THE NORTH ATLANTIC OSCILLATION HAS AN EFFECT ON WESTERN TURKEY'S PRECIPITATION PATTERN

Merih BOZBURA
bozbura@itu.edu.tr, No: 601161005

Abstract. The North Atlantic Oscillation (NAO) is a large-scale natural climate variability that has important impacts on the weather and climate of the North Atlantic region and surrounding continents, especially Europe. Strong positive phase is disposed to be connected with lower than normal temperatures over southern Europe and the Middle East and it is disposed to be associated with lower than normal precipitation over southern and central Europe. Unlike strong positive phase of NAO , negative phase is disposed to be connected with higher than normal temperature over southern Europe and the Middle East and it is disposed to be associated with higher than normal precipitation over southern and central Europe. In this study, relationship between NAO and annual precipitation of Turkey will be conducted. Exploratory data analysis, linear regression, and principle component analysis are applied to precipitation and NAO indexes datasets to assess whether there is a relationship between NAO and precipitation or not. Only, Marmara region rejects the null hypothesis.The provinces with significant p-values and grouped provinces that are obtained from principle component analysis are nearly same except few provinces. Also, Marmara region has significant p-value. In the light of these informations, precipitation is affected by North Atlantic Oscillation can be said.

1 Introduction

The North Atlantic Oscillation (NAO) is a large-scale natural climate variability that has important impacts on the weather and climate of the North Atlantic region and surrounding continents, especially Europe. NAO is a difference in sea-level pressure of Azores High and Subpolar Low. The positive phase of NAO indicates higher than normal heights and pressure over Europe and the negative phase indicates lower than normal heights and pressure over Europe. Strong positive phase is disposed to be connected with lower than normal temperatures over southern Europe and the Middle East and it is disposed to be associated with lower than normal precipitation over southern and central Europe as it is shown in Figure 1. Unlike strong positive phase of NAO , negative phase is disposed to be connected with higher than normal temperature over southern Europe and the Middle East and it is disposed to be associated with higher than normal precipitation over southern and central Europe as it is shown in Figure 2 (Climate Prediction Center Internet Team, 2012). Therefore, NAO is expected to have an impact on Turkey. Weather and climate conditions are controlled by NAO in Mediterranean basin along with Turkey. The geographical and temporal variations and anomalies in the precipitation for Turkey is strongly associated with strong NAO phases (Türkeş and Erlat, 2003). In this study, relationship between NAO and annual precipitation of Turkey will be conducted. Exploratory data analysis, linear regression, and principle component analysis are applied to precipitation and NAo indexes datasets to assess whether there is a relationship between NAO and precipitation or not.

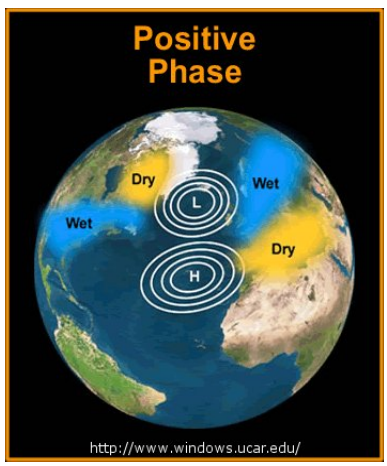

Figure 1. Positive phase of NAO (Global Patterns, n.d).

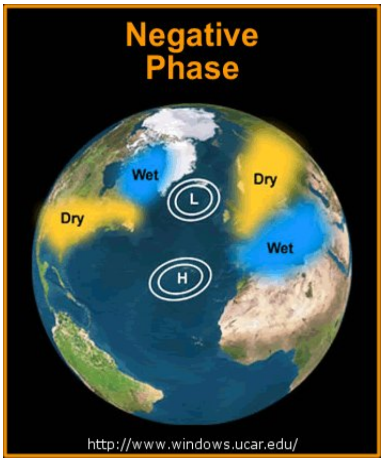

Figure 2. Negative phase of NAO (Global Patterns, n.d).

2 Data and Method

Hypothesis is determined as there is a relationship between precipitation and NAO for this paper. The North Atlantic Oscillation Indexes and annual precipitation of Turkey datasets are used in this paper. Annual precipitation is an observation data and it includes all provinces of Turkey. These datasets are between 1970 and 2012. Firstly, exploratory data analysis are applied to understand fundamentals features of precipitation data. Secondly, linear regression is applied to see whether there is a relationship between NAO and precipitation or not. Finally, principle component analysis is applied to understand effect of NAO on precipitation. This study is conducted with R.

3 Exploratory Data Analysis

Exploratory data analysis is an approach to analyse a data. Exploratory data analysis supports to find out essential properties of the data. Shapiro-Wilk normality test is applied the precipitation data. P-values of Marmara, Ege, Karadeniz, Akdeniz, Ic Anadolu, Dogu Anadolu, and Guneydogu Anadolu regions are respectively 0.001137, 2.65e-15, 2.2e-16, 1.884e-12, 3.409e-13, 2.2e-16, and 3.668e-06. So, the null hypothesis that is the samples come from a normal distribution is rejected. Histogram of Marmara region has right-skewed distribution as it is stated in the Figure 3. A few larger values bring the mean upwards. It is the closest region to normal distribution compared to the other regions and it is expected by looking Shapiro-Wilk normality test results.According to Figure 4, Figure 5, Figure 6, Figure 7, Figure 8, and Figure 9, other regions also have right-skewed distribution.

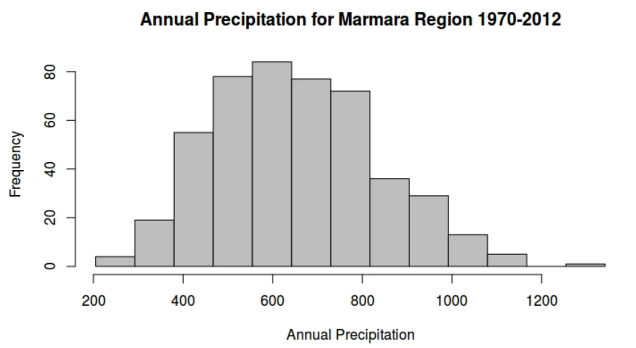

Figure 3. Histogram of Marmara region.

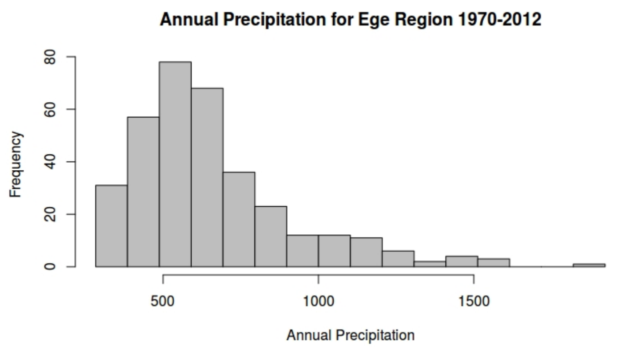

Figure 4. Histogram of Ege region.

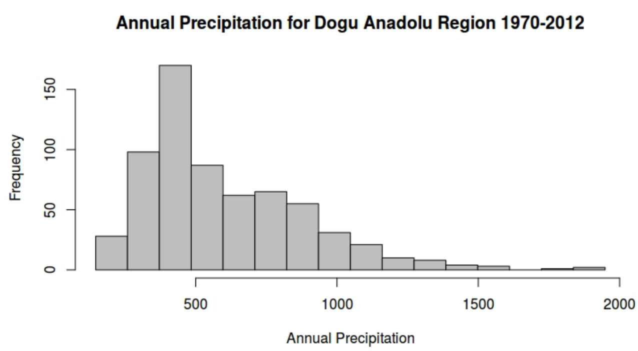

Figure 5. Histogram of Dogu Anadolu region.

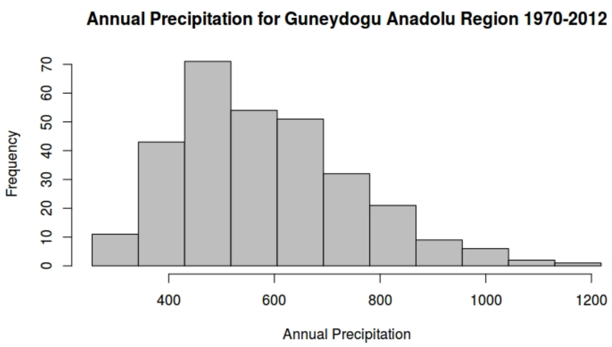

Figure 6. Histogram of Guneydogu Anadolu region.

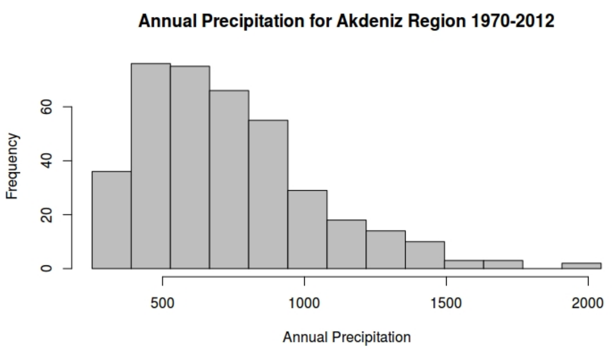

Figure 7. Histogram of Akdeniz region.

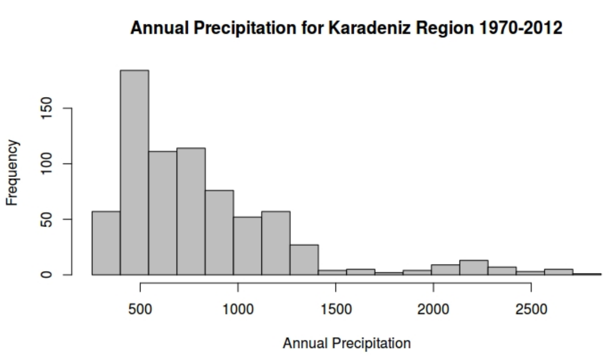

Figure 8. Histogram of Karadeniz region.

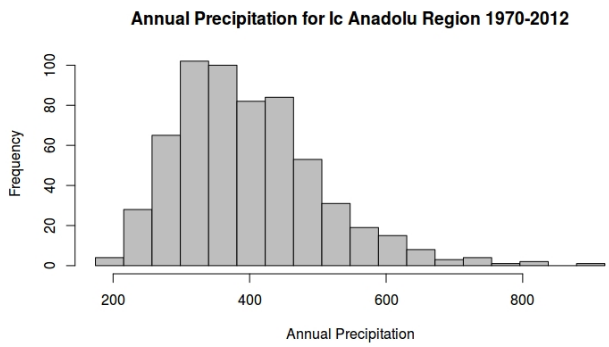

Figure 9. Histogram of Ic Anadolu region.

There are eighty one provinces in Turkey and as it is expected, 53^{rd} province which is Rize has the highest precipitation in Turkey between 1970 and 2012 as it is stated in Figure 10. Also, Iğdır has the lowest precipitation. Average precipitation of Turkey is around 630 mm.

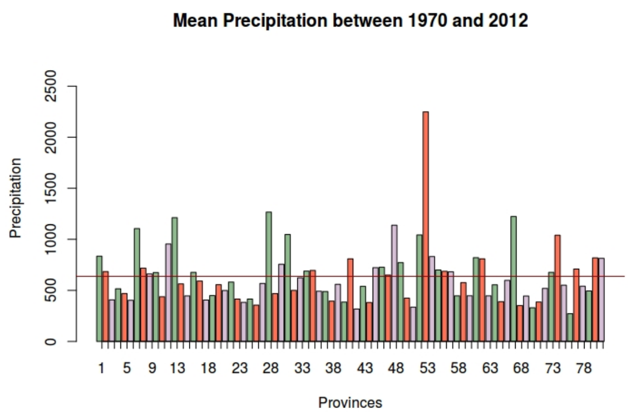

Figure 10. Average barplot for all provinces.

4 Linear Regression

Turkey has seven regions and linear regression is applied between each region and NAO index. P-values of Marmara, Ege, Karadeniz, Akdeniz, Ic Anadolu, Dogu Anadolu, and Guneydogu Anadolu regions are respectively 0.003816, 0.1888, 0.05551, 0.4469, 0.7868, 0.9588, and 0.7791. Only, Marmara region rejects the null hypothesis which is there is no relationship between precipitation and NAO. However other regions do not reject the null hypothesis. The summary of the linear model between Marmara region and NAO index is shown in Figure 11. Marmara region is the independent variable and NAO index is the dependent variable in this linear model. The equation of the model is $NAOindex = 7.4118 - 0.0015 * Canakkale + 0.0015 * Edirne - 0.007 * Istanbul + 0.0032 * Tekirdag + 0.0063 * Yalova - 0.0008 * K\text{ß}rklareli - 0.0019 * Bal\text{ß}kesir - 0.0044 * Bilecik - 0.0035 * Bursa + 0.0014 * Kocaeli - 0.0045 * Sakarya.$

In addition to the Marmara region, Karadeniz and Ege regions are chosen to examine and linear regression is applied between each provinces of these three regions and NAO indexes. Provinces with significant p-values are obtained since they will be compared principle component analysis results to understand that pattern on precipitation belongs to NAO or not. Provinces with significant p-values are Balıkesir, Bilecik, Bursa, Canakkale, Edirne, Istanbul, Kocaeli, Sakarya, Tekirdag, Yalova, Kırklareli, Amasya, Kastamonu, Rize, Izmir, Kütahya, and Manisa.

```
Residuals:
    Min      1Q  Median      3Q     Max
-3.3659 -0.8634  0.0453  0.7355  3.3252

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  7.4117935  1.6522810   4.486 9.32e-05 ***
ckle        -0.0015119  0.0024180  -0.625   0.5364
edir         0.0015837  0.0038671   0.410   0.6850
ist         -0.0069934  0.0034079  -2.052   0.0487 *
tkrd         0.0032244  0.0028871   1.117   0.2726
yalv         0.0063374  0.0036696   1.727   0.0941 .
kirk        -0.0008512  0.0033302  -0.256   0.8000
blke        -0.0019170  0.0022720  -0.844   0.4053
blck        -0.0044137  0.0044600  -0.990   0.3300
brsa        -0.0034628  0.0044263  -0.782   0.4400
keli         0.0014010  0.0032915   0.426   0.6733
skry        -0.0044772  0.0027584  -1.623   0.1147
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.547 on 31 degrees of freedom
Multiple R-squared:  0.5439, Adjusted R-squared:  0.3821
F-statistic: 3.361 on 11 and 31 DF,  p-value: 0.003816
```

Figure 11. The summary of the linear model between Marmara region and NAO index.

The R^2 of the model is 0.5439 and it is not too small. So, the model explains 54% of the variability of the dependent variable which is NAO index.

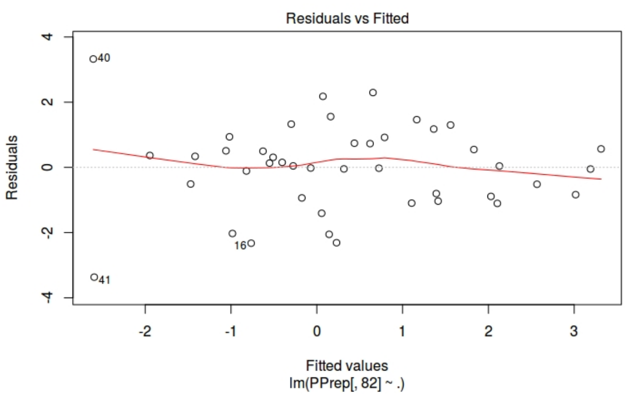

Figure 12. Residual vs Fitted plot.

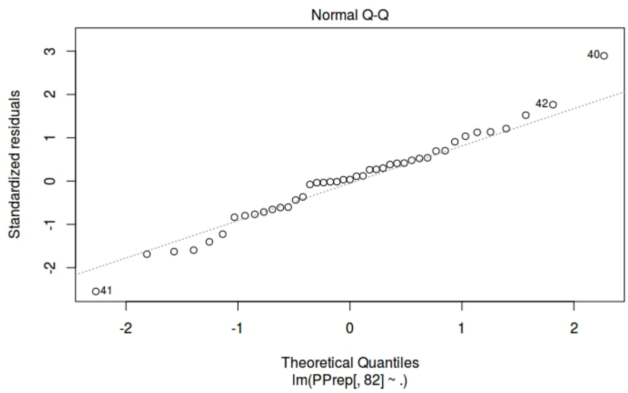

Figure 13. Normal Q-Q plot.

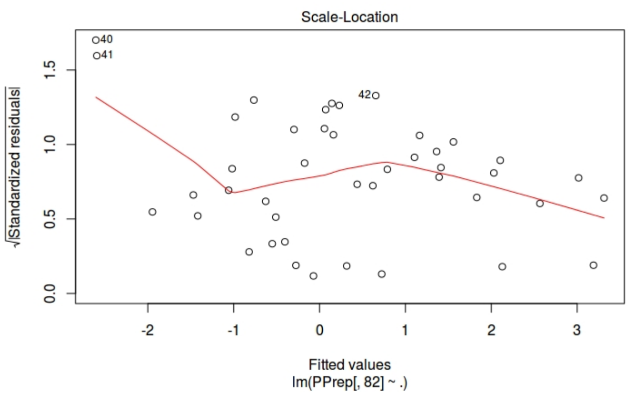

Figure 14. Scale Location plot.

Residual is the difference between fitted and actual dependent point. Fitted point is a predicted point by the model.There should be no discernible pattern around zero residual versus fitted plot. As it is stated in Figure 12, residuals and fitted values are almost randomly distributed around the zero line. Therefore, the model is not very good, but it fits the data. According to Figure 13, residuals nearly follow the normal distribution. Also, scale-location residuals are randomly distributed and there is no discernible pattern as it is shown in Figure 14.Cook's distance is used to find dominant points in independent variables. These points are far from the other points. In this case, 29^{th} and 40^{th} points are little away from the other points and 41^{st} point are far from the other points as it is shown in Figure 15. If the Cook's distance is greater than 0.5, that point can be influencial, so it is worthy to examine (Identifying Influential Data Points, n.d). Cook's distance of 41^{st} point is greater than 0.5. If exceeding point equals about two times of average of data, it should be examined (Jacoby, n.d). 41^{st} point is not greater or equal to the two times of average of data. So, it is not worthy for investigating.

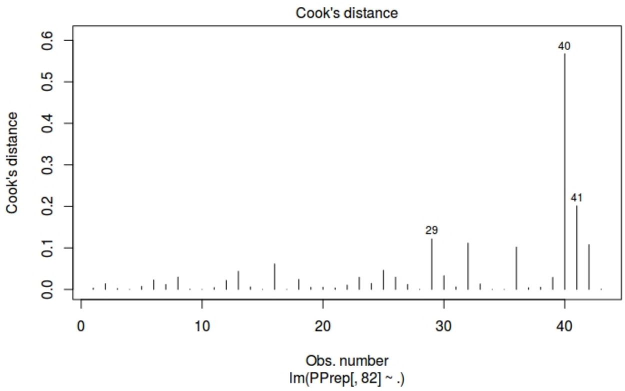

Figure 15. Cook's distance plot.

5 Principle Component Analysis

Principal Component Analysis(PCA) is the oldest and the most famous multivariate statistical technique (Abdi and Williams, 2010).The goal of principal components analysis is to clarify the maximum amount of variance with less number of principal components. It is used for reducing dimension of very big datasets in addition to keep most of the information in the big dataset. R function of principle component analysis pulls the data normal distribution with centring and scale arguments. Before applying PCA, precipitation data is scaled and standardized since the scale between NAO indexes and precipitation is large. Firstly, scree plots or percent variance plots are examined to determine the number of principle components that are enough to explain dataset. According to the Kaiser's Criterion, if eigenvalues are greater than 1, these PCs can be taken (Habing, 2003). Secondly, interpreting how much do variables contribute to the principle components by looking the loadings. Finally, interpreting and understanding distribution of the variables in this classification by examining the scores.

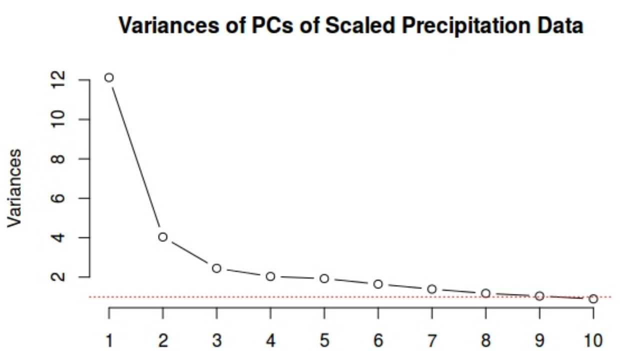

Figure 16. Scree plot of scaled precipitation.

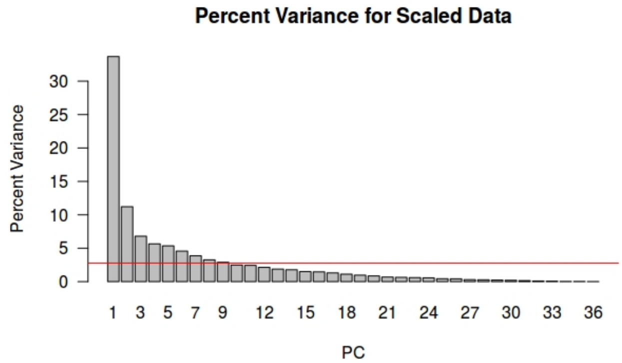

Figure 17. Percent variance plot of scaled precipitation.

According to Figure 16 and Figure 17, first nine principal component explains the important part of the dataset. The horizontal red line on the Figure 17 represents that if each variable devoted equally, they would devote 2.8% to the total variance since there are thirty six variables. Fisrt nine PCs represent 77.32% of the data as it is seen on Figure 17 or cumulative variance in R code.

```
           PC1        PC2         PC3         PC4         PC5          PC6
amsy 0.19283777  0.119763046 -0.03005605  0.28866468 -0.118832484  0.01536832
artv 0.05226084  0.264700611 -0.16404973  0.08828010  0.090917379 -0.25086379
brtn 0.10246691  0.102942227  0.23433216 -0.43092785  0.005865238  0.22750527
bolu 0.19794247 -0.008038566 -0.04359529 -0.03859983  0.144306162 -0.15594896
bybr 0.08957615  0.260417898  0.03122757  0.16478895  0.026550438  0.33649288
dzce 0.12998977 -0.072638956  0.11479633  0.07209496  0.465716324 -0.05614788
            PC7       PC8        PC9         PC10        PC11        PC12
amsy -0.308915957  0.0103766 -0.15831858  0.06485237 -0.08679631 -0.02413529
artv  0.378560195  0.1051176  0.20409231 -0.05343447 -0.08499438 -0.34240889
brtn -0.007291841 -0.1360664 -0.14005068  0.23096118 -0.17807854 -0.04297476
bolu -0.223777985  0.2694080  0.08123273 -0.02121548  0.10364551 -0.17639561
bybr  0.049641197 -0.1400139  0.02503747 -0.33552876  0.06749622 -0.22681061
dzce -0.117971754  0.2219801 -0.10274952 -0.02636045  0.11103084 -0.05734660
            PC13        PC14        PC15         PC16        PC17        PC18
amsy  0.232370506  0.01984874  0.005875968  0.02852258 -0.06692604 -0.11372587
artv -0.014534806  0.08343316  0.151164577 -0.14948521  0.04142609  0.30164855
brtn -0.112211765  0.18525763  0.069714588 -0.02916501  0.07142305  0.21761181
bolu -0.380407674 -0.16667922 -0.056719922  0.06972157 -0.14940535 -0.07118022
bybr  0.261749031 -0.22302055 -0.295300242 -0.03862699  0.10793592 -0.08791713
dzce -0.002443074  0.22273658 -0.420368629  0.20570489  0.09611663  0.08682555
           PC19        PC20        PC21        PC22        PC23        PC24
amsy  0.12500803  0.18938429 -0.12448131 -0.13665001 -0.11135995 -0.03093511
artv -0.07150660  0.32465824  0.04483325 -0.06779759  0.15462779  0.05228155
brtn  0.12559604 -0.05496912 -0.31651080 -0.03537053  0.16795779 -0.15717926
bolu  0.44644394 -0.06209016 -0.12398088  0.20972742  0.26651262  0.02246975
bybr -0.05736425 -0.10635022 -0.35260157  0.39234898  0.09152966  0.01644350
dzce -0.06571385  0.19091448 -0.04759756 -0.16673993 -0.16474965 -0.12837233
           PC25        PC26        PC27        PC28         PC29        PC30
amsy  0.11808717  0.34778134 -0.03193633 -0.47437977 -0.021264502 -0.13661897
artv  0.14575982 -0.14917760 -0.16759713 -0.26833236  0.157831381 -0.08522787
brtn  0.21438479 -0.09176541  0.13192405 -0.03144300 -0.064342191 -0.07595714
bolu  0.14400988 -0.01143379 -0.04339387 -0.04349219  0.147987469  0.11236757
bybr -0.05936514 -0.11609637 -0.06822225 -0.03292171 -0.003408239  0.02049329
dzce -0.11802220 -0.02200175 -0.17529602  0.26114143  0.206103508 -0.17865650
            PC31        PC32         PC33        PC34        PC35        PC36
amsy  0.295981374  0.05976773  0.116714972 -0.16013108 -0.18399502 -0.03037759
artv  0.007748655  0.02019195 -0.117462817 -0.05509367  0.10620078  0.06911158
brtn  0.071832040 -0.24932162 -0.010080901 -0.34088314  0.10453760 -0.07720621
bolu -0.138883297 -0.05494048 -0.009733375  0.10840534 -0.28428358 -0.15651497
bybr -0.072212082  0.06518384  0.032431917 -0.06528853  0.15407330 -0.03597567
dzce  0.011973330 -0.07198673  0.111800414 -0.17386594  0.05213561  0.10470732
```

Figure 18. Heads of loadings of scaled precipitation.

In the Figure 18, positive loadings represents positively correlated principle component and variable and negaitve loadings represents negatively correlated principle component and variable. Variables with large loadings have big effect on principle component. For example, Canakkale, Istanbul, Tekirdağ, Yalova, Bilecik, Bursa, Kocaeli, Sakarya, Aydın, Denizli, and Izmir have large loadings in PC1 and

Artvin, Bayburt, Giresun, Gümüşhane, Rize, Ordu, Samsun, Tokat, Trabzon, Sinop have large loadings in PC2 in this case. If PC2 is wanted to plot, plot would looks like time series of Artvin, Bayburt, Giresun, Gümüşhane, Rize, Ordu, Samsun, Tokat, Trabzon, Sinop since these provinces contributed more than other provinces for PC2. It is the same for other PCs and the other loadings can be seen by looking the R code or Figure 18. .

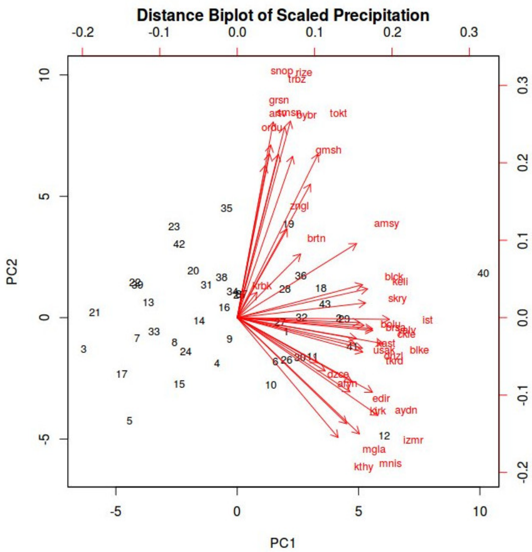

Figure 19. Biplot of scaled precipitation.

Scores are the position of each variables in this new coordinate system of PCs. Points with similar features can be grouped by looking biplots. Points that are close together correspond to variables that have similar scores on the components. According to Figure 19, Kütahya, Manisa, Izmir,Edirne, Kırklareli are near each other and Bolu, Yalova, Bursa, Istanbul, Balıkesir, Canakkale, Tekirdağ, Kastamonu, Bilecik, Sakarya, and Kocaeli are near each other. These provinces are the provinces with significant p-values. Also, Bartın, Zongudak, Karabük, Gümüşhane, Tokat, Sinop, Trabzon, Giresun, Artvin, Samsun, Bayburt, and Ordu are near each other and Düzce, Afyon, Aydın are near each other and these provinces are not the provinces with significant p-values. Kütahya, Manisa, Izmir,Edirne, Kırklareli are near each other, so they can be used interchangeably. Therefore, dimension can be reduced. Reducing dimension can be applied for all the groups that mentioned above.

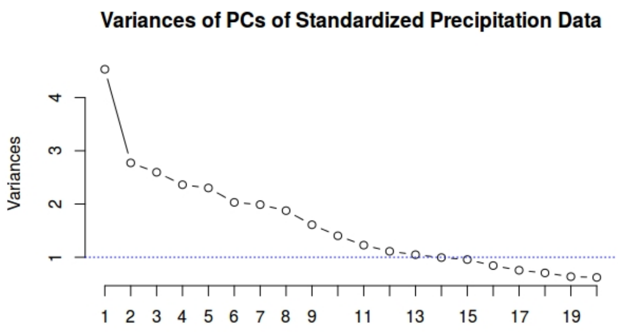

Figure 20. Scree plot of standardized precipitation.

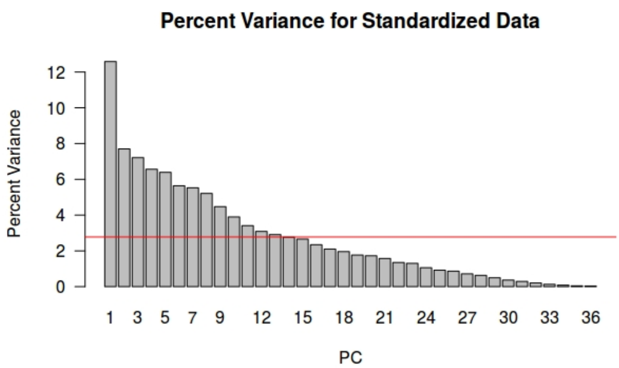

Figure 21. Percent variance plot of standardized precipitation

According to Figure 20 and Figure 21, first fourteen principal component explains the important part of the dataset. The horizontal red line on the Figure 21 represents that if each variable devoted equally, they would devote 2.8% to the total variance since there are thirty six variables. Fisrt fourteen PCs represent 77.37% of the data as it is seen on Figure 21 or cumulative variance in R code.

```
            PC1        PC2        PC3        PC4        PC5         PC6
amsy 0.190917370 -0.10882687  0.05141870 -0.11659648  0.220187450 -0.25046787
artv 0.159170028  0.02972431 -0.18571539 -0.11462989 -0.140285531  0.31035921
brtn 0.005780693  0.28296529  0.02580156  0.10206587  0.052659907 -0.14063213
bolu 0.160411960 -0.22707174 -0.15904520 -0.06488459 -0.107606497 -0.03303453
bybr 0.045016140  0.18812854  0.21268659 -0.05394806 -0.327256015 -0.12985121
dzce 0.094492233 -0.39511466 -0.23924820 -0.14497623 -0.002692687  0.14709476
            PC7        PC8        PC9       PC10         PC11         PC12
amsy  0.01748991  0.01213203 -0.27856946  0.21312894 -0.087358986  0.117439806
artv -0.14442985  0.33112145  0.14963759  0.08900313  0.001781240  0.124396705
brtn  0.49259503  0.01345546  0.05561997 -0.06381486  0.105273497  0.016205012
bolu  0.10652162 -0.12393069  0.10217536  0.35844286 -0.417142049  0.155072487
bybr -0.06257204  0.16788589  0.07138236 -0.01134035 -0.391994736  0.157649241
dzce  0.07768172 -0.07581670 -0.05795075 -0.03312606 -0.005393209 -0.008722816
           PC13        PC14        PC15        PC16         PC17         PC18
amsy  0.17472283  0.02620003 -0.22286777 -0.03033826 -0.038174213  0.369600182
artv -0.25969954 -0.19189503  0.10964632 -0.01949750  0.004816732  0.103712347
brtn -0.11293913 -0.01525968  0.23637692  0.18380731 -0.095529354  0.052935052
bolu -0.01734649 -0.06228804  0.19543190 -0.12601097 -0.089632050 -0.009093564
bybr -0.19487672  0.04107363 -0.06091066  0.04027815  0.016201017  0.112959642
dzce -0.03440088  0.08706296 -0.11026700  0.01147127  0.033232617  0.011323984
           PC19         PC20        PC21       PC22        PC23        PC24
amsy  0.14412477  0.184129806 -0.12714884 -0.2657740 -0.04747856  0.09828023
artv -0.07091178 -0.112446291 -0.17924217 -0.1432906  0.03677026  0.14263509
brtn  0.06702760 -0.040535305 -0.03118077 -0.1711518  0.17833425  0.31242018
bolu  0.01902587  0.004799248 -0.23231978  0.1903225  0.09647022 -0.02617809
bybr -0.18910647  0.078022833  0.07241137  0.2645319 -0.25202600  0.20050404
dzce  0.17364625 -0.370284655  0.08478111  0.1546718 -0.35948907 -0.11925784
           PC25         PC26        PC27        PC28        PC29        PC30
amsy  0.31867750 -0.077626150 -0.07825560 -0.07390232 -0.12696349  0.11004685
artv -0.20846507  0.004965736  0.02778470  0.22427209  0.07697261  0.08520367
brtn  0.08404681 -0.013922353  0.03318232  0.01246166  0.15898656 -0.04613209
bolu -0.05430131  0.182362539  0.08480912 -0.07221634 -0.20449100 -0.13650194
bybr  0.19786551 -0.002441514  0.01320489 -0.30288884  0.04294197 -0.03057855
dzce  0.20736142 -0.022178132 -0.22474182  0.10693012  0.22207538 -0.14512111
           PC31        PC32        PC33        PC34        PC35        PC36
amsy  0.13267186  0.09247458 -0.15371842 -0.19128517  0.26857472  0.01379708
artv -0.04021543 -0.02453061 -0.07407727 -0.38066880  0.33677243 -0.20151842
brtn  0.23732163 -0.24239367 -0.09986665 -0.28213367 -0.28725166  0.14647348
bolu  0.13057672 -0.08866328  0.35453719 -0.05280296 -0.07519773  0.28599726
bybr  0.03764037 -0.11324785 -0.14821205  0.07954768 -0.11301744 -0.32660509
dzce  0.03267417 -0.09527214 -0.19355754 -0.23409775 -0.27532734 -0.02381403
```

Figure 22. Heads of loadings of standardized precipitation.

As it is mentioned before, variables with large loadings have big effect on principle component and some of them are shown in Figure 22. For example, Gümüşhane, Tekirdağ, Yalova, Kırklareli, Izmir, Manisa, Muğla, Uşak, Bursa, Balıkesir, and Aydın have large loadings in PC1 and Bartın, Bolu, Düzce, Istanbul, Kırklareli, Afyon, Muğla, Bayburt, Giresun, and Rize have large loadings in PC2 in this case. The other loadings can be seen by looking the R code or Figure 22.

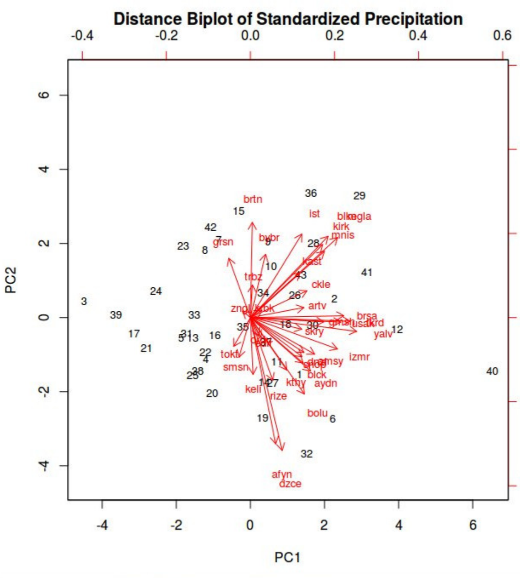

Figure 23. Biplot of standardized precipitation.

According to Figure 23, Istanbul, Balıkesir, Kırklareli, Manisa, Kastamonu, Canakkale are near each other and Bursa, Tekirdağ, Gümüşhane, Yalova, Sakarya, and İzmir are near each other and Amasya, Bilecik, Kütahya, Bolu are near each other and Edirne, Kocaeli, and Rize are near each other. These provinces are the provinces with significant p-values. Also, Zonguldak, Giresun, Trabzon, Bartın, and Bayburt are near each other and Düzce, Afyon, Aydın are near each other and Tokat and Samsun are near each other. These provinces are not the provinces with significant p-values. Istanbul, Balıkesir, Kırklareli, Manisa, Kastamonu, Canakkale are near each other, so they can be used interchangeably. Therefore, dimension can be reduced. Reducing dimension can be applied for all the groups that mentioned above.

In this section, principle component analysis are examined and provinces that are near each other are obtained. Also, provinces have been grouped by means of principle component analysis and which provinces can be used interchangeably have been specified.

6 Conclusions

Exploratory data analysis is applied to understand the datasets and linear regression analysis is applied to find out whether there is a relationship or not between precipitation and NAO index datasets. Only, Marmara region rejects the null hypothesis is found out and to investigate NAO's effect, Ege and Karadeniz regions are choosen in addition to Marmara region. Principle component analysis are applied to three region which contains thirty six provinces. As it is seen in section 4 and section 5, the provinces with significant p-values and grouped provinces that are obtained from principle component analysis are nearly same except few provinces. Also, Marmara region has significant p-value. In the light of these informations, precipitation is affected by North Atlantic Oscillation can be said.

References

Abdi, H. and Williams, L. J.: Principal component analysis, Wiley Interdisciplinary Reviews: Computational Statistics, 2, 433–459, doi:10.1002/wics.101, 2010.

Climate Prediction Center Internet Team, .: North Atlantic Oscillation (NAO), http://www.cpc.ncep.noaa.gov/data/teledoc/nao.shtml, 2012.

Global Patterns, .: Global Patterns Arctic and North Atlantic Oscillations (AO and NAO), http://climate.ncsu.edu/climate/patterns/NAO.html, n.d.

Habing: Exploratory Factor Analysis, http://people.stat.sc.edu/habing/courses/530EFA.pdf, 2003.

Identifying Influential Data Points, .: Identifying Influential Data Points, University Lecture, https://onlinecourses.science.psu.edu/stat501/node/340, n.d.

Jacoby, .: Regression III:Advanced Methods, University Lecture, http://polisci.msu.edu/jacoby/icpsr/regress3/lectures/week3/11.Outliers.pdf, n.d.

Türkeş, M. and Erlat, E.: Precipitation changes and variability in Turkey linked to the North Atlantic oscillation during the period 1930-2000, International Journal of Climatology, 23, 1771–1796, doi:10.1002/joc.962, 2003.