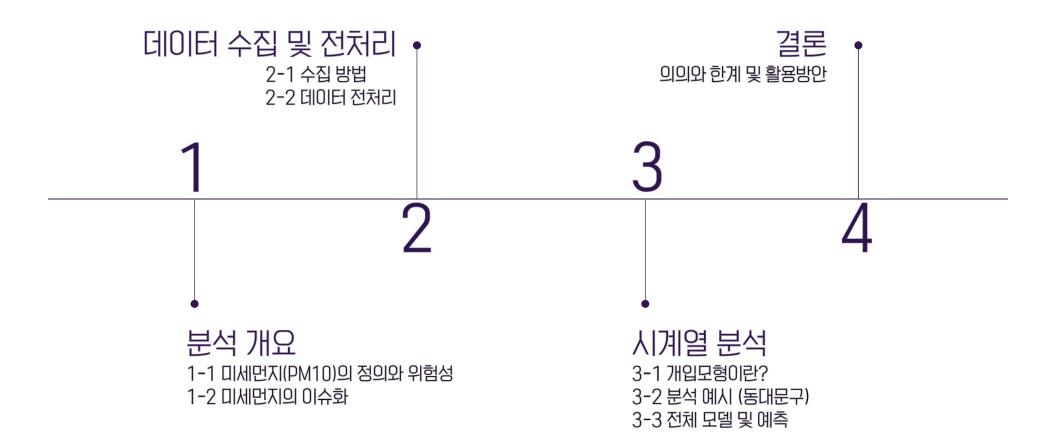


목차



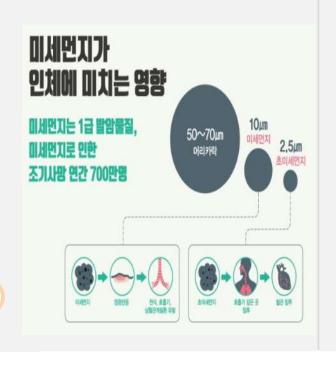
1.개요

1-1 미세먼지(PM10)의 정의와 위험성



· のかなり(PM10)を?

입자의 크기가 지름 10㎞ 이하인 대기오염물질 중 하나.





출처: 환경부

1.개요

1-2 미세먼지의 이슈화

기가간바스보신사 연합뉴스

수도권 미세먼지 치솟아…비상저감조치, 이틀 연속 시행될 듯 (본문문기 👶 설정

기사입력 2018-01-17 15:46 기사원문









미세먼지 '나쁨'에도 마스크 없이...(서울=연합뉴스) 이재희 기자 = 미세먼지 '나쁨'을 보인 17일 오후 서울 경복궁 수문장

10번중 3번 '오보' … 미세먼지 예보 오락가락

대기농도 '나쁨'에 '보통'으로 발령 … 정확도 70여% 12명 4개조 1일 3명 측정 … 관계기관 "인력 예산부족"

2016년 02월 15일 00:05 월요일

But, 제대로 예측되지 않고 있음



시계열 모델을 통한 미세먼지 예측

출처: 연합뉴스, 인천일보





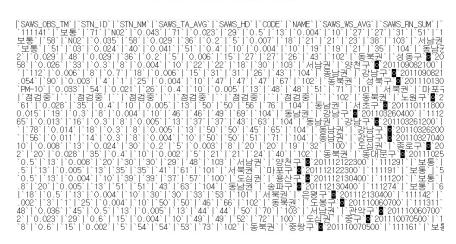
원본데이터

42대기환경정보의 원본데이터로 반출가능

반출가능

서울시 빅데이터 캠퍼스에서 대기환경정보 자료 반출

TV_OP_1HRTMS.txt



-기간: 2008-01-01 00:00 ~ 2017-09-11 11:00 (시간별)

-변수: 측정 날짜, 통합대기환경지수, 미세먼지 농도, 오존 농도, 일산화탄소 농도, 권역명 등

-obs: 21112697H

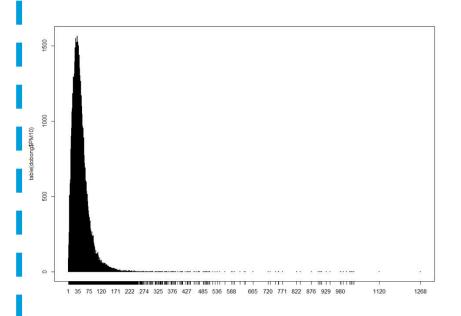
출처: 서울특별시 빅데이터 캠퍼스 웹사이트

2.데이터 수집 및 전처리

2-2 데이터 전처리

- ① 구별로 나누기 【
 - ☑ 강남구.csv
 - ☑a 강동구.csv
 - ☑ 강북구.csv
 - ☑ 강서구.csv
 - 집 관악.csv
 - ☑ 광진구.csv
 - 지 구로.csv
 - 집 금천구.csv
 - ☑ 노원구.csv
 - ☑ 도봉구.csv
 - ☑ 동대문구.csv
 - ☑ 동작구.csv
 - □ 마포구.csv

② **결측치(약 2%) 처리**: single imputation (모든 구가 skewed to right 이므로 median 사용)





```
gangnam<-read.csv("강남구.csv")|
gangnam$PM10[is.na(gangnam$PM10)==TRUE]<-median(gangnam$PM10,na.rm=T)
length(gangnam$PM10[is.na(gangnam$PM10)==TRUE])
write.csv(gangnam, file="C:/Users/Owner/Desktop/gangnam.csv")
```

2.데이터 수집 및 전처리

2-2 데이터 전처리

③ 기간 설정

: 2017-01-01 ~ 2017-09-11

④ 시간별 미세먼지 농도

→일별 평균 미세먼지 농도로 변환

1	Α	В	С	D	Е	F
1	DATE	dobong	dongdaemoon	dongzak	eunpyung	gangseogi
2	2017-01-01	77.3	86	70.8	69.8	75.6
3	2017-01-02	112.5	124.9	91.2	101.1	104.3
4	2017-01-03	93.4	97.6	67.8	78.5	79.9
5	2017-01-04	52.9	53.5	51.3	47.4	60.6
6	2017-01-05	42.1	45.4	34.3	33.7	55.3

```
gangnam<-read.csv("gangnam.csv")</pre>
gangnam<-gangnam[substr(gangnam$MSRDATE,3,4)==17,]
time_index <- seq(from=as.POSIXct("2017-01-01 00:00"),
                   to=as.POSIXct("2017-09-11 11:00"),by="hour")
eventdata <-xts(gangnam[,1],order.by=time_index)
spl<-split(eventdata,month(time_index))</pre>
spll_1<-split(spl$'1',day(time_index))</pre>
list_1<-lapply(spll_1,mean)
spll_2<-split(spl$'2',day(time_index))</pre>
list_2<-lapply(spll_2,mean)
spll_3<-split(spl$'3',day(time_index))</pre>
list_3<-lapply(spll_3,mean)
spll_4<-split(spl$'4',day(time_index))
list_4<-lapply(spll_4,mean)
spll_5<-split(spl$'5',day(time_index))</pre>
list_5<-lapply(spll_5,mean)
spll_6<-split(spl$'6',day(time_index))
list_6<-lapply(spll_6,mean)
spll_7<-split(spl$'7',day(time_index))</pre>
list_7<-lapply(spll_7,mean)
spll_8<-split(spl$'8',day(time_index))</pre>
list_8<-lapply(spl1_8,mean)
spll_9<-split(spl$'9',day(time_index))</pre>
list_9<-lapply(spll_9,mean)
last<-c(unlist(list_1),unlist(list_2),unlist(list_3)</pre>
         ,unlist(list_4),unlist(list_5),unlist(list_6)
         ,unlist(list_7),unlist(list_8),unlist(list_9))
gangnam<-na.omit(last)
dust<-cbind(dobong,dongdaemoon,dongzak,eunpyung,gangbuk,
             gangdong, gangnam, gangseo, gumcheon, guro, gwanak,
             gwanggin, jongro, joong, joongrang, mapo, nowon,
             seocho, seodaemoon, seongbuk, seongdong, songpa,
            yangcheon, yongsan, youngdeungpo)
write.csv(round(dust,1), file="C:/Users/Owner/Desktop/dust.csv")
```

← dust.csv (최종 데이터)
row: 254 col:25

3.시계열 분석 3-1 개입 모형이란?

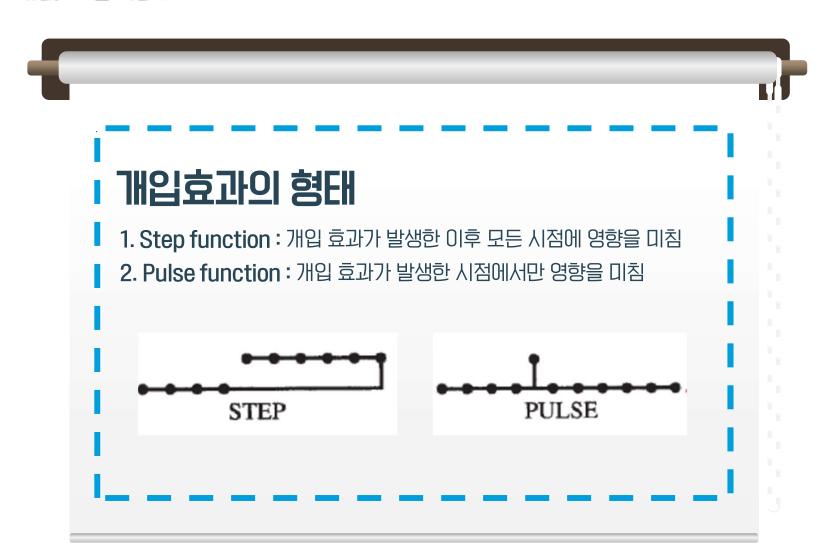


740/0132?

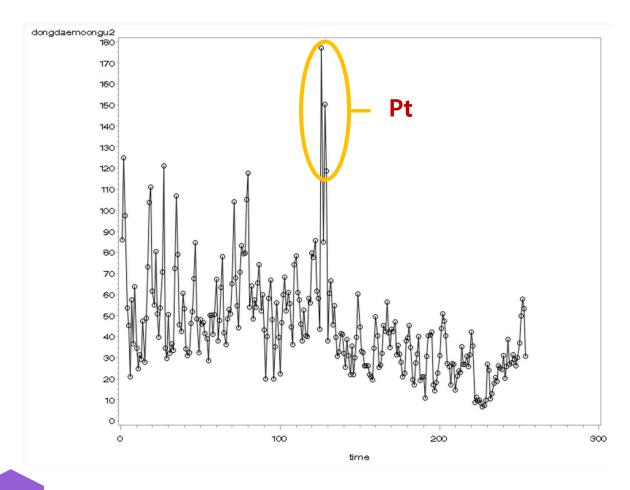
시계열의 변화를 야기시키는 <mark>외부의 충격</mark> 개입으로 인해 시계열의 변동이 생기게 되면 기존의 시계열 분석 방법으로는 <u>만족할만한 결과를 얻을 수 없다</u>



3-1 개입 모형이란?

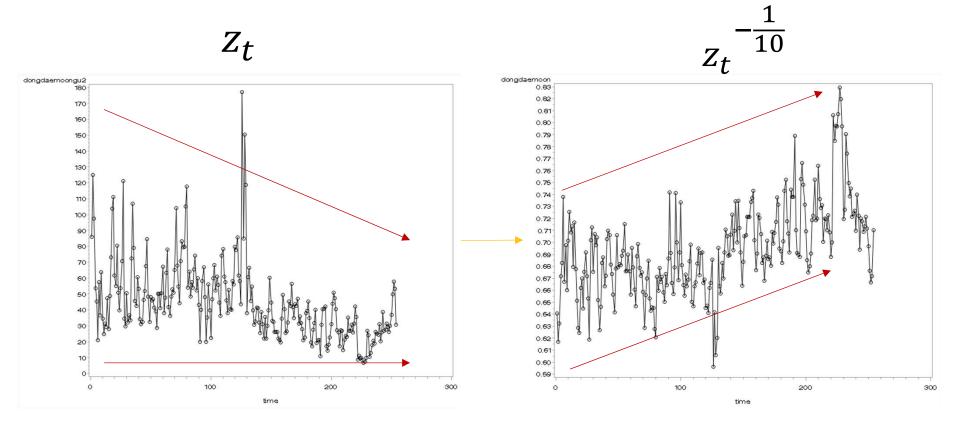


3-2 분석 예시 (동대문구)





3-2 분석 예시 (동대문구)





분산이 갈수록 줄어들어 <mark>분산안정화</mark> 실시

3-2 분석 예시 (동대문구)

Autocorrelation Check for White Noise										
To Lag Chi-Square DF Pr > ChiSq Autocorrelations										
6	24.73	6	0.0004	-0.165	-0.193	-0.084	-0.093	0.115	-0.053	
12	26.13	12	0.0103	0.001	0.041	0.021	-0.040	-0.024	-0.030	
18	33.94	18	0.0128	0.036	-0.021	-0.087	0.068	0.106	-0.060	
24	47.66	24	0.0028	0.005	-0.021	-0.060	0.113	0.035	-0.176	



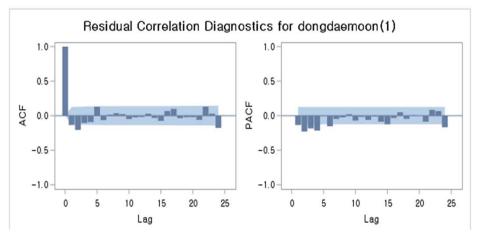
Autocorrelation 존재하는 지 확인

Dickey-Fuller Unit Root Tests											
Туре	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F				
Zero Mean	0	-0.0935	0.6610	-0.15	0.6312						
Single Mean	0	-63.3943	0.0015	-6.05	<.0001	18.29	0.0010				
Trend	0	-91.4147	0.0006	-7.42	<.0001	27.54	0.0010				



Unit root test 결과, 차분 실시 $\rightarrow \nabla z_t^{-\frac{1}{10}}$

3-2 분석 예시 (동대문구)





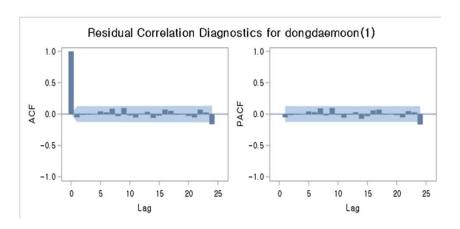
잔차 모형의 p, q 정하기

Conditional Least Squares Estimation										
Parameter Estimate		Standard Error	t Value Approx		Control of the last of the las	Lag	Variable	Shift		
MU	0.0003856	0.00007612	5.07		<.0001	0	dongdaemoon	0		
MA1,1	0.37776	0.06098	6.19		<.0001	2	dongdaemoon	0		
MA1,2	0.88239	0.04847	18.20		<.0001	3	dongdaemoon	0		
MA1,3	-0.26015	0.06730	-3.87		0.0001	5	dongdaemoon	0		
AR1,1	-0.23788	0.05109	-4.66		<.0001	1	dongdaemoon	0		
AR1,2	0.63723	0.06951	9.17		<.0001	3	dongdaemoon	0		
NUM1	0.04199	0.01899	2.21		0.0280	0	pt1	0		



Estimator가 모두 유의하면서 AIC, SBC가 가장 작은 모델을 선정

3-2 분석 예시 (동대문구)



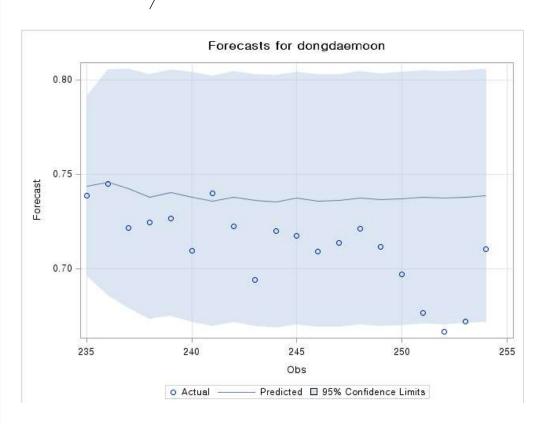
Autocorrelation Check of Residuals										
To Lag Chi-Square DF Pr > ChiSq Autocorrelations										
6	1.45	1		0.2292	0.050	-0.010	-0.004	0.003	0.046	0.030
12	7.10	7		0.4184	0.088	-0.030	0.098	-0.019	-0.051	0.008
18	10.90	13	П	0.6194	0.038	-0.058	-0.024	0.075	0.055	-0.010
24	20.70	19	П	0.3537	0.002	-0.027	-0.050	0.074	0.028	-0.160
30	21.98	25	П	0.6370	0.028	-0.006	0.003	-0.040	-0.016	-0.042
36	26.16	31		0.7138	0.058	0.035	0.060	0.070	-0.029	-0.01
42	28.68	37		0.8344	0.006	0.071	0.009	-0.006	0.053	0.019
48	34.78	43		0.8097	0.022	0.060	0.093	-0.079	0.014	0.02

모델이 제대로 적합됐는지 확인

3-2 분석 예시 (동대문구)

	Forecasts for variable dongdaemoon										
Obs	Forecast	Std Error	95% Confid	ence Limits	Actual	Residual					
235	0.7435	0.0243	0.6959	0.7911	0.7386	-0.0049					
236	0.7456	0.0305	0.6858	0.8054	0.7449	-0.0007					
237	0.7423	0.0324	0.6789	0.8057	0.7214	-0.0209					
238	0.7379	0.0330	0.6731	0.8026	0.7242	-0.0137					
239	0.7401	0.0333	0.6749	0.8053	0.7265	-0.0136					
240	0.7377	0.0338	0.6715	0.8039	0.7096	-0.0281					
241	0.7357	0.0339	0.6693	0.8021	0.7397	0.0040					
242	0.7378	0.0339	0.6713	0.8043	0.7222	-0.0156					
243	0.7360	0.0341	0.6692	0.8028	0.6942	-0.0419					
244	0.7354	0.0341	0,6685	0.8022	0.7200	-0.0154					
245	0.7371	0.0341	0.6703	0.8040	0.7174	-0.0198					
246	0.7358	0.0342	0.6688	0.8028	0.7089	-0.0269					
247	0.7359	0.0342	0.6690	0.8029	0.7134	-0.0226					
248	0.7373	0.0342	0.6703	0.8043	0.7211	-0.0161					
249	0.7363	0.0342	0.6693	0.8034	0.7114	-0.0249					
250	0.7369	0.0342	0.6698	0.8039	0.6969	-0.0400					
251	0.7378	0.0342	0.6708	0.8048	0.6765	-0.0613					
252	0.7372	0.0342	0.6702	0.8043	0.6664	-0.0708					
253	0.7379	0.0342	0.6709	0.8050	0.6718	-0.0661					
254	0.7386	0.0342	0.6716	0.8057	0.7100	-0.0286					

Confidence interval 안에 대부분의 예측 값이 포항됐으므로 제대로 예측됐다고 볼 수 있음



예측 값과 실제 값을 비교

3-3 전체 모델 및 예측

□ PROC ARIMA DATA=new:

identify var=ngangnamgu(1) crosscor=(Pt(1)) stationarity=(adf=(0) dlag=1); estimate q=(2) p=(1,2) input=(Pt) noconstant; forecast back=20 lead=20;

run:

□ PROC ARIMA DATA=new:

identify var=nyoungdeungpogu(1) crosscor=(Pt(1)) stationarity=(adf=(0) dlag=1); estimate q=(2) p=(1,2) input=(Pt) noconstant; forecast back=20 lead=20;

run;

□ PROC ARIMA DATA=new:

 $\label{eq:constraint} \begin{array}{l} \text{identify var=nyongsangu(1) crosscor=(Pt(1)) stationarity=(adf=(0) dlag=1) ;} \\ \text{estimate q=(2,3,5) p=(1,2,3,4) input=(Pt);} \end{array}$

forecast back=20 lead=20;

run:

□ PROC ARIMA DATA=new:

 $\label{eq:constant} \begin{array}{l} \text{identify var=nyangcheongu(1) crosscor=(Pt(1)) stationarity=(adf=(0) dlag=1) ;} \\ \text{estimate q=(2,3) p=(1,2,6) input=(Pt) noconstant;} \end{array}$

forecast back=20 lead=20;

run;

□ PROC ARIMA DATA=new;

 $\label{eq:cosscor} \begin{tabular}{ll} identify var=nsongpagu(1) crosscor=(Pt(1)) stationarity=(adf=(0) dlag=1); \\ estimate q=(2,3,5) p=(1,3) input=(Pt) noconstant; \\ \end{tabular}$

forecast back=20 lead=20;

run:

□ PROC ARIMA DATA=new:

 $\label{eq:constant} \begin{tabular}{ll} identify var=nseongdonggu(1) crosscor=(Pt(1)) stationarity=(adf=(0) dlag=1) ; \\ estimate q=(2,3) p=(1,4) input=(1\$Pt) noconstant; \\ \end{tabular}$

forecast back=20 lead=20;

run:

□ PROC ARIMA DATA=new:

 $\label{eq:constraint} \begin{array}{ll} identify \ var=nseongbukgu(1) \ crosscor=(Pt(1)) \ stationarity=(adf=(0) \ dlag=1) \ ; \\ estimate \ q=(2) \ p=(1\ 2\ 6) \ input=(Pt); \end{array}$

forecast back=20 lead=20;

□ PROC ARIMA DATA=new:

identify var=nseodaemoongu(1) crosscor=(Pt(1)) stationarity=(adf=(0) dlag=1); estimate q=(2) p=(1 2) input=(Pt) noconstant; forecast back=20 lead=20;

run:

□ PROC ARIMA DATA=new;

identify var=nseochogu(1) crosscor=(Pt(1)) stationarity=(adf=(0) dlag=1); estimate q=(1 2 5) p=(2) input=(Pt) noconstant; forecast back=20 lead=20;

run:

□ PROC ARIMA DATA=new:

identify var=nnowongu(1) crosscor=(Pt(1)) stationarity=(adf=(0) dlag=1); estimate q=(2) p=(1 2) input=(Pt) noconstant; forecast back=20 lead=20;

run:

□ PROC ARIMA DATA=new:

identify var=nmapogu(1) crosscor=(Pt(1)) stationarity=(adf=(0) dlag=1); estimate q=(2) p=(1 3 4) input=(Pt) noconstant; forecast back=20 lead=20;

run:

□ PROC ARIMA DATA=new:

identify var=njoongranggu(1) crosscor=(Pt(1)) stationarity=(adf=(0) dlag=1); estimate q=(1 2) p=(1 2 4) input=(Pt); forecast back=20 lead=20;

run:

□ PROC ARIMA DATA=new;

identify var=njoonggu(1) crosscor=(Pt(1)) stationarity=(adf=(0) dlag=1); estimate q=(2) p=(1 2) input=(Pt) noconstant; forecast back=20 lead=20;

run:

□ PROC ARIMA DATA=new:

identify var=ngangbukgu(1) crosscor=(Pt(1)) stationarity=(adf=(0) dlag=1); estimate q=(2) p=(1 2 4) input=(Pt) maxiter=500; forecast back=20 lead=20;

3-3 전체 모델 및 예측

□ PROC ARIMA DATA=new:

identify var=ngangdonggu(1) crosscor=(Pt(1)) stationarity=(adf=(0) dlag=1); estimate q=(1 2) p=(1 2 3) input=(Pt) noconstant maxiter=500; forecast back=20 lead=20;

run:

□ PROC ARIMA DATA=new;

identify var=ndobonggu(1) crosscor=(Pt2(1)) stationarity=(adf=(0) dlag=1); estimate q=(1 2) p(1 2 3) input=(Pt2);

forecast back=20 lead=20;

run:

□ PROC ARIMA DATA=new:

identify var=ndongdaemoongu(1) crosscor=(Pt3(1)) stationarity=(adf=(0) dlag=1); estimate q=(2 3 5) p=(1 3) input=(Pt3);

forecast back=20 lead=20;

run:

□ PROC ARIMA DATA=new:

identify var=ndongzakgu(1) crosscor=(Pt2(1)) stationarity=(adf=(0) dlag=1); estimate q=(2 3 5) p=(1 2 3 4 6) input=(Pt2) noconstant; forecast back=20 lead=20;

run:

□ PROC ARIMA DATA=new:

identify var=neunpyunggu(1) crosscor=(Pt2(1)) stationarity=(adf=(0) dlag=1); estimate q=(2 3 5) p=(1 2 3 4) input=(Pt2);

forecast back=20 lead=20;

run:

□ PROC ARIMA DATA=new:

identify var=ngangseogu(1) crosscor=(Pt2(1)) stationarity=(adf=(0) dlag=1); estimate q=(2) p=(1 2) input=(Pt2) noconstant; forecast back=20 lead=20;

run:

□ PROC ARIMA DATA=new:

identify var=ngumcheongu(1) crosscor=(Pt2(1)) stationarity=(adf=(0) dlag=1); estimate q=(2) p=(1 2) input=(Pt2) noconstant; forecast back=20 lead=20;

run:

□ PROC ARIMA DATA=new:

identify var=ngurogu(1) crosscor=(Pt2(1)) stationarity=(adf=(0) dlag=1); estimate q=(2) p=(1 3 4) input=(Pt2) noconstant; forecast back=20 lead=20;

run:

□ PROC ARIMA DATA=new;

identify var=ngwanakgu(1) crosscor=(Pt2(1)) stationarity=(adf=(0) dlag=1); estimate q=(1) p=(1 2) input=(Pt2) noconstant; forecast back=20 lead=20;

run;

□ PROC ARIMA DATA=new:

identify var=ngwangjingu(1) crosscor=(Pt2(1)) stationarity=(adf=(0) dlag=1); estimate q=(1 2) p=(2) input=(Pt2);

forecast back=20 lead=20; run;

□ PROC ARIMA DATA=new:

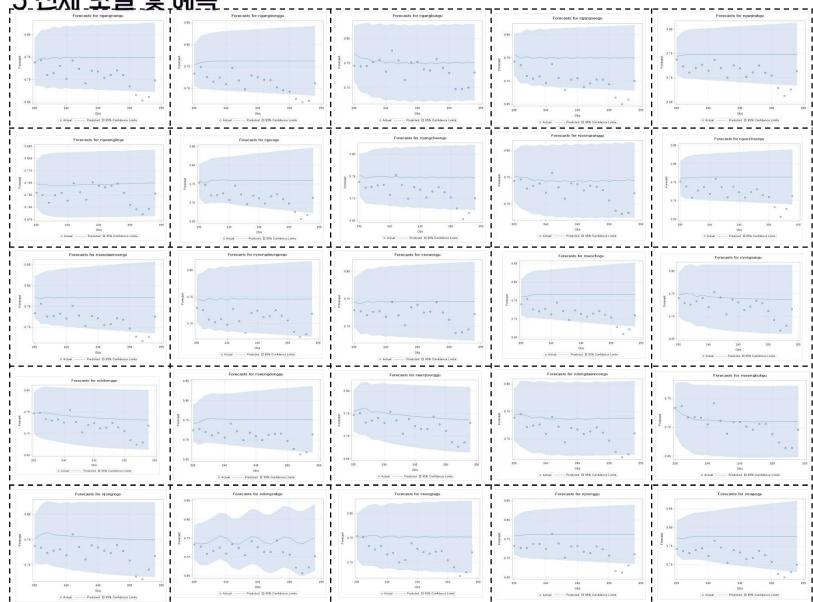
identify var=njongrogu(1) crosscor=(Pt2(1)) stationarity=(adf=(0) dlag=1); estimate q=(2 3 5) p=(1 2 3 4 6) input=(Pt2); forecast back=20 lead=20;

run:

구별 25개 모델 설정

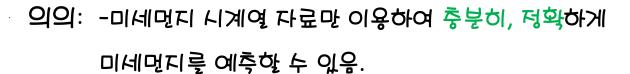
3-3 전체 모델 및 예측

forecasting



결론

4 의의와 한계 및 활용방안



한계: -시간벽 데이터를 이용하지 않아, 시간벽 미세먼지 예측이 불가능

- 종종 신립구간에서 벗어난 예측리가 존대

확용방안: 전이하수 답음모형은 추가하여 미세먼지의 원인은 파악하면 보다 근본적인 해결책은 제시한 수 있은 것.

참고



