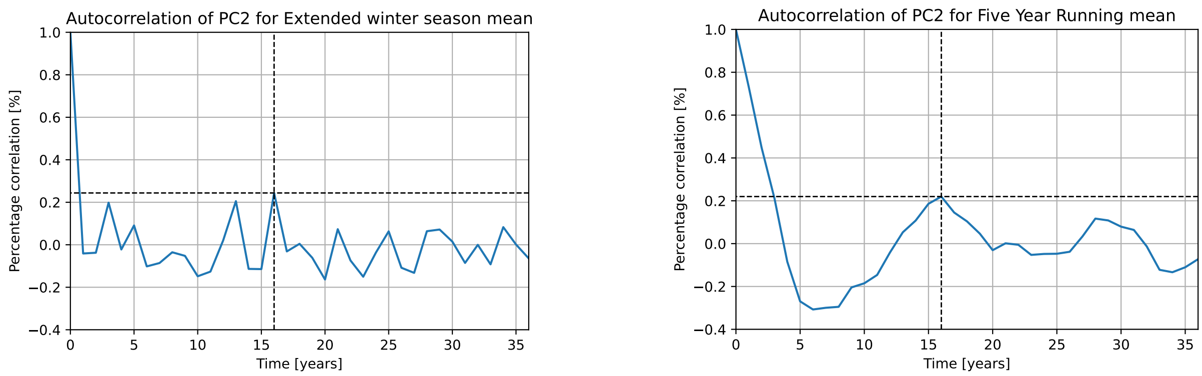
Figure 3 represents the normalized PCs for the extended winter seasonal mean and the five-year-running mean, meaning the time evolution of the two leading EOFs over a seaonal mean and a decadal mean. The PCs represents the index of the associated EOF and shows the negative and positive phase of the respective pattern. The PCs indicates the relative loading for each month of each EOF over the time period analysed. The extended winter seasonal mean show some periodic behavior with many of the negative phases being grouped together. This is more apparent in the five-year-running mean where we see a clear periodic pattern. The period appears longer in the PC1 compared to PC2. The dashed line is a linear fit. Only for PC2 of the five-year-running mean do we see a trend in the data. Furthermore, the trend is only apparent in the positive variations.

Figure 4 shows the auto-correlation of the secondary PCs. For the extended winter seasonal mean we see two peaks around 14 and 16 years and similarly for the five-year-running mean, however this peak is better defined. The indication of a period of $\approx$ 16 years is further promoted by the smaller second peak around 30 years.



(a) Extended Winter Seasonal mean (b) Five-Year-Running mean

Figure 3: The normalized PCs for the respective EOFs for the extended winter seasonal mean and five-year-running mean for the time period 1950-2022 for the North Atlantic region. The dashed black lines show the trend of the index.

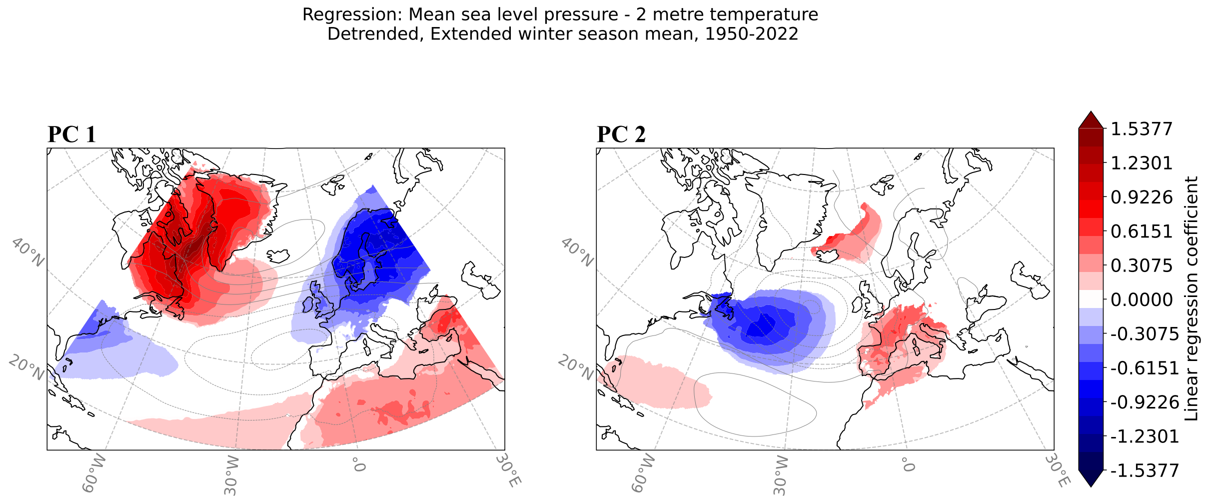


(a) Extended Winter Seasonal mean (b) Five-Year-Running mean

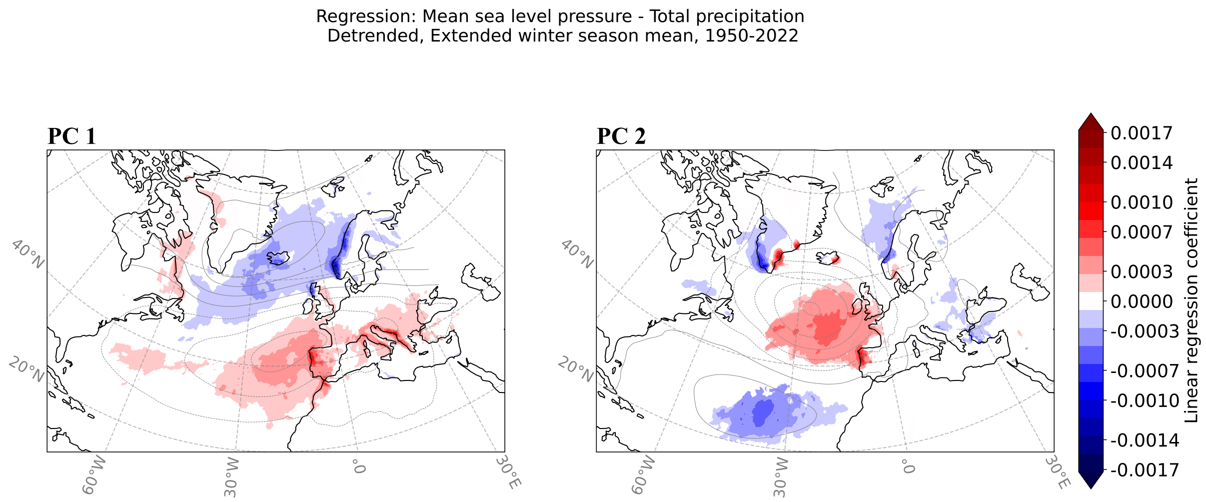
Figure 4: The auto-correlation for the second leading PCs for both the extended seasonal mean and the five-year-running mean. The correlation is calculated with $c_k = \sum_n c_{n+k} \cdot \overline{c_n}$ with c being the PCs.

4.4 PCR analysis

This regression analysis seeks to determine the correlations between surface weather conditions and the modes of variability. The goal of doing PCR analysis is to predict the behavior of one variable (the independent variable) based on fluctuations in one or more related variables (the dependent variable). On figure 5 the PCs of the mean sea level pressure serves as the independent variable. The meteorological variables temperature and precipitation are used as the dependent variables respectively. Focusing on the second PC, the regression coefficient for the msl and temperature regression plot, figure 5 (a), exceeds 0.5 over Western Europe and over Northeast of Iceland indicating that some amount of the two meter temperature variability in these regions during the extended winter season can be explained by the EOF2 pattern. On figure 5 (b), the total precipitation anomalies shows strong positive regression at the Southeast tip of Greenland. These two observations are used to determine a sub-region for identifying years with extreme weather conditions.



(a) Two meter temperature



(b) Total precipitation

Figure 5: PCR result for the msl and a) two meter temperature anomalies and b) precipitation anomalies for the extended seasonal mean from 1950 to 2022 covering the North Atlantic region. Only results associated with a 0.3 correlation is shown.

The correlation map and regression map for the msl over the extended winter season (NDJFM) including the first four EOFs can be found in the appendix in figure 9 and 10.

4.5 Precipitation and temperature quartiles

From the correlation and regression maps on figures 9 and 10, a sub-region is defined. This analysis focuses on the second mode, thus the second PC is in focus. The sub-region is found as the region, where we see a strong correlation between the msl and variable of interest. This is done for both temperature and precipitation identifying the four quartiles. The quartiles is found by the 25th and 75th percentile of the temperature and precipitation data. The plots for this approach is placed in the appendix. The analysis is based on the extended winter seasonal mean to obtain results based on the winter season as a whole.

4.5.1 The second leading EOFs for each quartile

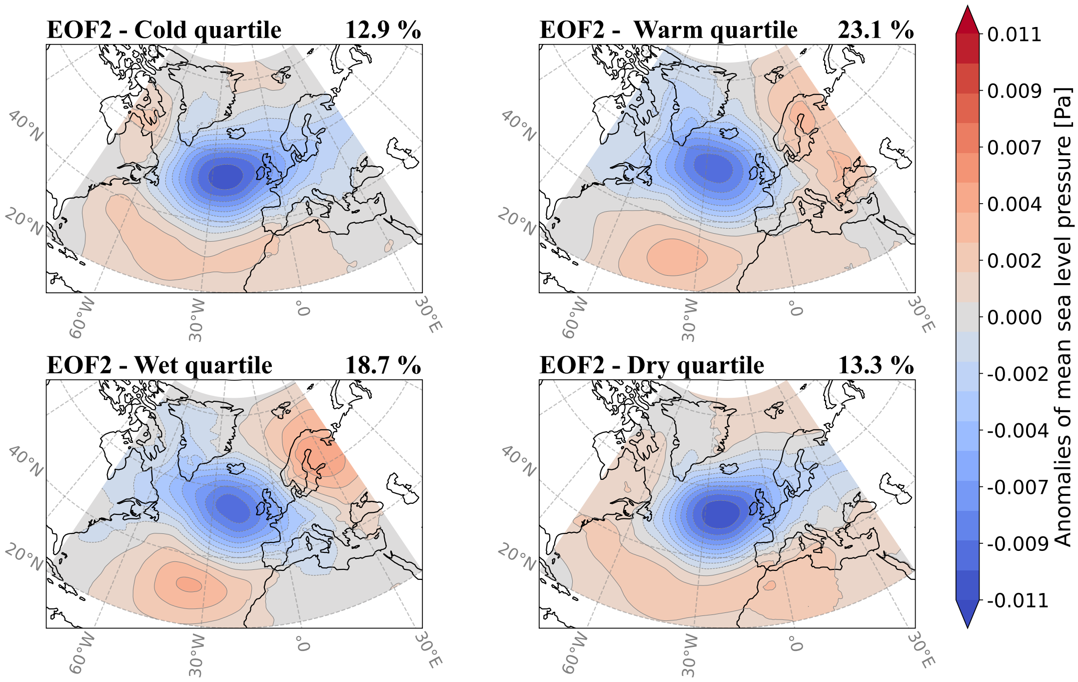
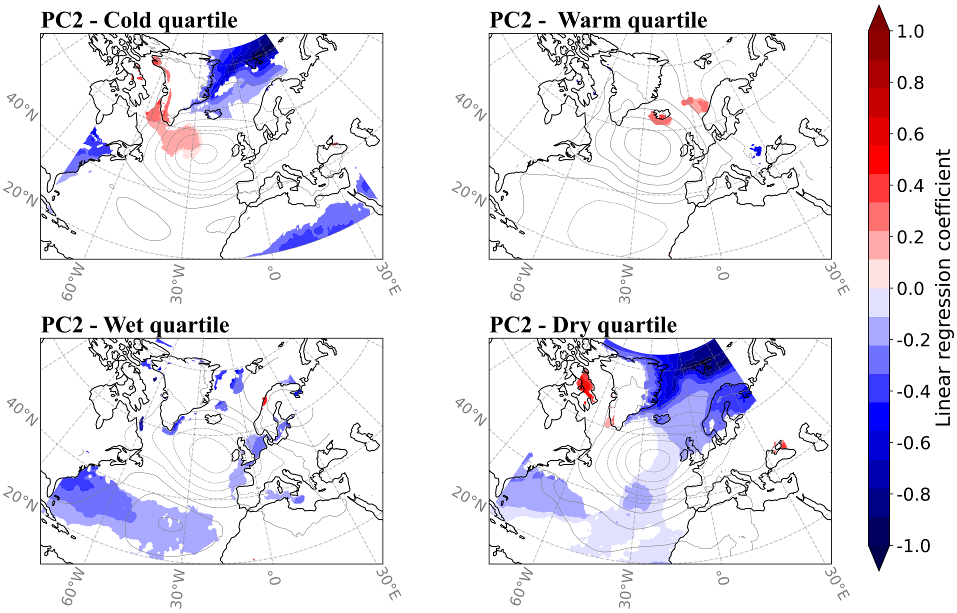


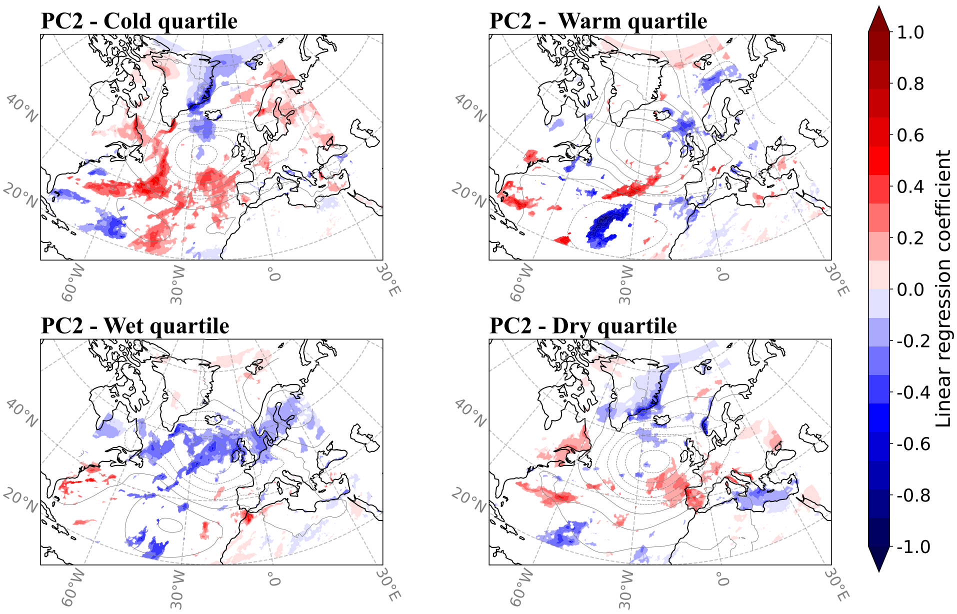
Figure 6: The second leading EOF for msl for each quartile over the extended winter seasonal mean for the time period 1950-2022 of the North Atlantic region.

All four leading EOFs for the precipitation and temperature quartiles is displayed on figure 14 in the appendix. Figure 6 shows the spatial pattern of the second leading EOF for all temperature and precipitation quartiles with a center of high negative anomalies at location (55°N, 25-30°W). The cold and dry quatiles features an opposite sign anomalie band south of the monopole center of action. Stretching between (20-40°N, 65-10°W) for the cold quartile and (20-45vN, 80°W-15°E) for the dry quartile. The warm and wet quartile has similar patterns with a tripole-looking structure. The negative center of action is nearly equal to the one shown in the aforementioned quartiles, however with a weaker amplitude and having more resemblance with the pattern shown on figure 2 and 1. The warm quartile has the largest explained variance and the wet quartiles the second largest. The two other quartiles, dry and cold has approximately the same explained variance at a significantly lower value.

4.5.2 PCR analysis



(a) Temperature pattern



(b) Precipitation pattern

Figure 7: PCR results of PC2 for msl and a) temperature anomalies b) precipitation anomalies for each temperature and precipitation quartile in the North Atlantic region for the time period 1950-2022. The grey contour lines indicate the EOF2 pattern. Only results associated with a 0.3 correlation is shown.

On figure 7 the warm quartile shows high positive PCR results for the temperature anomalies around Southern Iceland and for the total precipitation scattered around the Southeast of the North Atlantic region. The wet quartiles has a negative regression field located at

20-40N, 25-75W. For precipitation the wet quartile shows a negative filed of PCR around Northern Europe. The warm years show significantly few regression fields for both precipitation and temperature. However the temperatures anomalies over Iceland shows a high positive regression coefficient. Cold and dry quartile show generally fields of regression around Greenland and Arctic for both temperature and precipitation anomalies.

5 Discussion

5.1 EOF analysis

5.1.1 Eigenvalue spectrum

On figure 8 the eigenvalue spectrum is calculated. From this figure, we can read how much of the variance each eigenvalue explains. Each eigenvalue belongs to each EOF pattern. For EOFs with very close neighboring eigenvalues the explained variance it contains can be randomly mixed between the overlapping eigenvalues and thus making it hard to distinguish each mode. This means, the interpretation can then be misleading. This phenomena is referred to as degeneracy, when we can no longer interpret the eigenvalues as an independent pattern with a physical mode. Since the first four eigenvalues are well separated and non degenerated, the first four eigenvalues are reasonable to use in an analysis. This thesis will however mainly focus on the first two. In the appendix figures that contains all four EOF patterns are shown.

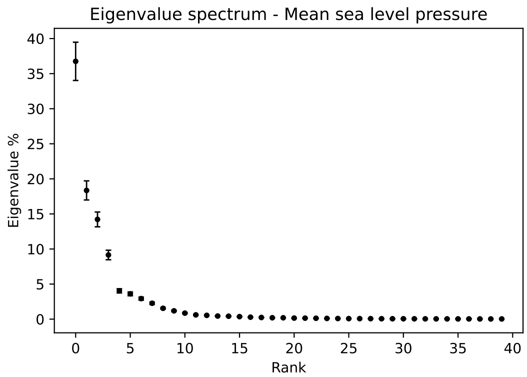


Figure 8: Eigenvalue spectrum for the first 40 eigenvalues together with each eigenvalues uncertainty, given by eq. 5

The EOF analysis provides us with a tool to understand and establish the base knowledge of the mode of variability over the North Atlantic. However the EOF analysis is purely a mathematical method for detecting the dominant spatial pattern with high explained variance in the data sets. In the following a physical interpretation of the EOF patterns presented in the results is given to connect the statistical result from the analysis with physical knowledge of atmospheric modes of variability.

5.1.2 Mean sea level pressure

On both figures the first EOF show the familiar pattern with the IL and AH which is the characteristics of the NAO pattern. The second EOF is recongnised as the EAO with the characteristic pressure center south of Iceland at the North Atlantic Ridge. Since the EOF showing the EAO pattern explains about 20 percent of the variance in the extended winter season, it corresponds to one out of the 5 months of the extended winter season being impacted by the EAO. With the North Atlantic ridge blocking features from the EAO pattern, it can then alter the westerly flow and contribute to the change in the surface weather conditions.

The PCs indexes characterize the temporal variability of the atmospheric modes, meaning the variation over time. For the PCs in figure 3 we only see a periodic oscillation in the five-year-running mean. Looking at their autocorrelation shown on figure 4 the period of each phase is 16 maybe 13 years. On figure 3 only the five-year-running mean for PC2 shows a decreasingtrend in the index. However, to really make meaning out of the PC we need to interpret the EOFs and PCs together, which gives back the original signal.

5.1.3 Extended winter seasonal mean and 5 year running mean

The extended winter seasonal mean is used for the data analysis as it has a higher explained value than the EOF analysis for the raw winter months (NDJFM). The five-year-running mean is marginally higher in the total explained variance. However, most of the explained variance is accounted for by the first EOF. The second EOF explains significantly less than for the raw winter months and also less than the extended winter seasonal mean. In the appendix the third and fourth EOF for both means is displayed and it can be seen that the five-year-running mean shifts the order of the two mentioned EOFs compared to both the EOF of the raw winter months and the extended winter seasonal mean. The five-year-running mean has a larger explained variance for the first EOF1 implying it accounts for a decadal variability. The second EOF has significantly less explained variance implying it is variability of the annually winter seasons. The second leading pattern lose significance when using a longer timescale, since the explained variance drops.

Both the Extended seasonal mean, the raw winter months and the five-year-running mean reveals high explained variance in agreement with North's rule of thumb for feature extraction corresponding to physical modes.

5.1.4 Quartiles

The analysis for the temperature and precipitation quartiles is done to reveal the importance of the EAO in the years with a specific characteristic for the winter season (e.g. warm, cold, wet, dry). The cold quartile accounts for explained variance of 12.9% shown in figure 14. This result is implying that the EAO pattern might be of less importance in years with a cold winter. The same argument applies for the dry years. However the high explained variance for the warm quartile does show that years with positive temperature anomalies might be impacted by the EAO. Also the wet years show high explained variance for EOF2.

5.2 PCR analysis

Focusing on only the second mode reveals how the EAO pattern contributes to the anomalies of precipitation and temperature in the data over the North Atlantic from 1950-2022.

The regression value indicates the direction of the correlation between the surface condition at a certain location and the respective mode of variability. The regression coefficient is associated with the variation corresponding to one standard deviation along the respective PC for each grid point. We can therefor use it to detect the value of R, which indicates how the data of either precipitation or temperature is spread around the regression line in terms of sigma. For a high value of R around the direction of the PC2 there will be large anomalies. This reveal if there is a correlation between the anomalies observed and the EOF pattern. Displaying the strong positive and negative connections between the EAO index and the surface condition for the variable of interest, the results of this study shows strong connection at several regions on figure 5. Comparing with the EOF2 pattern shown as the grey contour lines on the figure, the total precipitation has high correlation Southeast of the EAOs characteristic center of action, and for temperature there is a strong negative regression Southwest of the center of action.

The results from the PCR analysis on figure 5 (b) indicates that there is a positive correlation between msl and temperature anomalies Northeast of Iceland. On figure 5 (a) the EAO shows strong correlation with the precipitation anomalies at the Southern tip of Greenland. These two regions serves as sub-regions for a further analysis of the winter with specific characteristics (eg wet, dry, warm, cold). The results is shown on figure 7. The result deviates significantly from the obtained results over the whole full period of time seen on figure 5. Possible explanations could be a mistake in the coding or since a large fraction of the temperature correlates around Arctic it might be relevant to look at sea ice variations or similar arctic dynamics.

5.3 Further work

Further work could look into other meteorological parameters such as wind speed to say something about the jet stream and storm-tracks, how the jet stream contributes to the climate formation and looking at how the two leading climate signals NAO and EAO impacts the jet stream. furthermore, looking into how the modes of variability contributes to the position and intensity of the jet stream. Another possible interest would be looking at more than one variable as the dependent correlation in the PCR analysis ie. how does the msl correlate with both temperature and precipitation anomalies together. Other possibilities are looking at the combined influence for EAO and NAO index for a more complete picture of the synoptic scale variations, combining effects of the NAO and EAO on the anomalies of meteorological variables.

6 Conclusion

This thesis presents the results of an EOF analysis for the msl in the North Atlantic region over the time period 1950-2022 focusing on the extended winter season, showing the four leading modes of variability.

The main purpose of the thesis is to examine the correlation between the modes of variability and the surface weather condition. This was obtained with a PCR analysis. Additionally looking at the extreme anomalies to gain knowledge of the impact of the climatic modes for the extreme winter seasons.

The EOF analysis showed explained variance around 20% for the the EOF2 indicating that the EAO is an important feature of the atmospheric circulation at least one out of five months during the extended winter season. The key feature explaining the anomalies seen during the winter season over the north Atlantic was shown with the PCR analysis revealing correlations between the EAO and the surface weather conditions. Especially the region over the North Atlantic ocean and Western Europe showed significant correlation patterns. Also during the analysis a 13-16 year periodic behaviour for the EAO was seen in the five-year-running mean.

The results shown in this project serves as a contribution to the understanding of the relation between large-scale atmospheric variability and anomalies in the meteorological variables over the North Atlantic. Concluding EAO as an additional but weaker driver compared to the NAO, allowing potentially better prediction for weather events and seasonal forecasting in the future.

References

Website to the climate data store. URL `https://cds.climate.copernicus.eu/cdsapp!/dataset/reanalysis-era5-single-levels-monthly-means-preliminary-back-extension?tab=overview`. Last accessed 19 January 2023.

Statistics review handout. URL `https://www.atmos.colostate.edu/ davet/at750/StatisticsReviewHandout.pdf`. Last accessed 3 January 2023.

A. Hannachi, I.T. Jolliffe, D.B. Stephenson. Empirical orthogonal functions and related techniques in atmospheric science: A review. *International Journal of Climatology*, 27 (9):1119–1152, 2007. doi: https://doi.org/10.1002/joc.1499.

Roger G. Barry and Richard J. Chorley. *Atmosphere, Weather and Climate*. Routledge, 29 West 35th Street, New York, NY 10001, 8 edition, 2003.

H Bjornsson and SA Venegas. A manual for eof and svd analyses of climatic data. *CCGCR Report*, 97(1):112–134, 1997.

Dragan Buric, Jovan Dragojlović, Milan Milenkovic, Ljubomir Popovic, and Miroslav Doderović. Influence of variability of the east atlantic oscillation (ea) on the air temperature in montenegro. *Thermal Science*, 22:211–211, 01 2017. doi: 10.2298/TSCI170710211B.

Center for Environmental Prediction. The north atlantic oscillation impact on temperatures, precipitation, snowfall, and snow depth. URL `http://www.cep.rutgers.edu/ oman/NAO.htm`. Last accessed 15 January 2023.

I. A. Renfew G.W.K. Moore. Cold european winters: interplay between the nao and the east atlantic mode. *Atmospheric Science Letters*, 13(1):1–8, 2012. doi: https://doi.org/10.1002/asl.356.

Michael Horswell Harry West, Nevil Quinn. The influence of the north atlantic oscillation and east atlantic pattern on drought in british catchments. *Front. Environ. Sci.*, 10, 2022. doi: https://doi.org/10.3389/fenvs.2022.754597.

Working Group I. Ipcc fourth assessment report: Climate change 2007, 3.6 patterns of atmospheric circulation variability, 2007. URL `https://archive.ipcc.ch/publications`$_{a}nd_{d}ata/ar4/wg1/en/ch3s3 - 6.html.Lastaccessed14January2023$.

Peter V. Hobbs John M. Wallace. *Atmospheric science: an introductory survey*. Elsevier Inc., Oxford, UK, 2 edition, 2006.

David B. Stephenson and Rasmus E. Benestad. *environmental statistics for climate researchers.*

Hans von Storch and Francis W. Zwiers. *Statistical Analysis in Climate Research*. Cambridge University Press, 1999.

Climate Prediction Center Internet Team. East atlantic, 2012. URL `https://www.cpc.ncep.noaa.gov/data/teledoc/ea.shtml`. Last accessed 19 January 2023.

A Appendix

A.1 The math behind EOF

First step of the EOF analysis is to form the data matrix. The composed space-time field, $X(t, s)$, represents the values of the field. The field is a scalar field. In this thesis mainly meteorological variables: msl, gp500 is used. The field depends on time, t, and the spatial position, s. The data matrix is represented as:

$$X = \begin{pmatrix} x_{11} & x_{12} & ... & x_{1p} \\ x_{21} & x_{22} & ... & x_{2p} \\ ... & ... & ... & ... \\ x_{n1} & x_{n2} & ... & x_{np} \end{pmatrix}$$

where the discrete time t_i and the spatial position s_j is combined as x_{ij} for $i = 1, ..., p$ and $j = 1, ..., n$. The data matrix contains the measurement of on of the meteorological variables aforementioned. The location is given by $x_1, x_2, ..., x_p$ and the time for observation is given as $t_1, t_2, ..., t_n$. For each time $t_j (j = 1, ..., n)$ the measurements at $x_i (i = 1, ..., p)$ is a map, the EOF pattern. The p columns of X represents the time series at a location n row. This way of ordering time and position related data into a matrix is the S-mode EOF analysis. Other types of EOF analysis will not be covered in this project.

Next step is to remove the timely mean from the data matrix. This is not strictly necessary for finding eigenvalues/eigenvectors. This does however allow the R to be interpreted as the covariance matrix. Removing the mean also makes it so we are looking at anomalies, which is diversions from the mean.

Step three, here we are going to define the covariance matrix, R, from the field data matrix, X.

$$R = X^t X \tag{12}$$

Step four: The eigenvalue problem. We find the eigenvalues and eigenvectors by solving the eigenvalue problem,

$$RC = CA \tag{13}$$

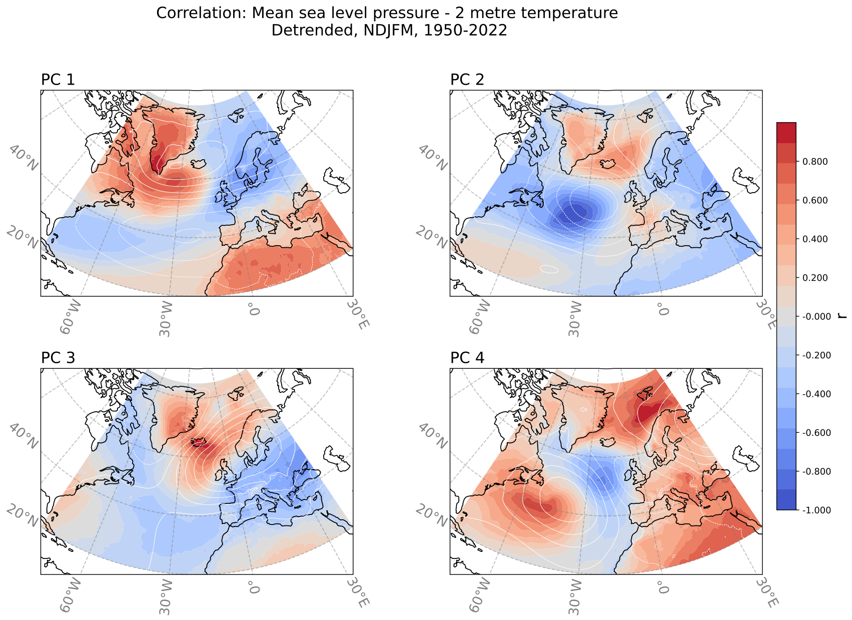
where A is the diagonal matrix of the eigenvalues, λ_i of R and the column c_i of C are the eigenvectors of R. Both matrices, A and C, are of the size $p \times p$. We then have multiple each eigenvalues λ_i with its corresponding eigenvector c_i. The eigenvectors are the EOFs, we are looking for. The eigenvectors are ordered according to size, so the EOF1 is the eigenvector with the largest eigenvalue, EOF2 is the eigenvector with the second largest eigenvalue and so on. The eigenvalues gives a measure of the fraction of the total variance in R explained by the respective mode. This is calculated with the trace of A, which is dividing each eigenvalue λ_i with the sum of all the eigenvalues.

The property of orthogonality. For the eigenvector matrix C it applies that:

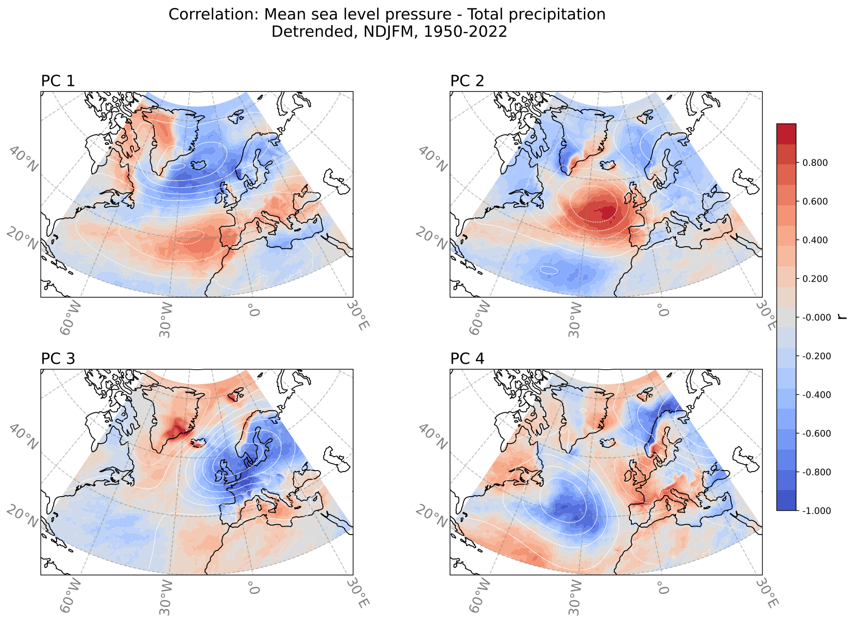
$$C^t C = CC^t = I \tag{14}$$

where I is the identity matrix. This states that the eigenvectors are orthogonal and therefor uncorrelated in space.

A.2 Correlation and regression map (NDJFM)

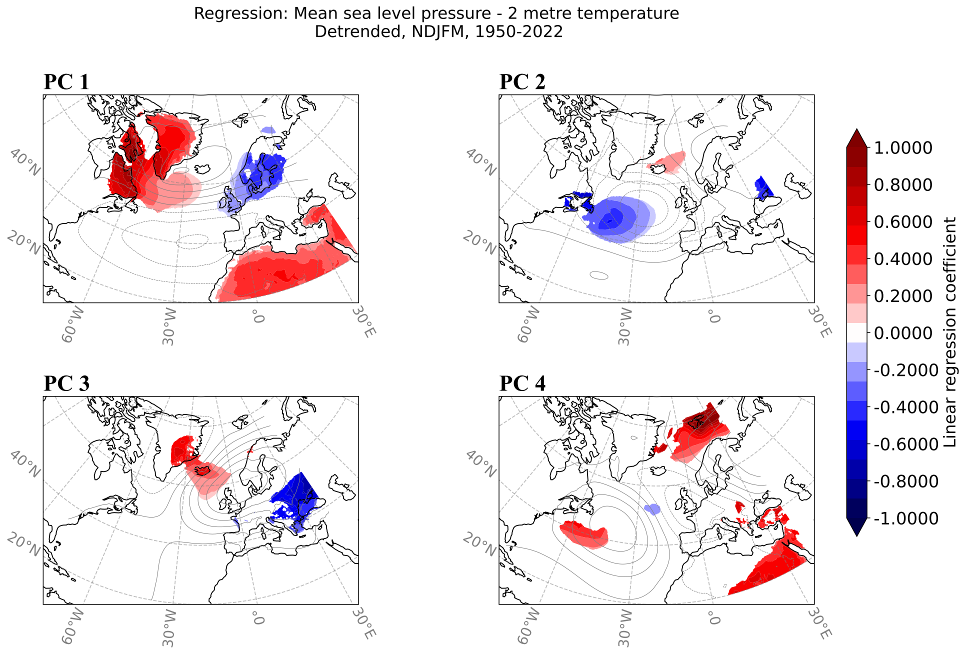


(a) Two meter temperature anomalies

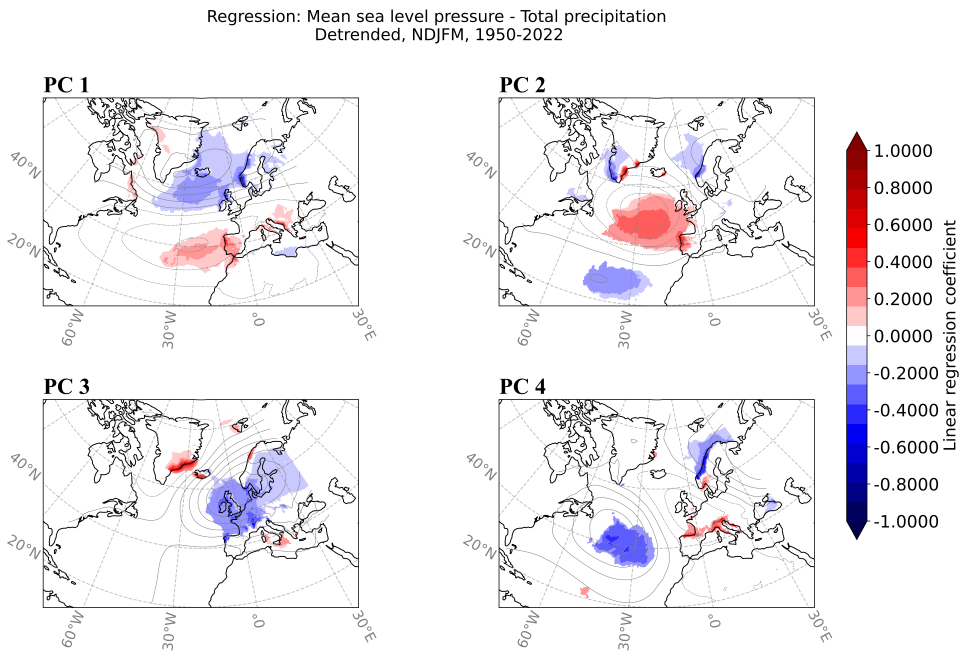


(b) Total precipitation anomalies

Figure 9: Correlation map for the raw winter months (NDJFM) calculated for the mean sea level pressure and a) two meter temperature anomalies and b) total precipitation anomalies in the North Atlantic region for the time period 1950-2022. The white contour lines indicate the EOF pattern associated with each PC.



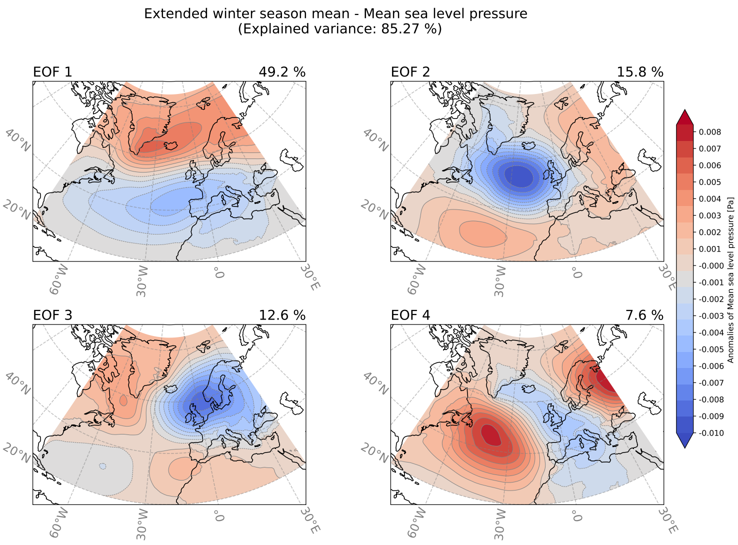
(a) Two meter temperature anomalies



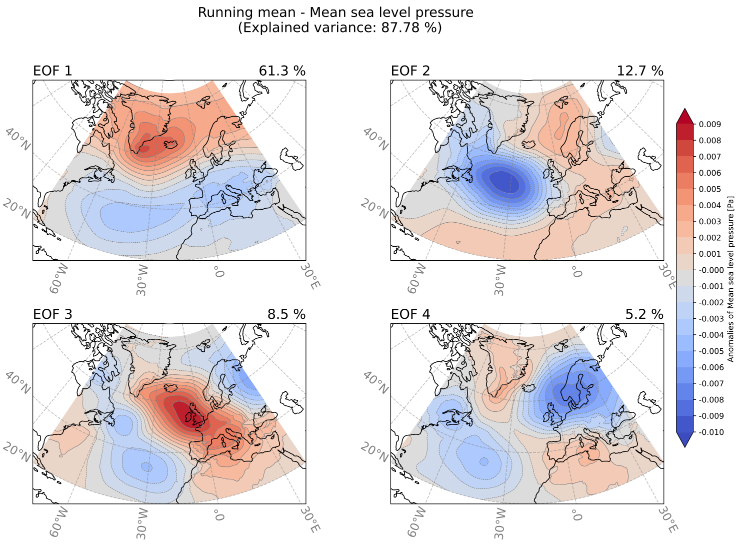
(b) Total precipitation anomalies

Figure 10: PCR map for the raw winter months (NDJFM) calculated for the mean sea level pressure and a) two meter temperature anomalies and b) total precipitation anomalies in the North Atlantic region for the time period 1950-2022. The grey contour lines indicate the EOF pattern associated with each PC. Only results associated with correlation coefficient over 0.3 is included.

A.3 Four leading EOF for extended winter seasonal mean and five-year-running mean



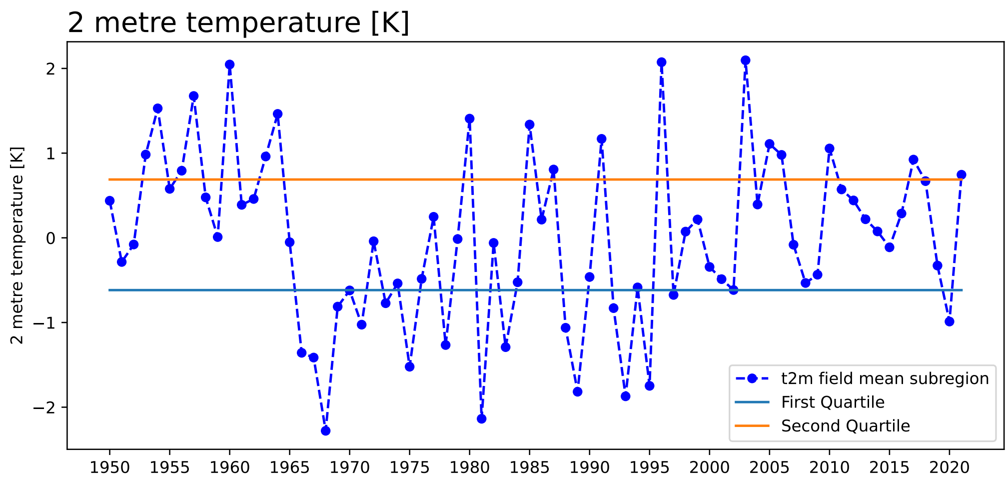
(a) Extended Winter Seasonal mean



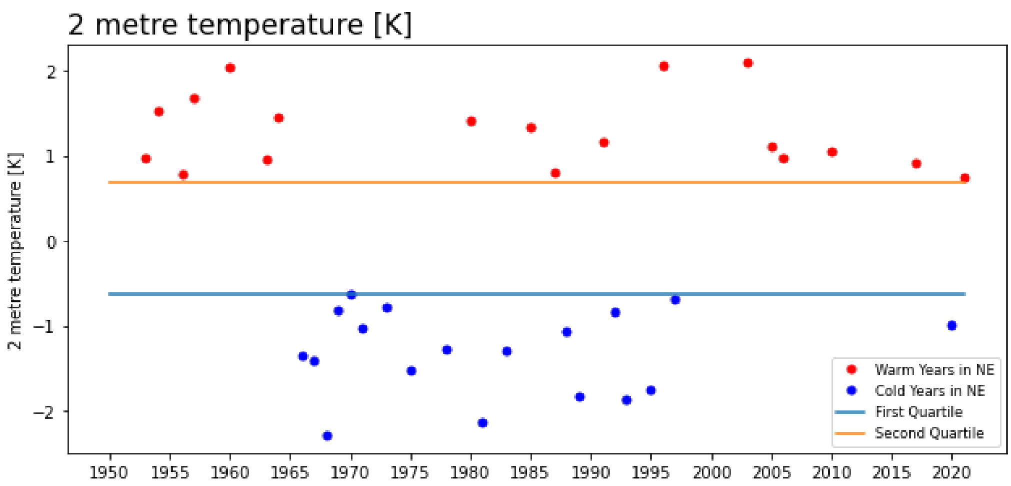
(b) Five-Year-Running mean

Figure 11: The four leading EOF calculated for the mean sea level pressure comparing a) the extended winter seasonal mean and b) the five-year running mean for the time period 1950-2022 for the North Atlantic region. Total explained variance for each analysis is displayed in the header. The explained variance for the individual EOF pattern is shown in the upper-right corner of each EOF sub-plot.

A.4 Temperature and precipitation quartiles

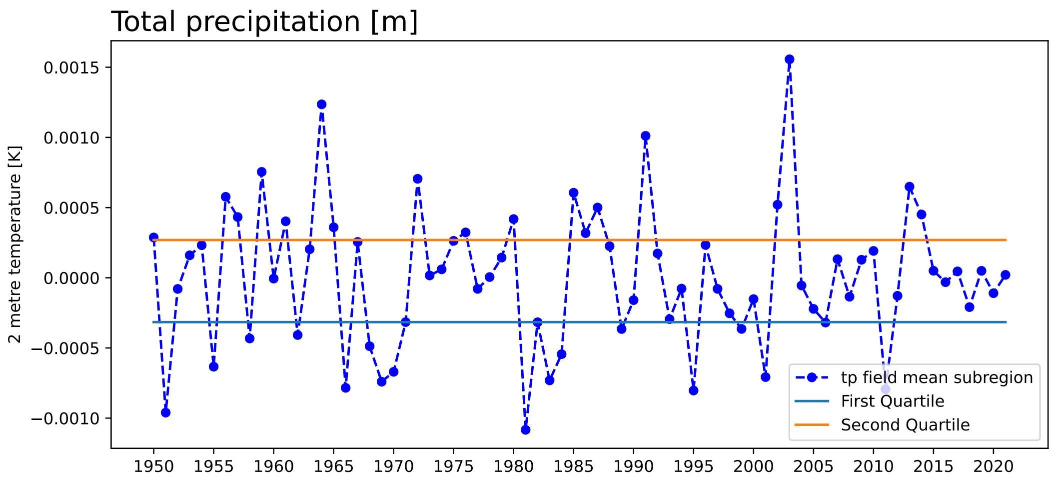


(a)

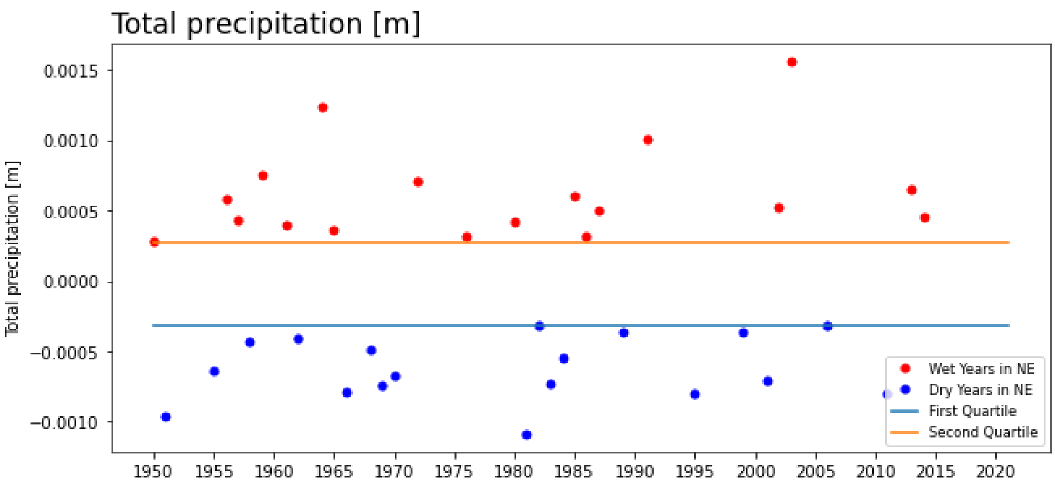


(b)

Figure 12: Defining the warm and cold years for the sub-region covering the time period 1950-2022 for the extended winter seasonal mean by finding the 25th and 75 the quartile of the temperature anomalies.

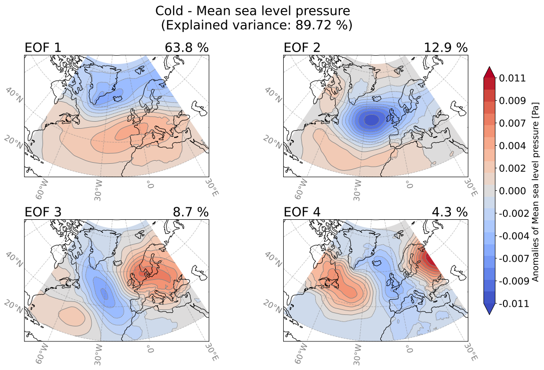


(a)

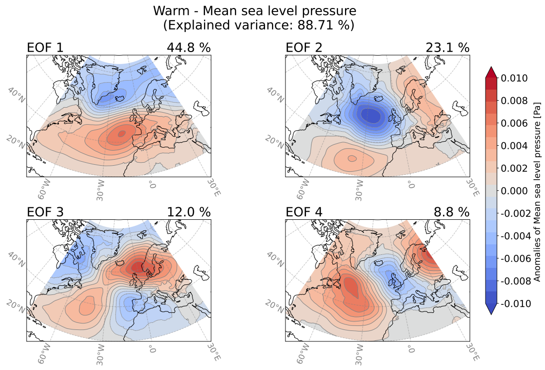


(b)

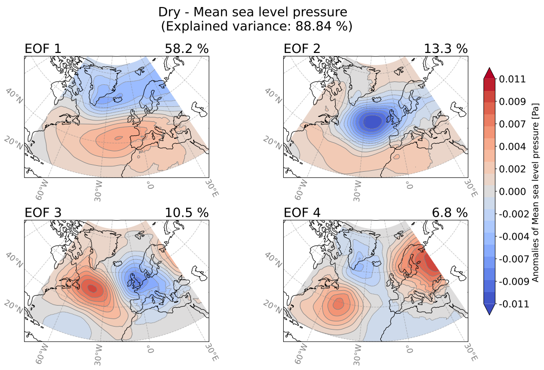
Figure 13: Defining the warm and cold years for the sub-region covering the time period 1950-2022 for the extended winter seasonal mean by finding the 25th and 75 the quartile of the temperature anomalies.



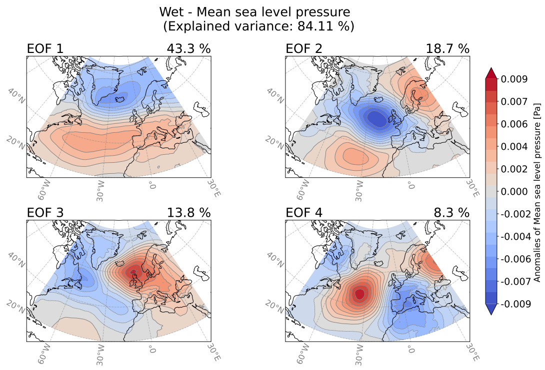
(a) Cold quartile



(b) Warm quartile

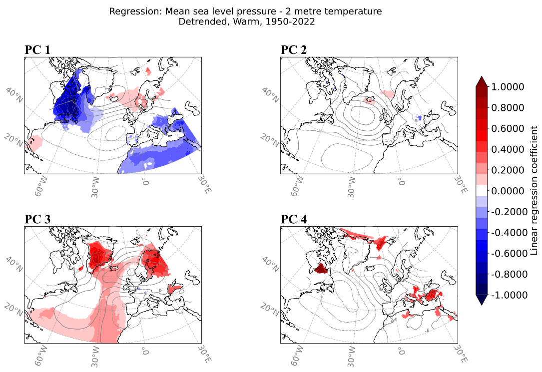


(c) Dry quartile

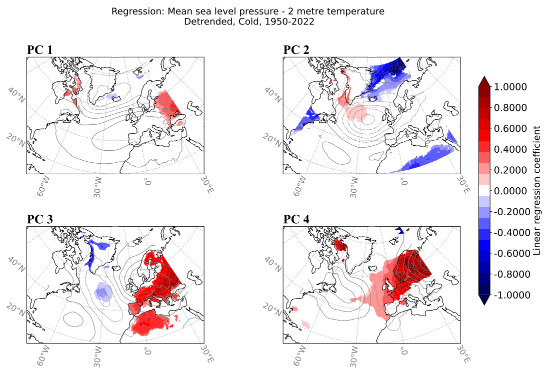


(d) Wet quartile

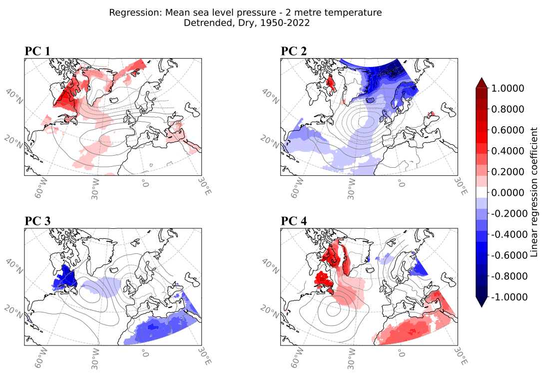
Figure 14: The four leading EOF of mean sea level pressure calculated for the different quartiles based on the extended winter seasonal mean over the North Atlantic region on the time period of 1950-2022. The total explaine variance is shown in the header of each plot and each individual EOF's explained variance is displayed of the upper-right corner.



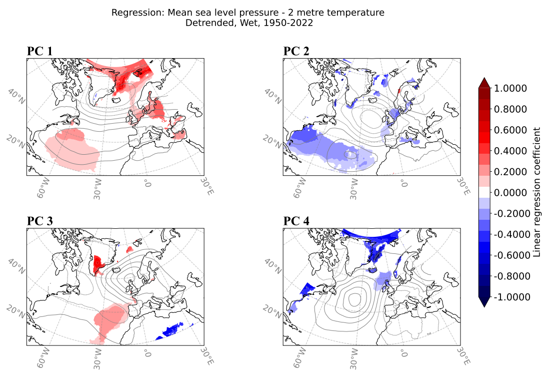
(a) Warm quartile



(b) Cold quartile

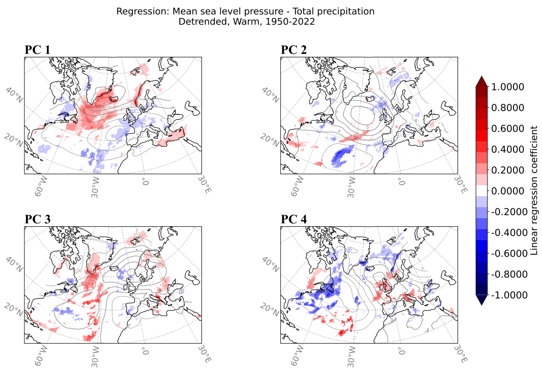


(c) Dry quartile

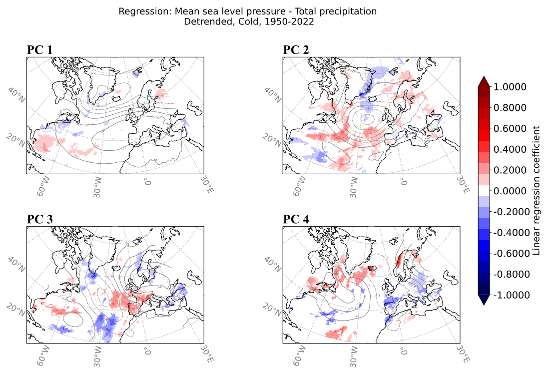


(d) Wet quartile

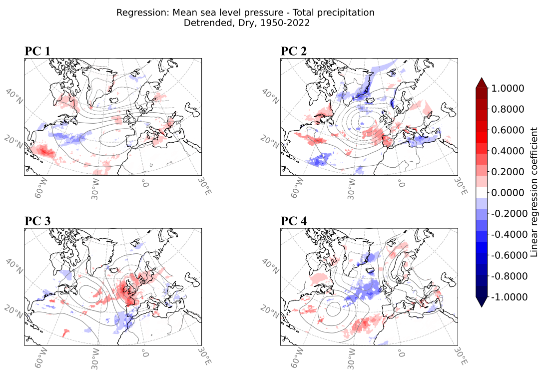
Figure 15: PCR analysis for mean sea level pressure and two meter temperature anomalies for each of the four quartiles over the North Atlantic region for the research period 1950-2022. Only correlation coefficients over the threshold of 0.4 are shown. The EOF pattern are shown as grey contour lines.Only results associated with a correlation over 0.3 are shown



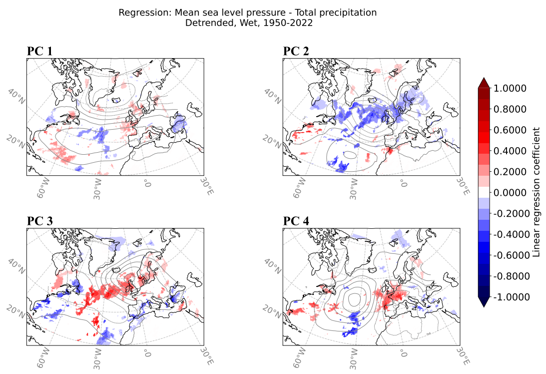
(a) Warm quartile



(b) Cold quartile



(c) Dry quartile



(d) Wet quartile

Figure 16: PCR analysis for mean sea level pressure and precipitation anomalies for each of the four quartiles over the North Atlantic region for the research period 1950-2022. Only correlation coefficients over the threshold of 0.4 are shown. The EOF pattern are shown as grey contour lines. Only results associated with a correlation over 0.3 are shown.