

한반도 기후 변화에 대한 시계열 분석

시계열 분석 기말 보고서

2021234640 이종현

서론

최근 세계 곳곳이 홍수 또는 가뭄, 폭염, 폭설 등으로 고통을 받고 있다. 기존과 달라진 기후 패턴에 세계 각지에서 예상치 못한 재난들이 쏟아지고 있으며, 이는 한반도라고 예외가 아니다. 최근 폭우로 인해 서울 강남이 침수된 바 있으며, 포항 등지에서도 침수 피해가 많이 보고되고 있다. 또한 가뭄 피해 역시 보고 되고 있어 기존의 한반도의 기후 패턴과는 다소 상이하다는 것을 피부로 체감하고 있다. 이에 따라 기후 및 환경을 보존하고 자원을 재순환하는 등의 움직임이 정부, 기업, 개인의 차원에서 활발히 이루어지고 있다.

본 프로젝트는 누적된 기상 데이터를 바탕으로 한반도의 기후 변화가 데이터 상에서도 확인이 되는지, 확인이 된다면 어떠한 패턴으로 변화하고 있으며 변화하는 추세를 바탕으로 예측을 진행하는 방식으로 진행하고자 한다. 본 프로젝트는 한반도의 대표적인 지역 2곳을 선정하여 분석을 진행하고자 한다. 가장 먼저 한국의 수도인 서울을 선정하였다. 서울은 가장 많은 인구가 거주하는 곳이기 때문에 사회, 문화적 중심지이다. 가장 많은 사람이 영향을 받은 곳인 동시에 정부 기관 등이 밀집하여 있기 때문에 정책에 대해 가장 빠르게 반응할 것이라 판단하였다. 또 다른 지역은 제주도 서귀포시를 선정하였다. 제주도 서귀포시는 제주특별자치도의 남단에 위치하는 시로 한라산의 이남에 위치한다. 영동, 영서 지방이 그러하듯 제주도의 서귀포시와 제주시는 다른 날씨 패턴을 보이게 된다. 이는 태백산맥의 영향을 받는 것처럼 한라산의 영향이 있다. 때문에 본 프로젝트에서는 제주도를 하나의 단일 지역으로 보기보다는 서울과 가장 먼 곳인 제주도 서귀포시를 타겟으로 하여 분석하고자 한다.

본 프로젝트는 다음과 같은 가설을 검증하고자 실시하였다.

- 서울과 서귀포시는 지구 온난화의 영향으로 평균 기온이 상승하고 있는가?
- 서귀포시는 기존 온대 기후에서 아열대 기후로 바뀌고 있는가?

방법 및 결과

데이터셋

기상 데이터 수집을 위해 많은 소스가 있으나 본 프로젝트에서는 국가 통계 포털(KOSIS)에서 제공하는 월별 기상 데이터와 서울 열린 데이터 광장에서 데이터를 수집하였다. 각 데이터는 모두 월별 평균 값으로 제공되어 있으며, 모든 집계 값은 월 별 기준이다. 제공 항목은 다음과 같다.

- 기온: 평균 기온, 평균 최고 기온, 극점 최고 기온, 평균 최저 기온, 극점 최저 기온
- 강수량
- 상대습도: 평균 습도, 최소 습도
- 평균 해면 기압
- 평균운량 (구름의 양)
- 일조시간
- 최심신적설
- 바람: 평균 풍속, 최대 풍속, 최대 순간 풍속

해당 데이터의 컬럼 중 본 프로젝트는 평균 기온, 평균 최고 기온, 평균 최저 기온, 강수량, 평균 습도만 분석에 이용하였다. 서울시의 데이터는 1961년 1월부터 2021년 12월까지 제공되었다. 하지만 제주도 서귀포시의 데이터는 1995년 1월부터 활용이 가능하여 본 프로젝트에서는 1995년 1월부터 2020년 12월까지 25개년 치의 데이터를 활용하였다.

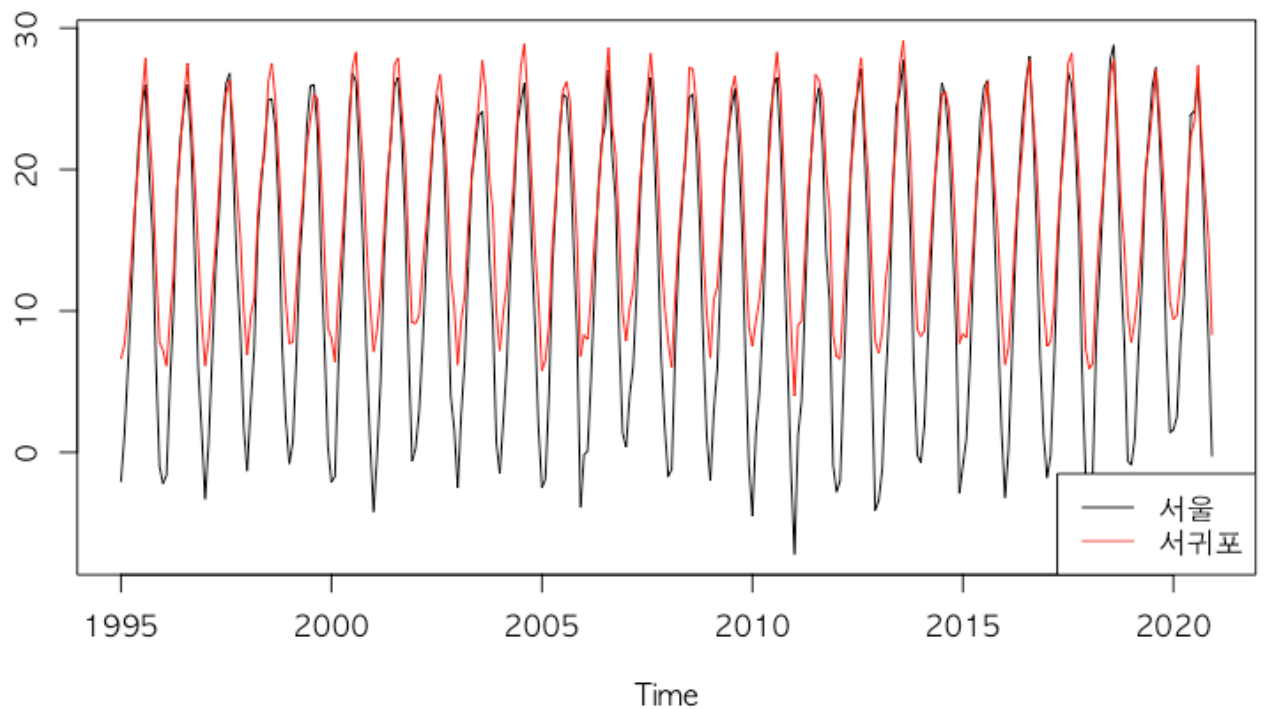
분석 전반

본 프로젝트의 목표는 실제 데이터를 바탕으로 가설을 검증하는 과정을 거치는 경험에 목표를 둔다고 판단하였기 때문에 자동으로 모수를 찾아주는 AutoARIMA 등의 방법은 제외하고 ACF, PACF의 패턴을 보고 판단하는 방법을 채택하였다. 분석은 R project (4.1.0) 을 이용하였으며, Apple M1 chip 프로세서 (Apple Inc, Macbook air 2021) 를 이용하여 분석되었다.

기온

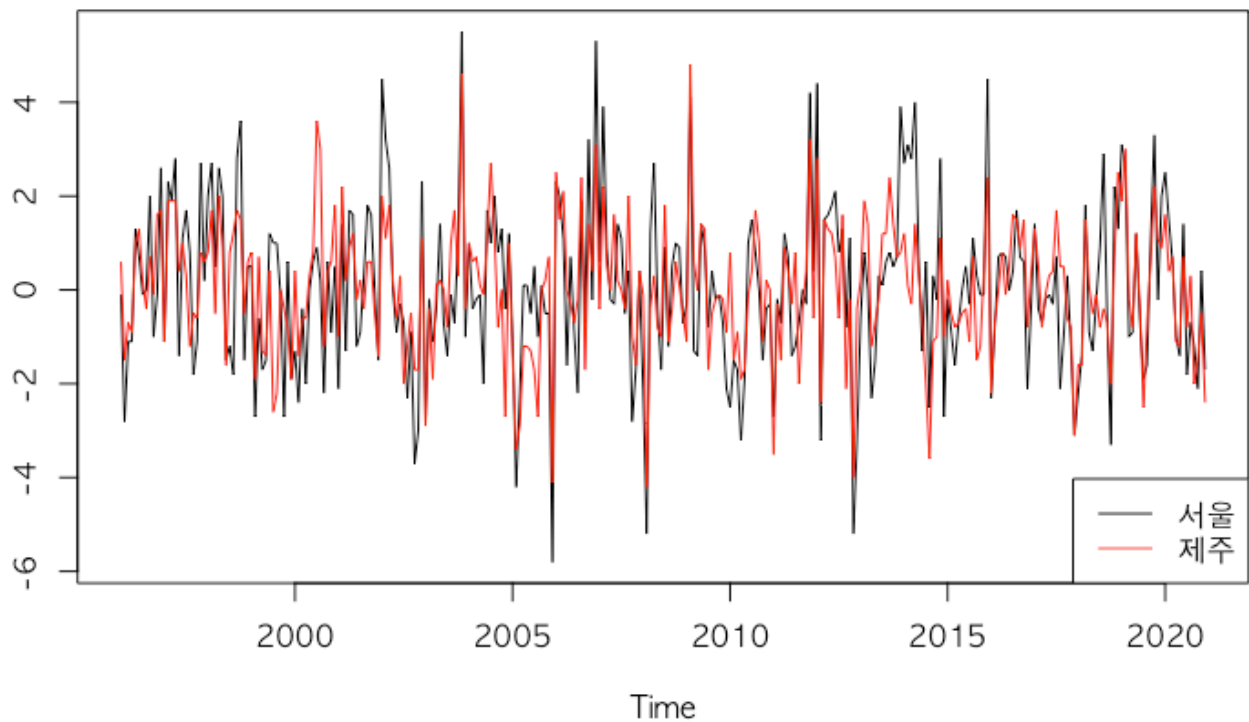
서울과 서귀포시의 기온 시계열 그림은 아래 그림 1에 나타나있다.

서울과 제주의 평균 기온 시계열 그림



서울의 평균 기온은 12.88 (9.92) 이었으며 서귀포의 평균 기온은 16.96 (6.84)로 서귀포의 평균 온도가 서울보다 높다는 점을 확인할 수 있었다. 시계열 그림에서 강한 상관관계가 확인되어 계산해본 결과 0.9777의 상관 계수가 확인되었다. 또한 선형 추세는 보이지 않았으나 강한 계절성이 확인되었다. 분산은 시간에 따라 증가하는 패턴이 보이지는 않았다. 이에 따라 계절 차분을 진행하였으며, 이때 주기는 12를 사용하였다. 계절 차분 후 시계열 그림은 다음과 같다.

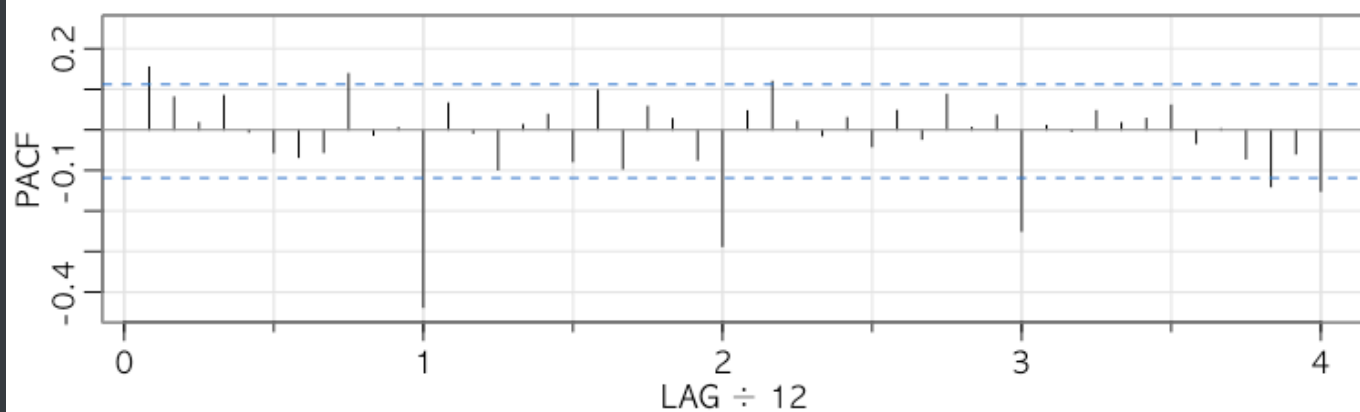
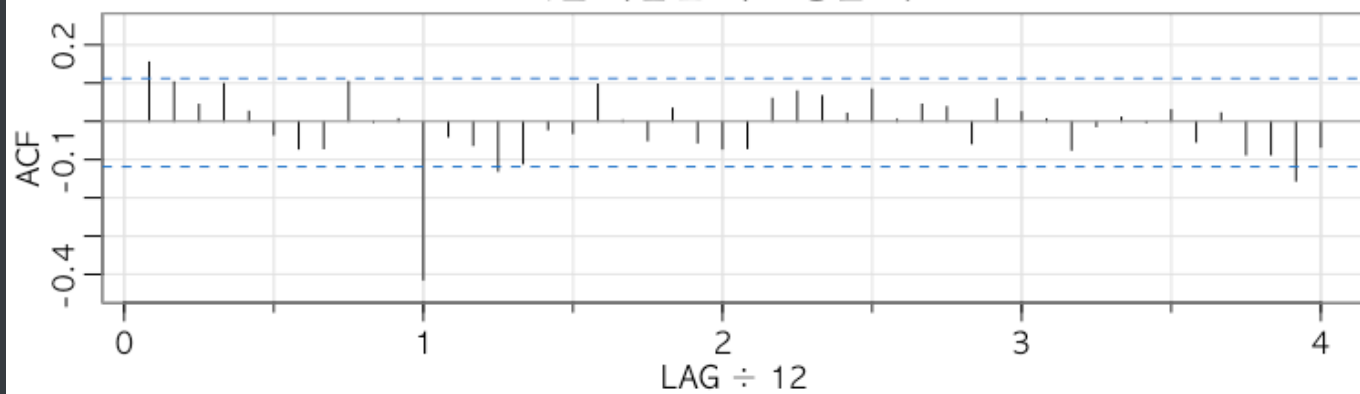
서울과 제주의 평균 기온 계절 차분 후 시계열 그림



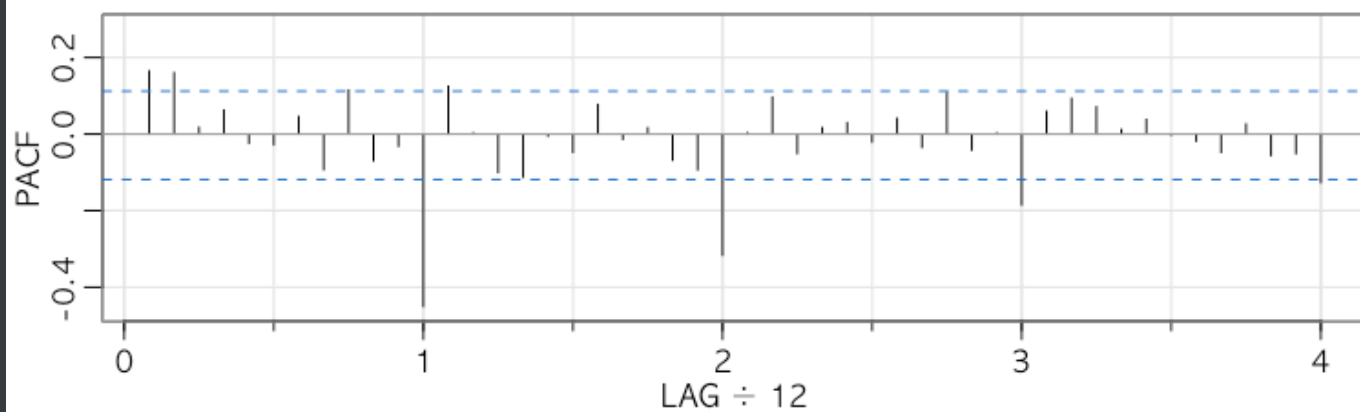
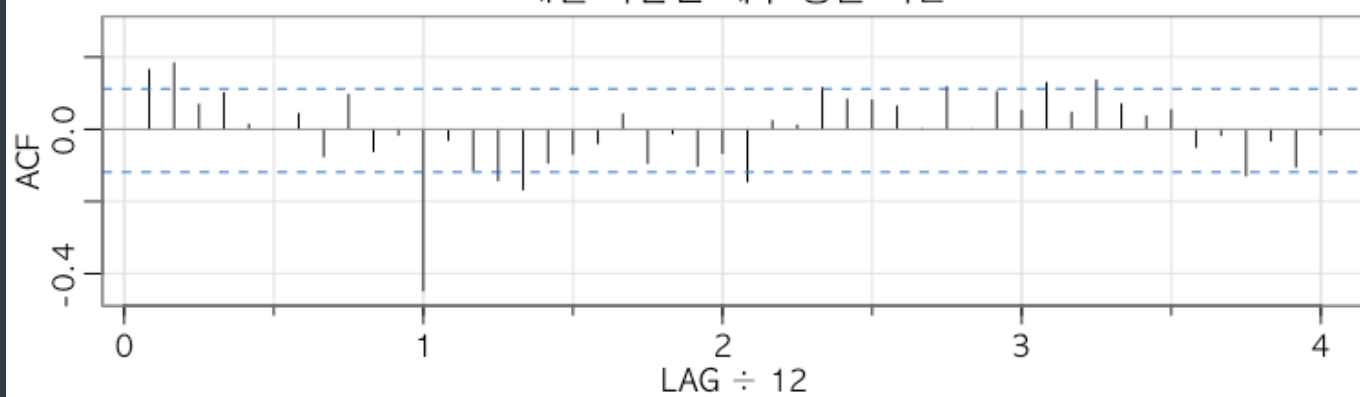
앞서 그림 1 보다 정상성을 확보한 것으로 보인다. 계절 차분된 시계열의 정상성 여부를 판단하기 위해 단위근 검정을 실시하였다. 서울의 경우, `adf.test` 를 이용하여 lag 가 0인 경우에 대해 Dicky fuller 단위근 검정을 실시한 결과, p 값이 0.01 로 나타나 귀무가설을 채택하였다. 즉, 시계열이 정상 시계열이 아니다. 이에 따라 SACF, SPACF 패턴을 함께 확인하였다.

각각 그림 3과 그림 4는 계절 차분된 서울 평균 기온에 대한 SACF, SPACF 패턴, 제주 평균 기온에 대한 SACF, SPACF 패턴이다.

계절 차분된 서울 평균 기온

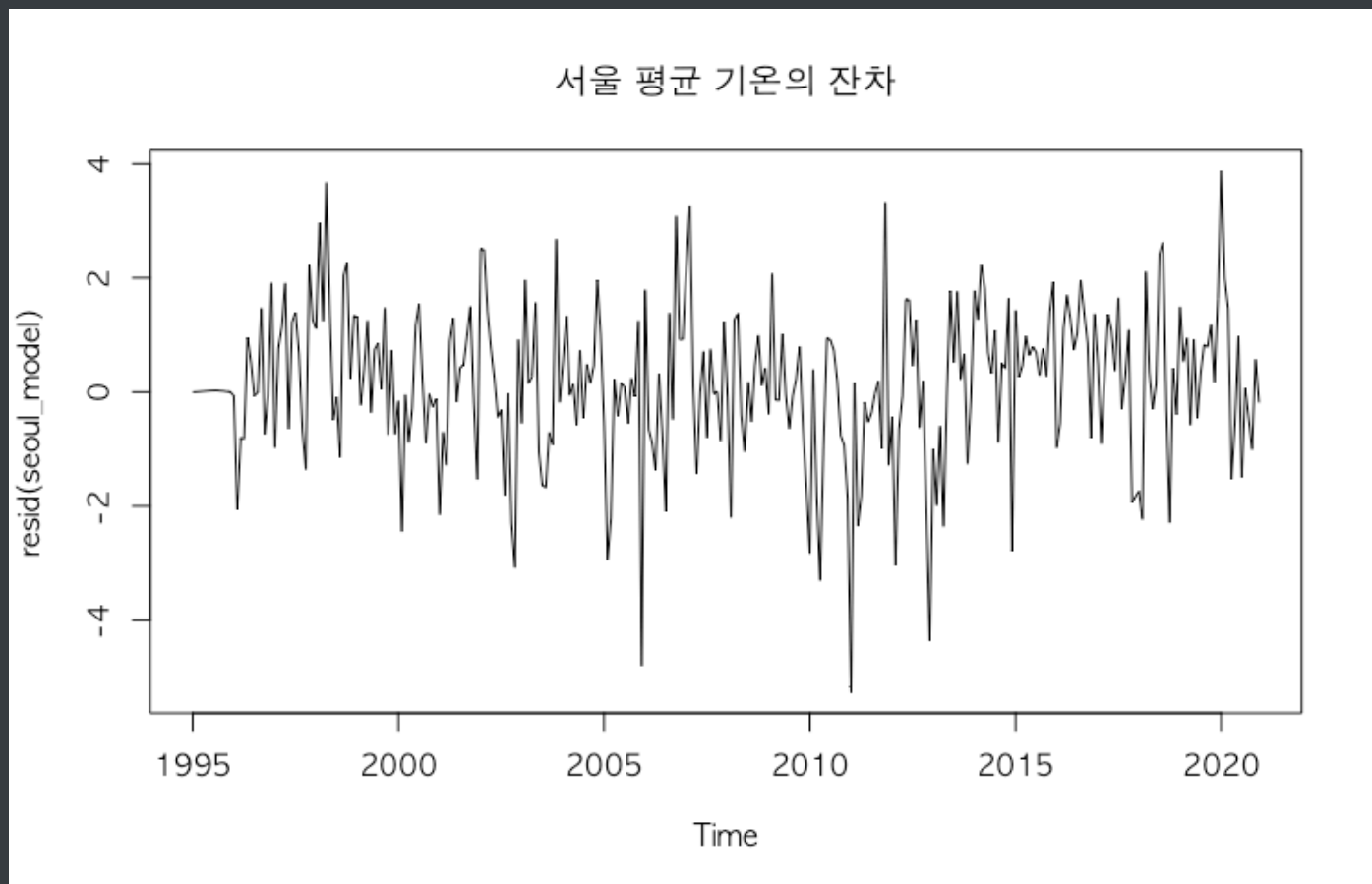


계절 차분된 제주 평균 기온

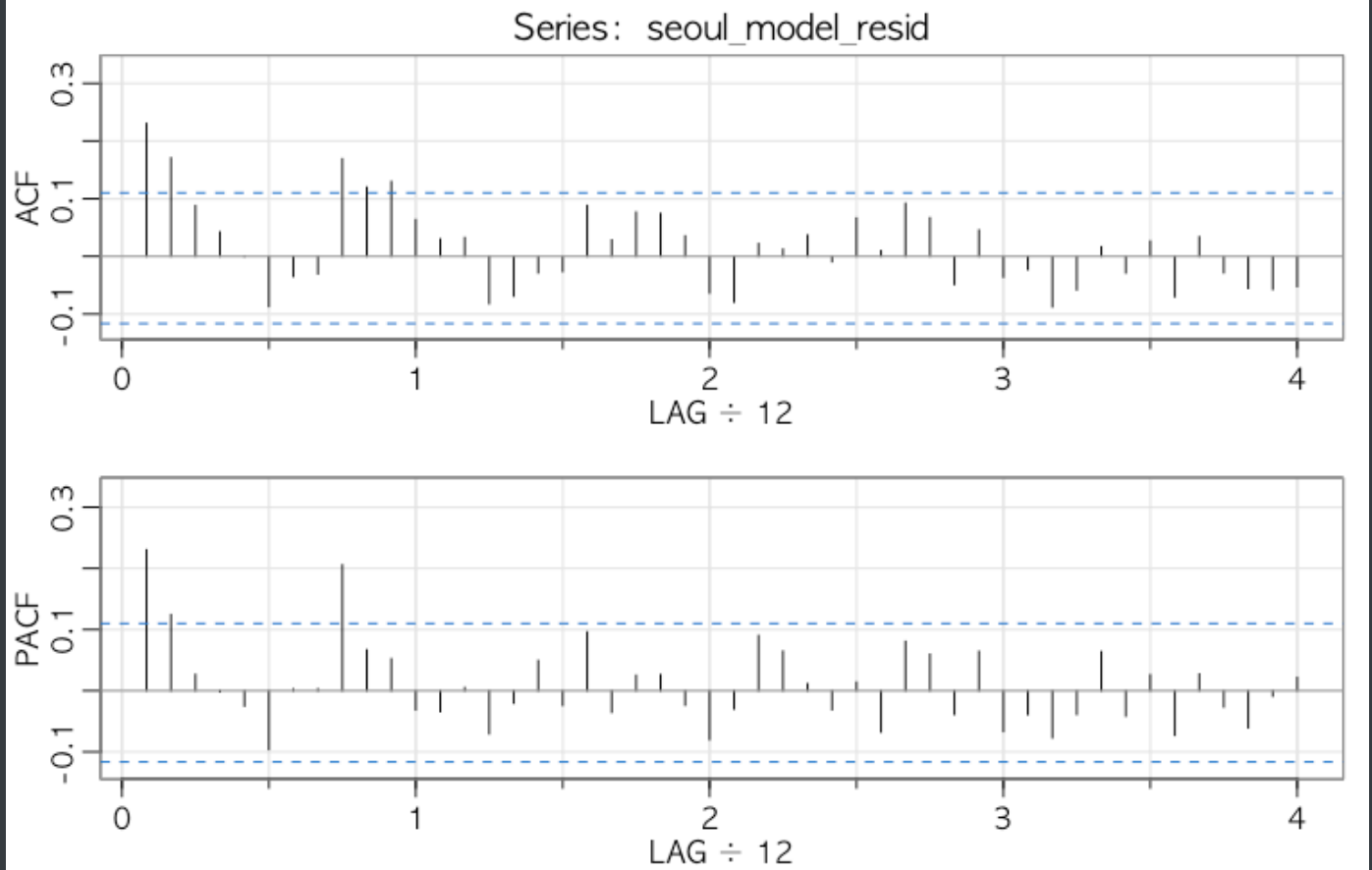


두 시계열 모두 PACF 의 12, 24, 36 에서 지수적으로 감소하는 패턴과 SACF 의 lag 12 에서 유의하게 높은 모양을 확인할 수 있다. 이는 교재의 10장 계절형 ARIMA 모형의 이론적 형태에 따라 추정해볼 수 있는데, 계절 성분 MA(1) 과정을 의심해볼 수 있다. (교재 그림 10-2)

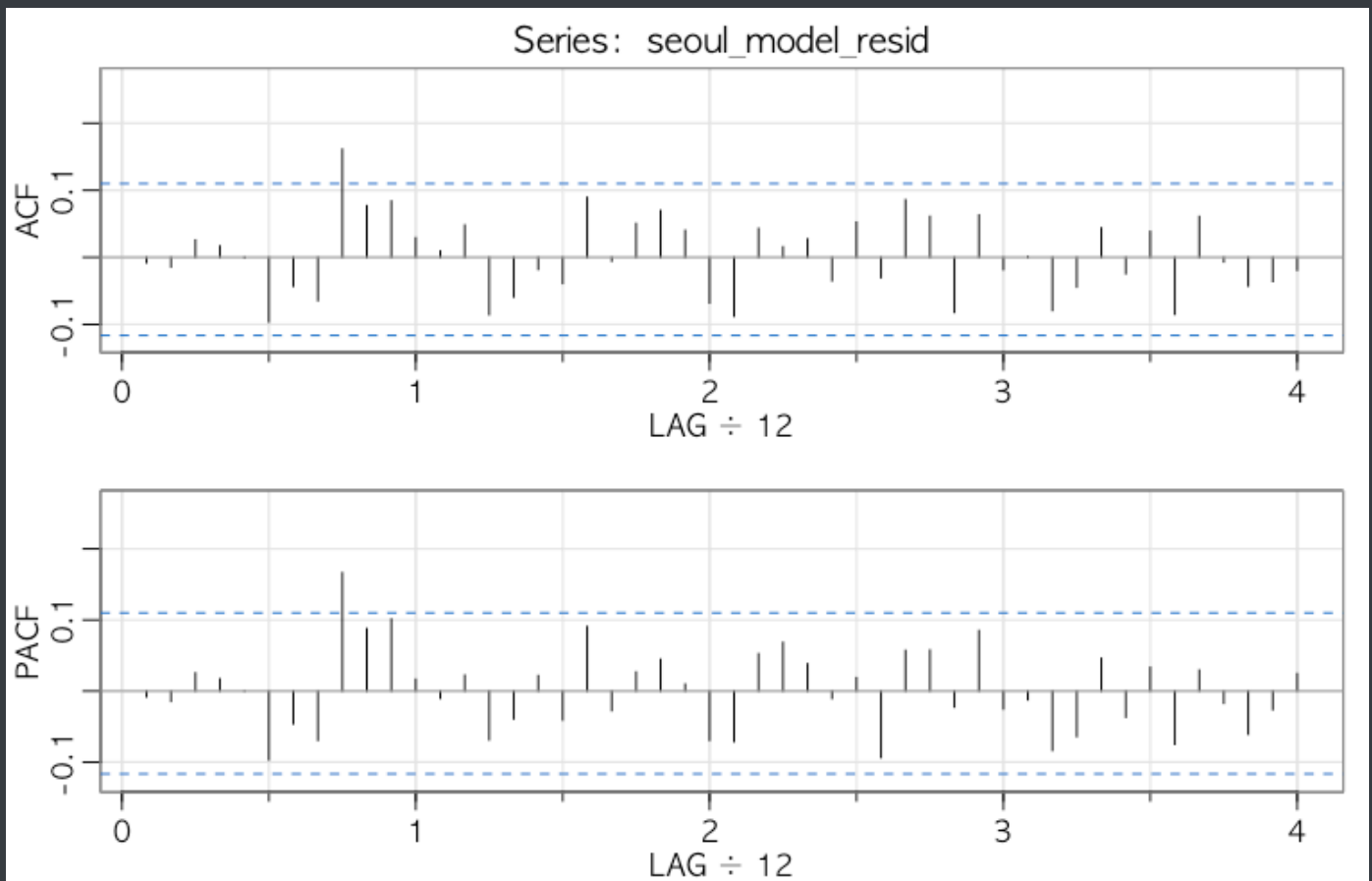
따라서 ARIMA 모형을 이용하여 $ARIMA(0, 0, 0)(0, 1, 1)_{12}$ 모형을 적합하였다. 서울의 결과는 다음과 같다.



잔차에 대해 포트 맨토 검정을 실시한 결과, lag 가 6, 12, 24 인 케이스에 대해 모두 p value 가 유의수준 0.05 보다 작게 나왔다. 즉, 해당 모형의 잔차는 백색잡음으로 볼 수 없다. RSACF와 RSPACF를 확인하였다.

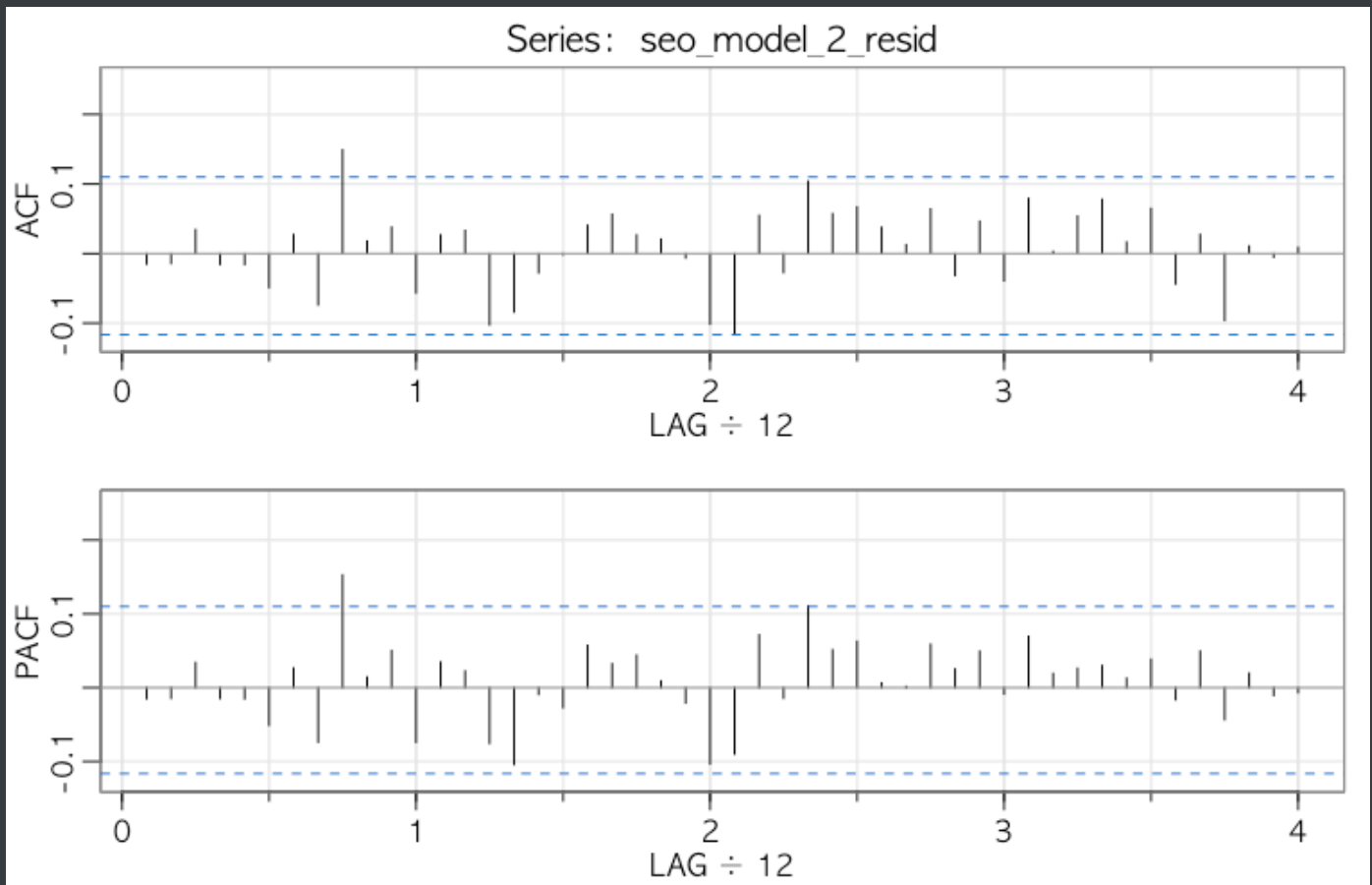


아직 잡아내지 못한 AR(2) 패턴이 남아있는 것으로 확인된다. 다시 모델을 $ARIMA(2, 0, 0)(0, 1, 1)_{12}$ 로 수정하여 적합을 진행하였다.



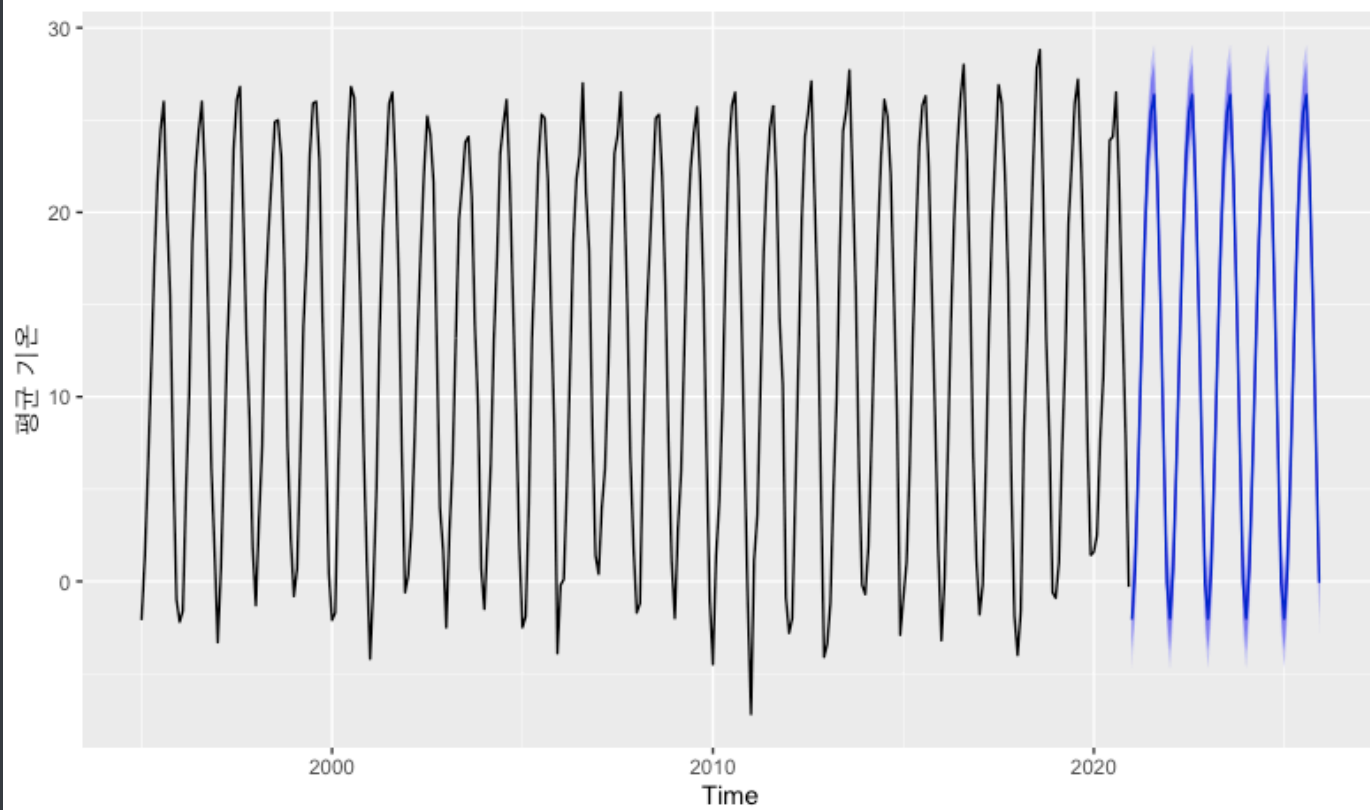
잔차가 상당히 안정적인 모습을 보이고 있고 Dicky Fuller 테스트의 결과, lag 가 6, 12, 24 모두에서는 백색잡음과 다르지 않았다. RSACF의 10 지점에서 치솟는 파트가 있기는 하지만 충분히 잘 적합된 모형으로 판단하여 추가적인 모수를 고려하지는 않았다. 모수: $\phi_1 : 0.2078, \phi_2 : 0.1339, O_1 : -0.9499$

서귀포시에 대해서도 동일한 패턴일 것임을 가정하여 $ARIMA(2, 0, 0)(0, 1, 1)_{12}$ 모델을 적합하였다. 모수: $\phi_1 : 0.1859, \phi_2 : 0.1859, O_1 : -1.00$

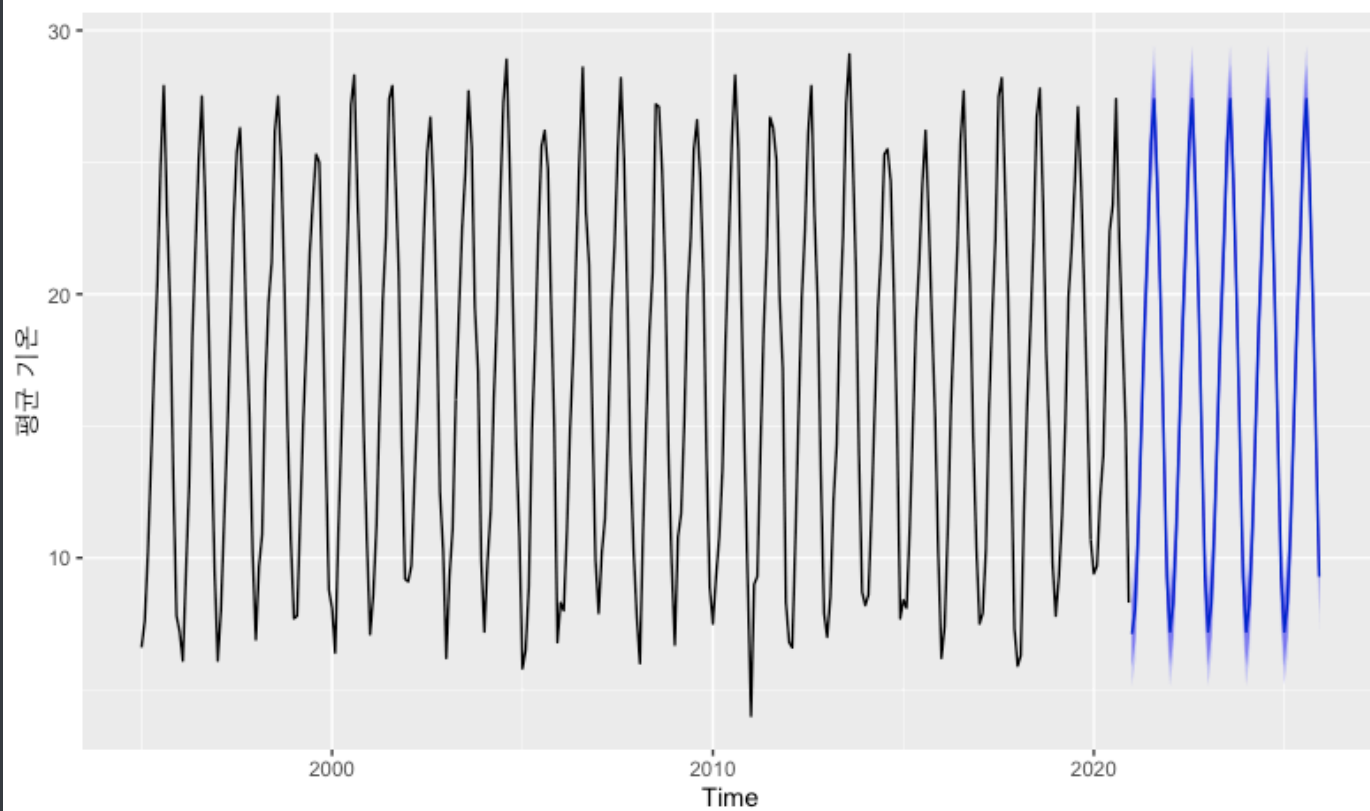


마찬가지로 lag 10에서 다소 치솟는 경향이 있으나 서울과 유사한 패턴으로 판단되어 추가 모델링을 진행하지는 않았다. 모델링의 결과를 바탕으로 5개년, 즉 60개월의 온도를 예측하였다.

서울의 평균 기온 예측



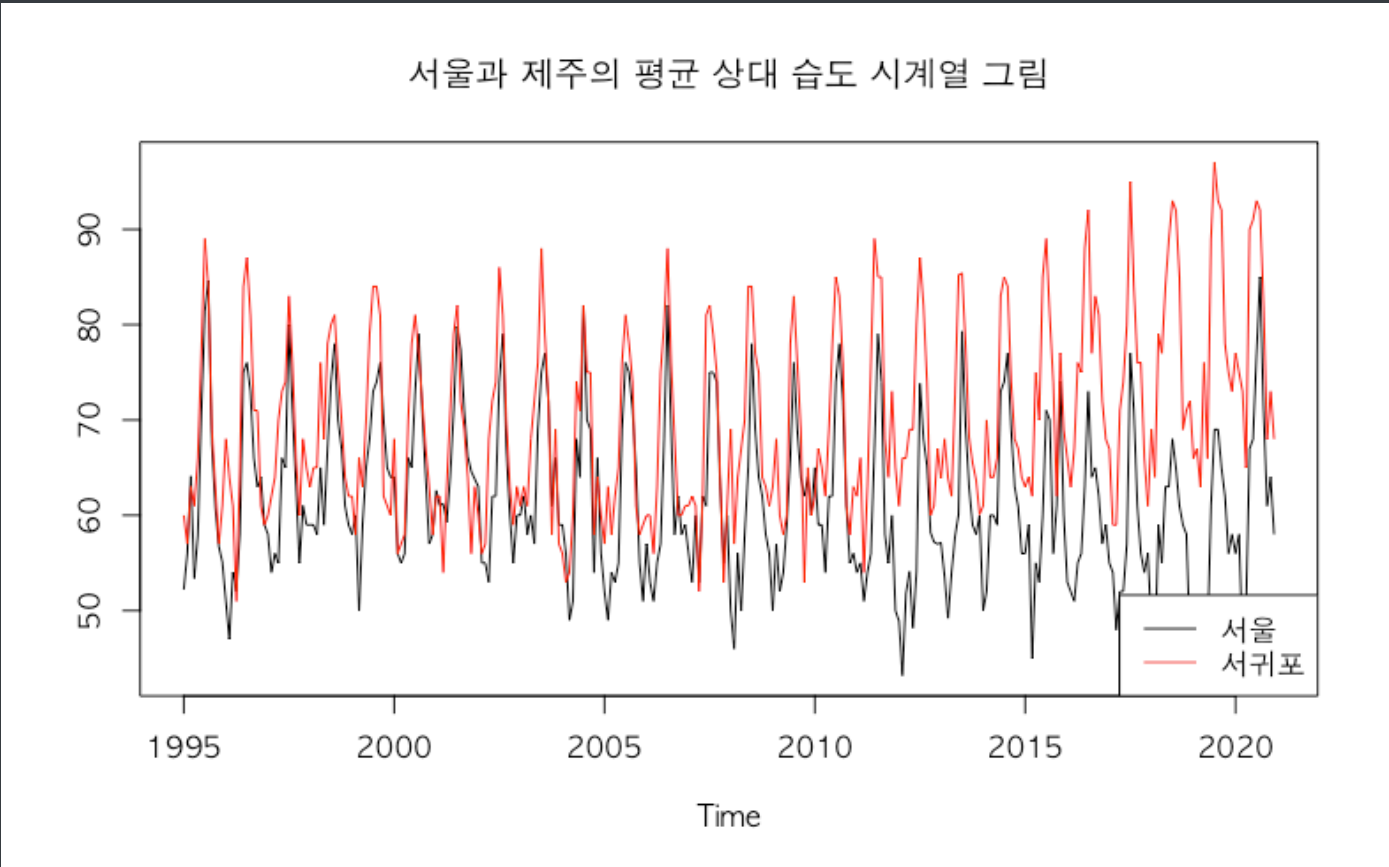
서귀포의 평균 기온 예측



두 지역을 예측하였을 때, 특별히 선형 추세는 확인되지 않아 예측된 값들은 과거의 데이터와 유사한 패턴으로 등장하는 것을 알 수 있었다. 즉 본 프로젝트에서 사용한 데이터로는 지구 온난화의 영향이 확인되지 않았다.

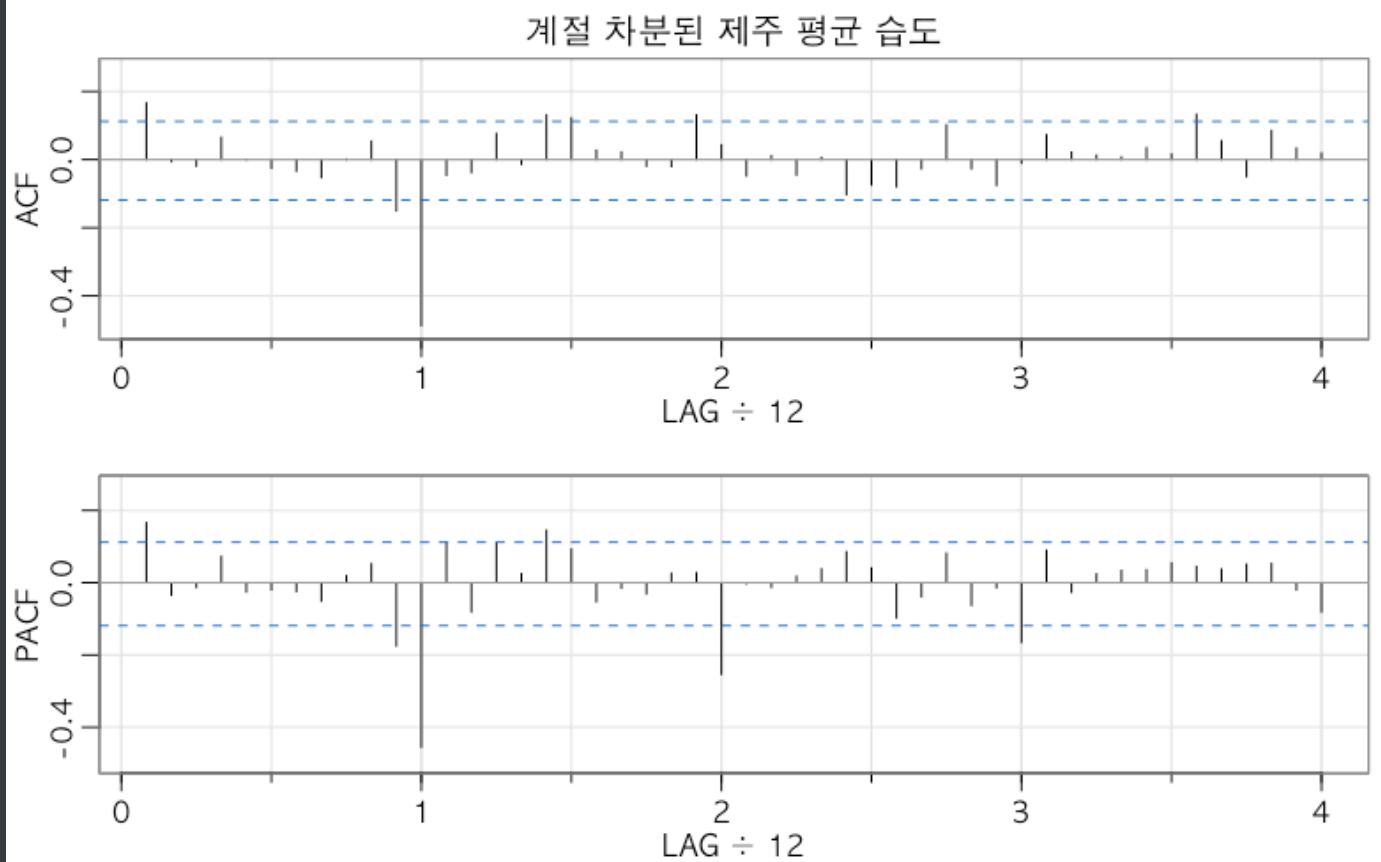
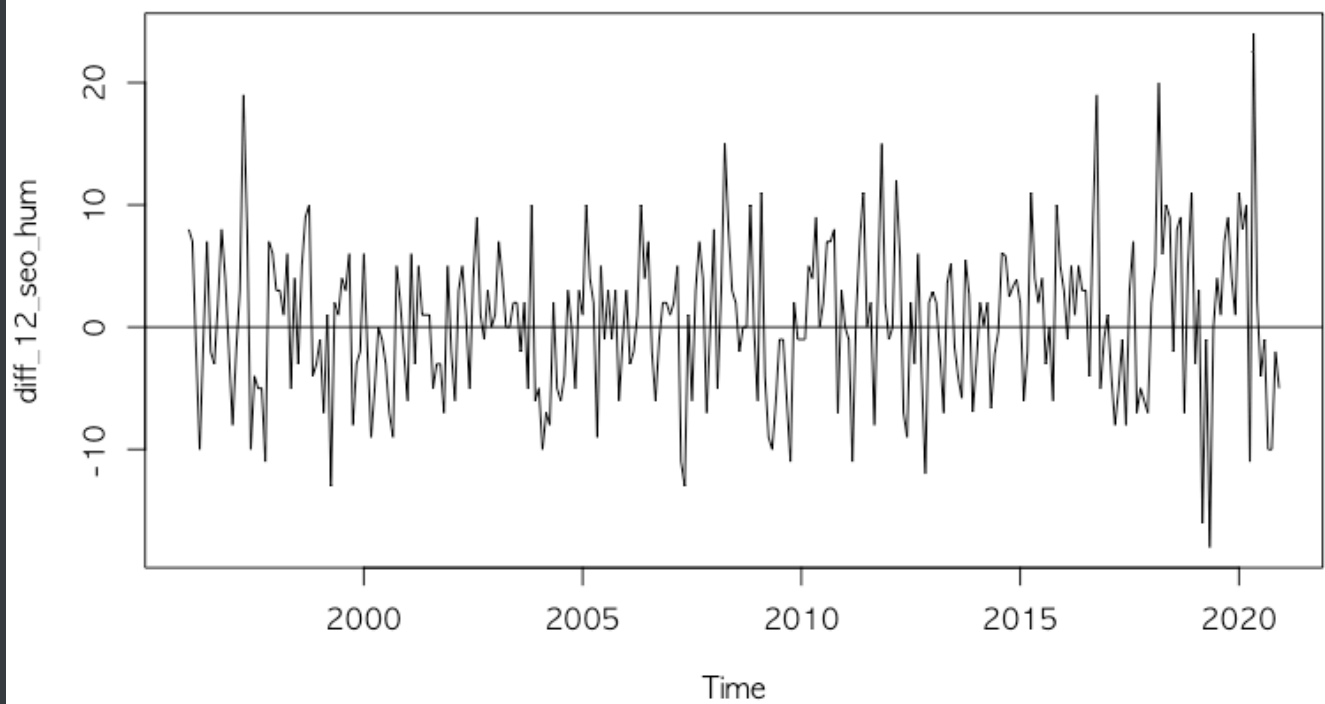
습도

다음 질문은 서귀포시는 아열대화되고 있는가에 대한 질문이다. 아열대는 고온 다습한 지역의 기후를 의미하며 제주도는 기존에 온대 지방으로 알려져 있었다. 습도에 대한 시계열 그래프는 다음과 같다.

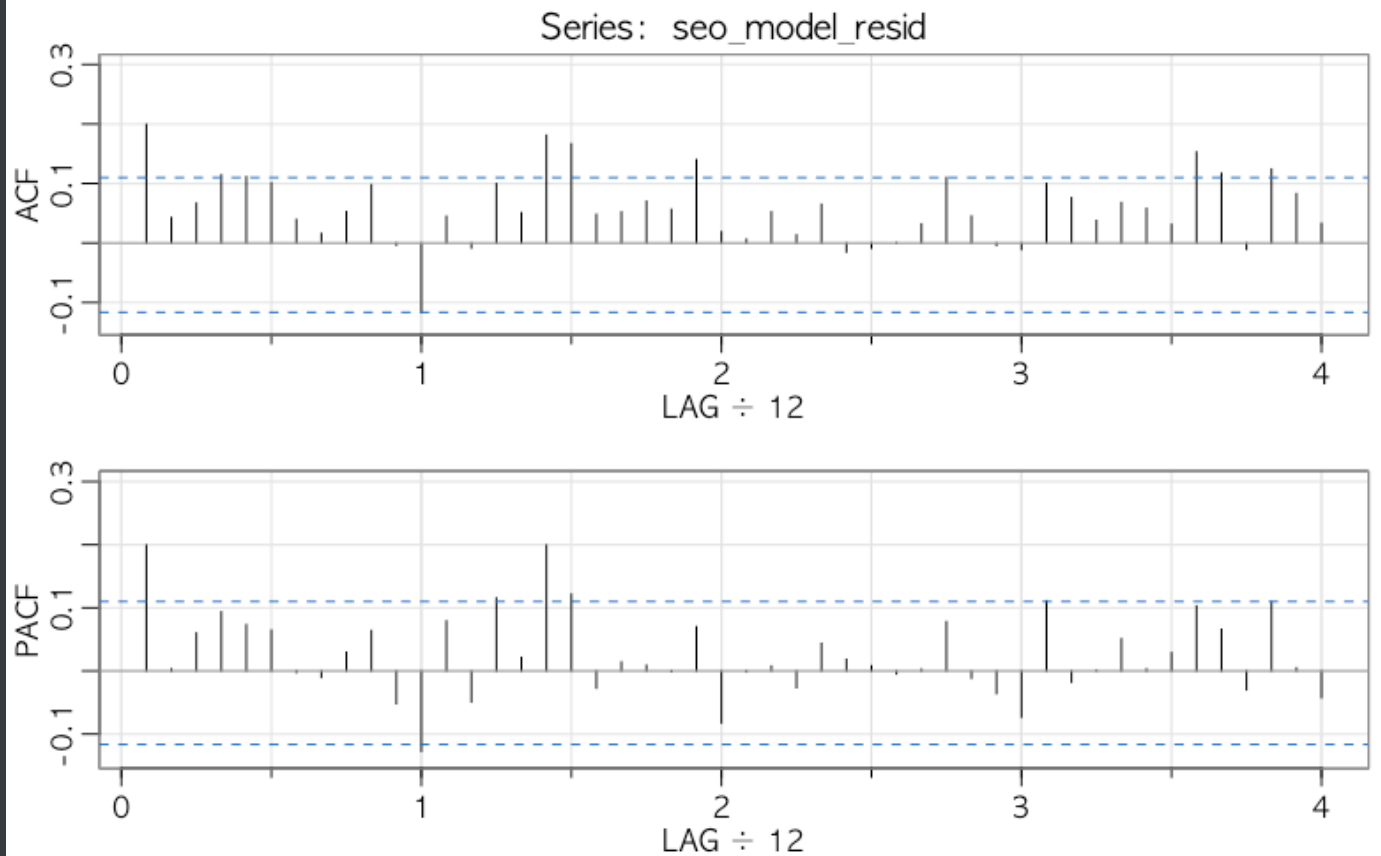


기온의 시계열 그래프와 동일하게 서울과 서귀포를 비교하였다. 이때 서울의 그래프에 비해 제주의 그래프에서 다른 추세가 보이는 것을 확인할 수 있다. 특히 2015년을 전후로 증가하는 양상을 보이고 있는 듯 하다. 개입 분석 등을 고려해볼 수 있을 것이다. 서울의 습도는 61.40 (8.48)이었으며 서귀포의 습도는 70.02 (9.91)로 서귀포가 더 습하였다.

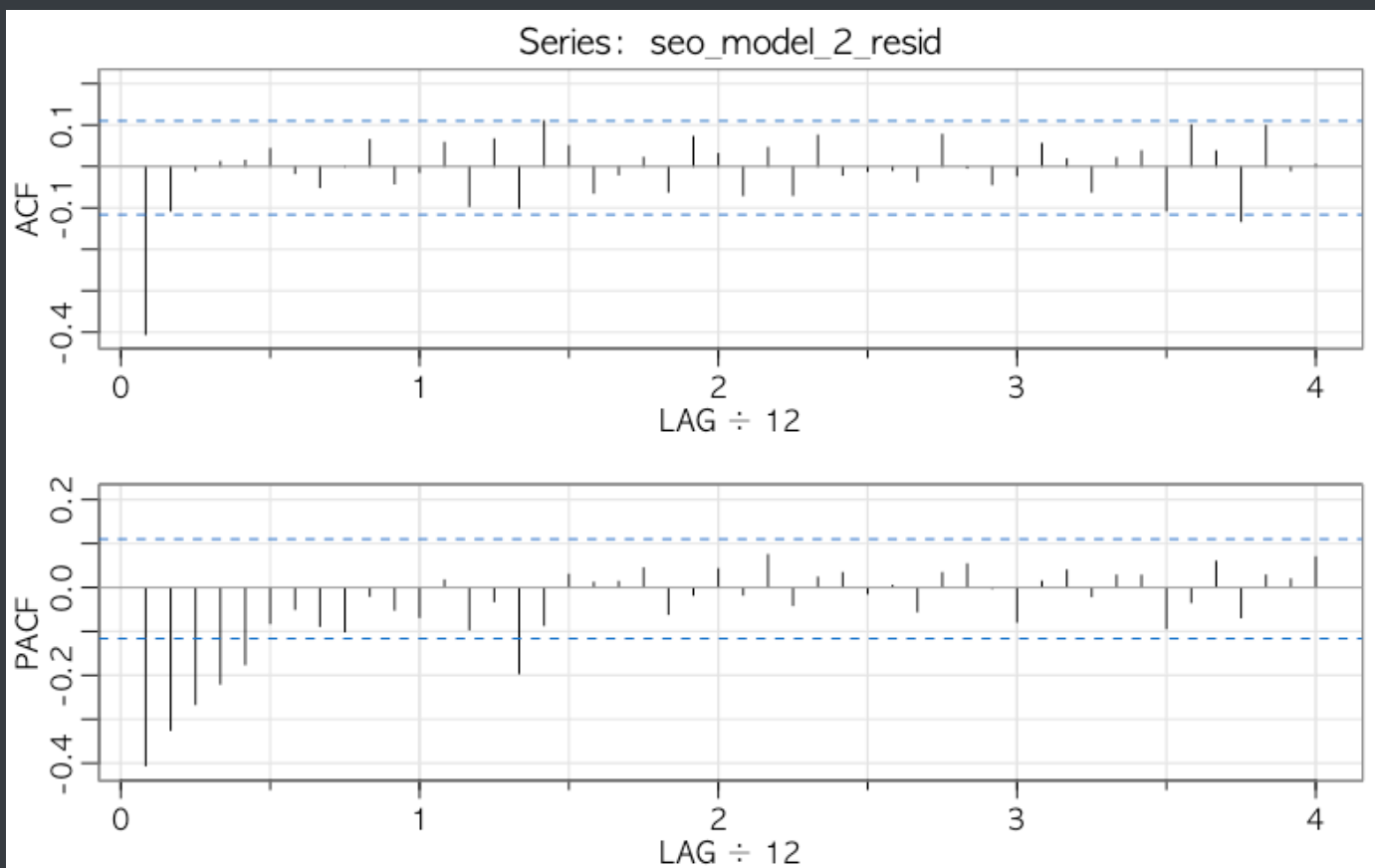
확연한 계절성이 드러나기 때문에 기온과 마찬가지로 습도에서도 계절 차분을 진행하였다. 계절 차분된 이후의 잔차와 RSACF, RSPACF 는 다음과 같다.



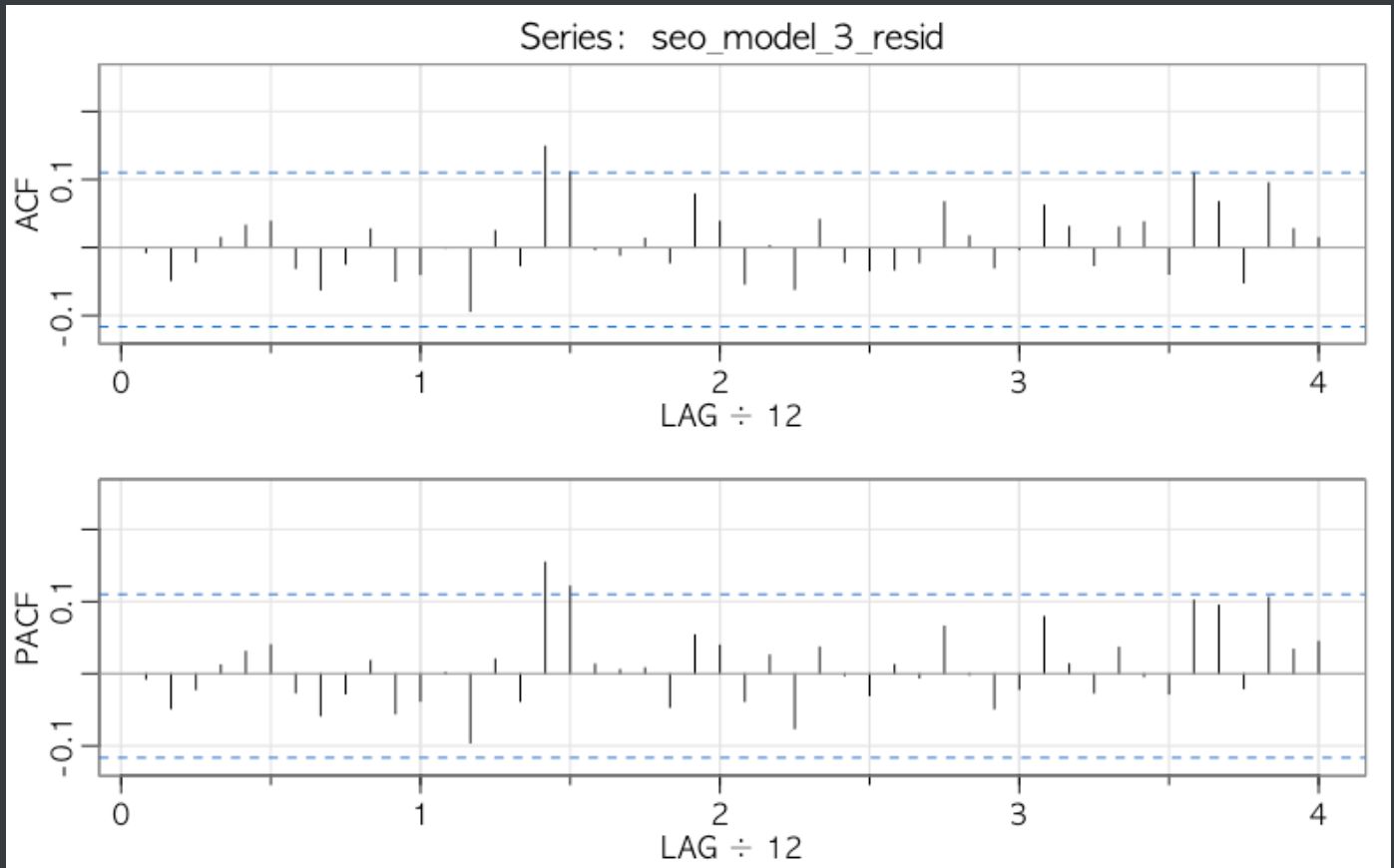
앞서 기온과 유사하게 계절 차원의 MA(1) 가 필요할 듯하다. \$ ARIMA(0, 0, 0)(0, 1, 1)_{12} \$ 을 적합 해본다.



RSACF 에서 뭉쳐있는 모양을 확인할 수 있었다. 1차 차분을 고려해볼만하다고 판단하였다.

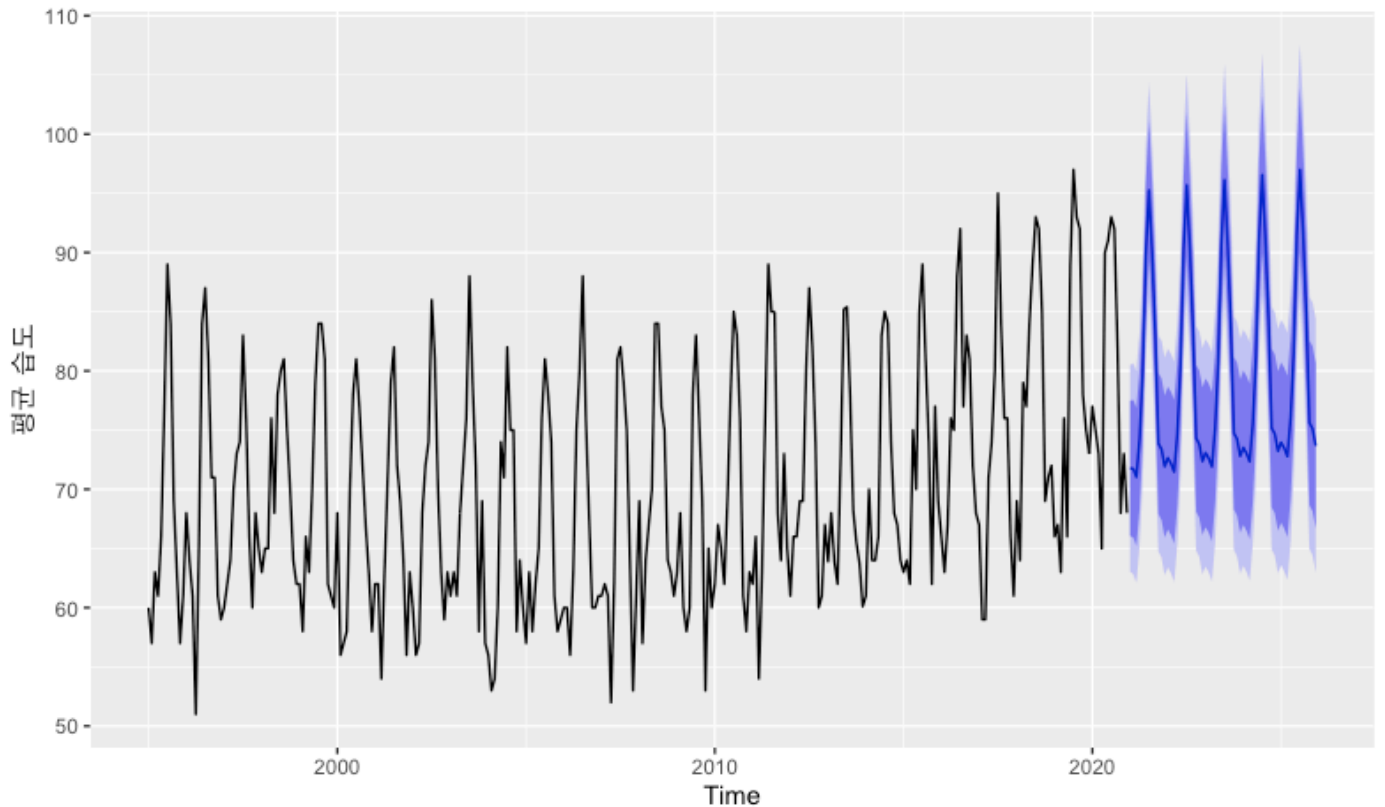


RSPACF 가 천천히 떨어지는 모양, RSACF 2에서 절단되는 모양이 식별되었다. 추가적으로 MA(1)를 추가로 도입한다. 최종 모델인 $ARIMA(0, 1, 1)(0, 1, 1)_{12}$ 의 적합 결과는 다음과 같았다. 잔차가 많이 안정된 모습이다. 포트 맨토 검사에서도 lag가 6, 12, 24인 지점 모두 백색잡음과 다르지 않다고 확인하였다. 모수: $\theta_1 : -0.8055, \theta_2 : -0.1145, O_1 : -1.00$.



이에 따라 최종적으로 모델을 바탕으로 예측한 값을 도출하였다. 앞서 기온과 마찬가지로 5개년의 습도를 예측하도록 하였다. 이번에는 앞서와 달리 습도가 해가 지날수록 상승하는 경향성이 확인되었다.

서귀포의 평균 습도 예측



최종 모델을 바탕으로 outlier detection 을 추가로 진행하였다. 이때 흥미롭게도 Level shift 가 발견되었다.

```

1 Call:
2 list(method = NULL)
3
4 Coefficients:
5      ma1      sma1      A040      A0262      LS279      A0293
6    -0.9382  -1.0000  12.5128  13.5767  6.0478 -14.7138
7 s.e.   0.0210   0.0625   4.1187   4.1294   1.5107   4.1297
8
9 sigma^2 estimated as 16.81:  log likelihood = -867.32,  aic = 1748.64
10
11 Outliers:
12   type ind time coefhat  tstat
13  1  A0  40   40  12.513  3.038
14  2  A0 262  262  13.577  3.288
15  3  LS 279  279   6.048  4.003
16  4  A0 293  293 -14.714 -3.563
    
```

LS는 279 시점, 즉 2018년 3월부터 무언가의 변화가 발생했음을 알 수 있다.

논의 및 결론

본 프로젝트는 월 단위로 수집된 데이터를 기반으로 서울과 서귀포시 지역의 온도 변화, 습도 변화 패턴을 탐지해보았다. 자료가 1995년부터 2020년까지 확보할 수 있어 그 이후의 데이터는 확인하지 못한 아쉬움이 있다. 특히 올해는 가뭄 등의 특징적인 날씨 패턴이 등장하였기에 아마 기존과는 다른 패턴을 보여줬을 것이라 생각되지만 본 프로젝트에서는 다루지 못하였다.

기온에서는 특별한 변화가 있다고 판단하지 못하였으나, 서귀포시의 습도 변화가 주목할만했다. 서귀포시의 습도는 바닷가 지방이라 원래도 높은 편이었으나 2018년 이후 높아지는 경향을 확인할 수 있었고, 이는 outlier detection 에서도 잡혔다. 지금 단계에서 원인은 알 수 없으니 제주도의 기후가 습해지고 있다는 것이 확인되었다.

물론 월 단위의 평균 값만을 가지고 분석하였기 때문에 많은 정보들이 누락되어 있을 것이다. 또한 단순히 평균값만을 사용하였기 때문에 특정 패턴이 잘 드러나지 않았을 것이라는 판단도 있다. 추가적으로 진행한 기온의 평균 최저점, 기온의 평균 최고점에서도 특별한 패턴이 드러나지는 않았다. 이는 각각 최저점과 최고점들을 모아서 평균을 낸 값을 의미하는데, 본 프로젝트에서 진행한 평균 기온과 크게 차이가 있는 것 같지 않아 다루지는 않았다.

습도의 증가 패턴을 추가적으로 제주도 제주시에서도 드러나는 패턴인지 확인할 필요성이 있다. 앞서 언급한 것처럼 제주도의 날씨는 한라산을 중심으로 크게 다르기 때문이다. 만약 제주시에서도 유사한 패턴이 발견된다면 이것은 제주도 전체의 기후 변화로 볼 수 있을 것이다. 실제로 제주도에서는 아열대화를 대비하여 농사 작물 다변화를 하고 있는데, 본 데이터에서 확인된 바로는 적절한 대책이라 할 수 있겠다.

참고 문헌

- [KOSIS](#)
- [서울열린데이터광장](#)
- [공공데이터포털](#)
- [기후변화 홍보포털](#)
- [제주 바다숲 사라져가고... "아열대 물고기 절반"](#)
- [제주섬, 아열대 기후 진입... 환경/생태 급변](#)
- [SAS/ETS와 R을 이용한 시계열 분석 5판](#)

- 한양대학교 통계학과 최정순 교수님 시계열분석 강의자료

Appendix: R code

```
1  rm(list=ls())
2
3  setwd("~/Workspace/2022-Fall_TimeSeriesAnalysis/final_project/")
4  par(family="AppleGothic")
5  library(astsa)
6  library(tidyverse)
7  library(tseries)
8  library(forecast)
9  library(tsoutliers)
10
11 data = read.csv("weather_data.csv")
12
13 seoul_df = data %>% filter(region == 0) %>% select(-c(region))
14 jeju_df = data %>% filter(region == 1) %>% select(-c(region))
15 seogwipo_df = data %>% filter(region == 2) %>% select(-c(region))
16
17 # 기온
18 seoul_temp_mean = ts(seoul_df$temp_mean, start=c(1995, 1), frequency=12)
19 seo_temp_mean = ts(seogwipo_df$temp_mean, start=c(1995, 1), frequency=12)
20
21 ts.plot(seoul_temp_mean, seo_temp_mean,
22         main="서울과 제주의 평균 기온 시계열 그림",
23         gpars=list(col=c("black", "red")))
24 legend("bottomright", legend=c("서울", "서귀포"), col=1:2, lty=1)
25
26 mean(seoul_temp_mean); sd(seoul_temp_mean)
27 mean(seo_temp_mean); sd(seo_temp_mean)
28
29 cor(seoul_df$temp_mean, seogwipo_df$temp_mean)
30
31 diff_12_seoul_temp = diff(seoul_temp_mean, lag=12)
```



```

32 diff_12_seo_temp = diff(seo_temp_mean, lag=12)
33
34 ts.plot(diff_12_seoul_temp, diff_12_seo_temp,
35         main="서울과 제주의 평균 기온 계절 차분 후 시계열 그림",
36         gpars=list(col=c("black", "red")))
37 legend("bottomright", legend=c("서울", "제주"), col=1:2, lty=1)
38
39 adf.test(diff_12_seoul_temp, k=0)
40 adf.test(diff_12_seo_temp, k=0)
41
42 acf2(diff_12_seoul_temp, main="계절 차분된 서울 평균 기온")
43 acf2(diff_12_seo_temp, main="계절 차분된 제주 평균 기온")
44
45 seoul_model = arima(seoul_temp_mean, order=c(2, 0, 0),
46                    seasonal=list(order=c(0, 1, 1), period=12))
47 seoul_model_resid = resid(seoul_model)
48 ts.plot(seoul_model_resid, main="서울 평균 기온의 잔차")
49
50 Box.test(seoul_model_resid, lag=6)
51 Box.test(seoul_model_resid, lag=12)
52 Box.test(seoul_model_resid, lag=24)
53 acf2(seoul_model_resid)
54
55 seoul_model_2 = arima(seoul_temp_mean, order=c(2, 0, 0),
56                      seasonal=list(order=c(0, 1, 1), period=12))
57 seoul_model_2_resid = resid(seoul_model_2)
58 ts.plot(seoul_model_2_resid, main="서울 평균 기온의 잔차")
59 Box.test(seoul_model_2_resid, lag=6)
60 Box.test(seoul_model_2_resid, lag=12)
61 Box.test(seoul_model_2_resid, lag=24)
62 acf2(seoul_model_2_resid)
63
64 seo_model_2 = arima(seo_temp_mean, order=c(2, 0, 0),
65                    seasonal=list(order=c(0, 1, 1), period=12))
66 seo_model_2_resid = resid(seo_model_2)
67 ts.plot(seo_model_2_resid, main="서귀포 평균 기온의 잔차")
68 Box.test(seo_model_2_resid, lag=6)

```

```

69 Box.test(seo_model_2_resid, lag=12)
70 Box.test(seo_model_2_resid, lag=24)
71 acf2(seo_model_2_resid)
72
73 autoplot(forecast(seoul_model_2, h=60)) +
74   labs(x="Time", y="평균 기온", title="서울의 평균 기온 예측") +
75   theme(axis.title.y=element_text(family="AppleGothic"),
76         plot.title=element_text(family="AppleGothic"),
77         plot.subtitle=element_text(family="AppleGothic"))
78
79 autoplot(forecast(seo_model_2, h=60)) +
80   labs(x="Time", y="평균 기온", title="서귀포의 평균 기온 예측") +
81   theme(axis.title.y=element_text(family="AppleGothic"),
82         plot.title=element_text(family="AppleGothic"),
83         plot.subtitle=element_text(family="AppleGothic"))
84
85 # 습도
86 seoul_hum_mean = ts(seoul_df$humidity, start=c(1995, 1), frequency=12)
87 seo_hum_mean = ts(seogwipo_df$humidity, start=c(1995, 1), frequency=12)
88
89 ts.plot(seoul_hum_mean, seo_hum_mean,
90         main="서울과 제주의 평균 상대 습도 시계열 그림",
91         gpars=list(col=c("black", "red")))
92 legend("bottomright", legend=c("서울", "서귀포"), col=1:2, lty=1)
93
94 mean(seoul_hum_mean); sd(seoul_hum_mean)
95 mean(seo_hum_mean); sd(seo_hum_mean)
96
97 diff_12_seo_hum = diff(seo_hum_mean, lag=12)
98 ts.plot(diff_12_seo_hum)
99 abline(h=0)
100 acf2(diff_12_seo_hum, main="계절 차분된 제주 평균 습도")
101
102 seo_model = arima(seo_hum_mean, order=c(0, 0, 0),
103                  seasonal=list(order=c(0, 1, 1), period=12))
104 seo_model_resid = resid(seo_model)
105 ts.plot(seo_model_resid, main="서귀포 평균 습도의 잔차")

```

```
106 acf2(seo_model_resid)
107
108 seo_model_2 = arima(seo_hum_mean, order=c(0, 1, 0),
109                     seasonal=list(order=c(0, 1, 1), period=12))
110 seo_model_2_resid = resid(seo_model_2)
111 ts.plot(seo_model_2_resid, main="서귀포 평균 습도의 잔차")
112 acf2(seo_model_2_resid)
113
114 seo_model_3 = arima(seo_hum_mean, order=c(0, 1, 1),
115                     seasonal=list(order=c(0, 1, 1), period=12))
116 seo_model_3_resid = resid(seo_model_3)
117 ts.plot(seo_model_3_resid, main="서귀포 평균 습도의 잔차")
118 acf2(seo_model_3_resid)
119
120 Box.test(seo_model_3_resid, lag=6)
121 Box.test(seo_model_3_resid, lag=12)
122 Box.test(seo_model_3_resid, lag=24)
123
124 autoplot(forecast(seo_model_3, h=60)) +
125   labs(x="Time", y="평균 습도", title="서귀포의 평균 습도 예측") +
126   theme(axis.title.y=element_text(family="AppleGothic"),
127         plot.title=element_text(family="AppleGothic"),
128         plot.subtitle=element_text(family="AppleGothic"))
129
130 outlier = tso(ts(seo_hum_mean), types=c("AO", "IO", "LS"),
131              tsmethod="arima", cval=3.0,
132              args.tsmethod=list(order=c(0, 1, 1),
133                                seasonal=list(order=c(0, 1, 1), period=12)))
133 outlier
```