

## 충청남도 서산시 기온의 통계적 모형 연구

이혼자<sup>1</sup>

<sup>1</sup>평택대학교 디지털응용정보학과

접수 2014년 9월 1일, 수정 2014년 9월 22일, 게재확정 2014년 10월 2일

### 요약

기온의 변화는 국가 정책에 여러 가지 영향을 준다. 본 연구에서는 충청남도 서산시 2003년 ~ 2012년 기온을 주위에서 쉽게 구할 수 있는 기상자료, 온실가스자료, 대기자료를 이용하여 자기회귀 오차 (autoregressive error) 모형으로 월별과 계절별로 분석하였다. 기온을 위한 기상자료로는, 풍속, 강수량, 일사량, 운량, 습도를 사용했고, 온실가스자료는 이산화탄소 ( $CO_2$ ), 메탄 ( $CH_4$ ), 아산화질소 ( $N_2O$ ), 염화불화탄소 ( $CFC_{11}$ ), 대기자료는 미세먼지 ( $PM_{10}$ ), 이산화황 ( $SO_2$ ), 이산화질소 ( $NO_2$ ), 오존 ( $O_3$ ), 일산화탄소 ( $CO$ )를 사용하였다. 분석 결과, 자기회귀오차모형으로 월별 기온을 39%-63% 정도 설명할 수 있다.

주요용어: 기상변수, 기온, 대기변수, 설명변수, 온실가스변수, 자기회귀오차모형.

### 1. 서론

기온변화는 최근 지구촌의 가장 중요한 이슈중 하나로 주목받고 있으며, 온난화는 인류의 생활을 위협할 수 있는 위기로 인식되고 있다. 특히 우리나라는 전 세계적인 온난화 속도보다 훨씬 빠르게 진행되고 있고, 지역별로 도시화나 산업화 정도, 지형적 위치에 따라 그 속도가 다르게 나타나고 있다. 또한 계절별로도 다르고 기상요인과 대기요인별로도 온난화 속도에 차이를 보이고 있다. 기온변화에 대한 중요한 원인의 하나가 대기조성 물질 중 온실가스의 증가이다. 따라서 기후변화와 지구 온난화에 전 지구촌이 공동으로 대응하기 위하여 국제사회는 1997년 교토 의정서를 채택하여 온실가스를 감축하기로 합의하였으며 당시 이산화탄소 ( $CO_2$ ), 메탄 ( $CH_4$ ), 아산화질소 ( $N_2O$ ), 수소불화탄소 ( $HFC_s$ ), 과불화탄소류 ( $PFC_s$ ), 육불화황 ( $SF_6$ ) 등 6가지 온실가스를 규제대상 물질로서 선정하였다 (Kim 등, 2008). 또한 온난화에 영향을 주는 대기요인으로는 이산화황 ( $SO_2$ ), 이산화질소 ( $NO_2$ ), 일산화탄소 ( $CO$ ), 오존 ( $O_3$ ) 등으로 알려져 있으며 (Song과 Yoon, 2011), 기상요인으로는 풍속 (wind speed), 일사량 (radiation), 운량 (amount of cloud), 강수량 (rainfall) 등이 영향을 많이 주는 것으로 알려져 있다. 우리나라 기상청에서는 세계온실가스자료센터 (WDCGG; World Data Centre for Greenhouse Gases, <http://gaw.kishou.go.jp>) 사이트에 안면도 온실가스 자료인 이산화탄소, 메탄, 이산화질소 등을 1999년 1월1일부터 공개하고 있다.

지금까지 기온 변화에 관해 다양한 각도에서 연구가 진행되어 왔다. 연구의 대부분이 서울, 부산, 제주 등 지역별로 기온변화 상태와 추세를 연구하고 있다. 그러나 기온에 영향을 주는 자료들인 기상자료, 온실가스자료, 대기자료를 이용한 기온 모형화에 관한 연구는 찾아볼 수 없다. 기온변화에 관한 국내 연구의 경향을 살펴보면, 대한민국 40년간의 기온변이추이와 표준기상데이터 비교분석 (Yoo와 Kang,

<sup>1</sup> (450-701) 경기도 평택시 용이동 111, 평택대학교 디지털응용정보학과, 교수. E-mail: esther@ptu.ac.kr

2010), 일일 최고기온의 변화에 대한 추정 (Ko, 2007), 일반화선형모형 (GLM) 날씨발생기를 이용한 서울 일일기온 모형 (Kim과 Kim, 2013), 한국의 기온변화와 기온변동성에 대한 통계적 연구 (Kim 등, 2011), 기후변화와 대기환경의 통합적 관리에 대한 고찰 (Song 등 2011), 기후관측소 분석에 관한 연구 (Kim과 Lee, 2013), 텍스트 마이닝을 활용한 기후변화관련 식품 (Bae 등 2013) 등 다양한 방법으로 기온변화에 관해 연구하였다.

본 연구에서는, 충청북도 서산시 기온자료 중 2003년 1월 1일부터 2012년의 12월 31일까지의 10년간 일 평균자료를 시계열 모형에 적합하고자 한다. 정확한 기온의 분석과 예측을 위해서는 기온과 연관된 기상, 온실가스 및 대기 자료를 포함시켜야 한다. 본 논문에서는 기온과 연관이 있고, 쉽게 자료를 구할 수 있는 5종류의 기상자료와 4종류 온실가스자료와 5종류 대기자료를 설명변수로 포함한 다변량 분석인 ARE (autoregressive error; 자기회귀오차) 모형으로 분석하였다.

본 연구의 목적은 충청남도 서산지역의 기온변화에 관한 모형을 온실가스자료, 대기자료, 기상자료 등을 이용하여 월별, 계절별로 분석하여 기후변화 대응전략을 수립할 수 있는 통계학적 기초자료로써 활용하고자 한다.

## 2. 기온자료 및 설명변수 자료

기온자료는 기상청의 충청남도 서산시 동문동 일 평균자료를 사용하였다. 기간은 2003년부터 2012까지의 1월 1일부터 12월 31일 까지 10년간 자료를 사용했다. 본 연구에서 1월에서 12월 까지 모든 자료를 사용한 이유는 월별로 기온 차이가 있기에 월별 분석이 중요할 것으로 생각되었기 때문이다.

설명변수로 사용한 기상자료, 온실가스자료와 대기자료는 기온과 연관이 있으면서, 쉽게 구할 수 있는 변수를 사용하였다. 모든 설명변수 자료는 기온과 같이 2003년~2012년 자료를 사용하였다. 기상자료는 기상청의 서산자료를 사용하였고, 5종류의 기상자료는 Song 등 (2011)을 참고로 하여 풍속 (wind speed), 상대습도 (relative humidity), 강수량 (rainfall), 일사량 (radiation), 운량 (amount of cloud)을 사용했으며 Table 2.1에 나타나 있다. 4종류의 온실가스자료는 충청남도 태안군 안면도에서 측정된 자료인 이산화탄소 ( $CO_2$ ), 메탄 ( $CH_4$ ), 아산화질소 ( $N_2O$ ), 염화불화탄소 ( $CFC_{11}$ )를 사용하였다.

**Table 2.1** List of explanatory variables

Variable	Contents	Unit	Notation
Temperature	Previous day 1 hour average surface temperature	$^{\circ}C$	$Y_{t-1}$
Wind	Daily maximum wind speed	$m/s$	$W_t$
Rainfall	Daily sum of precipitation	mm	$R_t$
Radiation	Daily sum of radiation	$MJ/M^2$	$D_t$
Cloud	Daily average cloud amount	-	$C_t$
Humidity	Daily average relative humidity	%	$H_t$
$CO_2$	Daily average $CO_2$	ppm	$E_t$
$CH_4$	Daily average $CH_4$	ppb	$B_t$
$N_2O$	Daily average $N_2O$	ppb	$J_t$
$CFC_{11}$	Daily average $CFC_{11}$	ppb	$F_t$
$PM_{10}$	Daily maximum $PM_{10}$	$\mu g/m^3$	$P_t$
$SO_2$	Daily maximum $SO_2$	ppb	$S_t$
$NO_2$	Daily maximum $NO_2$	ppb	$N_t$
$O_3$	Daily maximum $O_3$	ppb	$O_t$
$CO$	Daily maximum $CO$	0.1 ppm	$T_t$

참고로, 우리나라는 온실가스 측정을 위하여 1999년 1월부터 충남 태안군 안면도의 해안 국립공원에서 온실가스를 측정하여 GAW 세계온실가스센터로 자료를 송부하여 활용하고 있다. 5종류의 대기 자

료로는 미세먼지 ( $PM_{10}$ ), 이산화황 ( $SO_2$ ), 이산화질소 ( $NO_2$ ), 오존 ( $O_3$ ), 일산화탄소 ( $CO$ )를 사용했고, 국립환경과학원의 서산시 자료를 사용하였다. 사용된 온실가스자료, 대기 자료도 Table 2.1에 나타나 있다.

각 설명변수의 구체적인 설명은 Table 2.1에 나타나 있다. Table 2.1의 표기 (notation)는 4절의 모형식 변수표시이다.

### 3. 기온의 ARE 모형과 적합성

반응변수에 영향을 주는 설명변수가 있을 때는 설명변수를 분석에 이용하는 것이 효율적이다. ARE 모형은 설명변수를 분석에 사용할 수 있는 모형으로 시계열자료를 회귀모형에 적합 시킬 때 적합한 모형이다. 반응변수인 기온에 영향을 주는  $CH_4$  농도, 풍속, 강수량 등과 같은 설명변수를 사용하여 자료를 분석한다. ARE 모형은 전이함수 모형의 특수한 경우로, 오차항  $\epsilon_t$ 가 독립이 아니라 서로 상관관계를 갖게 되며, 특히 AR (autoregressive) 형태를 갖는다. 일반적인  $k$ 차 ARE 모형은 다음과 같다.

$$Y_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + \cdots + \beta_p X_{tp} + \epsilon_t, \quad t = 1, 2, \dots, n,$$

$$\epsilon_t = e_t - \phi_1 \epsilon_{t-1} - \phi_2 \epsilon_{t-2} \cdots - \phi_k \epsilon_{t-k} \text{ 이고 } e_t \sim i.i.d N(0, \sigma^2) \quad (3.1)$$

식 (3.1)에서 오차항  $\epsilon_t$ 는 서로 상관관계가 있고 AR( $k$ ) 형태를 갖는다. 또한 식 (3.1)에서  $Y_t$ 는 반응변수인 기온을 나타내며,  $X_{t,j}, j = 1, 2, \dots, p$ 는  $Y_t$ 에 영향을 주는  $CO_2$  농도나, 풍속, 등의 설명변수들을 나타낸다. ARE 모형에서는 반응변수의 과거시점 ( $Y_{t-1}, Y_{t-2}$  등)이 설명변수로 종종 사용되기도 한다. 위의 ARE 모형과 회귀모형의 차이점은 ARE 모형은 식 (3.1)에서 오차항  $\epsilon_t$ 가 서로 독립이 아니라 시차에 따른 자기 상관을 갖고 있으며 특히 AR( $k$ ) 모형의 형태를 갖는다는 점이다. 따라서 식 (3.1)을 ARE (자기회귀오차)모형이라 부르며 시계열자료에 회귀모형을 적합 시킬 경우 많이 사용된다.

ARE 모형들의 적합성을 측정하는데 있어 Lee (2012)를 참고하여 결정계수  $R^2$ 와 잔차 분석의 RMSE를 사용하였다. 결정계수  $R^2$ 은 기온  $Y_t$ 가 설명변수에 의해 어느 정도 설명되는지를 나타내는 지표로 값이 0이면 0% 설명되고 1이면 100% 설명됨을 의미한다. 따라서 값이 1에 가까울수록 설명력이 높은 좋은 모형이다. 예측값과 실제값의 분산정도를 나타내는 RMSE (root mean square error)의 공식은 다음과 같다.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (o_i - p_i)^2}, \quad (3.2)$$

$o_i$ 와  $p_i$ 는 각각  $i$ 일의 관측기온과 추정기온을 의미하며,  $n$ 은 통계량에 사용된 일수 (day)이다. RMSE는 값이 작을수록 좋은 모형이다.

### 4. 서산 기온의 ARE 모형

서산지역의 기온의 모형을 측정하기 위한 연구대상 기간은 2003~2012년의 10년간의 일 평균기온이다. 월별로 10년간의 그래프를 그려 복잡해지는 것을 피하기 위해 2003년과 2012년에 대해서만 월 평균온도를 그래프 (Figure 4.1)로 살펴보았다. 봄, 가을인 3월~5월, 9월~10월에는 2003년과 2012년의 기온이 거의 비슷하게 나타났고, 여름인 6월, 7월, 8월에는 2012년이 높게 나타난 반면, 겨울인 2월, 11월, 12월에는 2012년이 조금 낮게 나타났다.

4.1. 월별 기온에 관한 ARE 모형

월별 서산지역 기온모형을 분석하여 Table 4.1에 나타냈다. 월별 분석은 2003년~2012년의 10년간 자료를 모두 모형구축기간으로 사용했다. ARE 모형의 변수 선택은 14종류의 설명변수들 간의 다중공선성 (multi-collinearity)를 고려하면서 설명변수들 간의 다양한 모형을 분석한 후 RMSE 값이 작은 모형을 선택하였다. ARE 모형의 설명변수 차수 선택은 온실가스 변수인  $CO_2$ ,  $CH_4$ ,  $N_2O$ ,  $CFC_{11}$ 와 대기변수인  $SO_2$ ,  $NO_2$ ,  $CO$ ,  $O_3$ 는 당일자료 ( $t$ )를 사용하였고, 기상변수인 풍속, 습도, 강수량, 일사량, 운량은 기온과 더욱 밀접한 관계에 있을 것으로 사료되어 당일 자료 ( $t$ )와 전일자료 ( $t-1$ )를 교대로 사용한 후 적절한 값을 사용하였다.

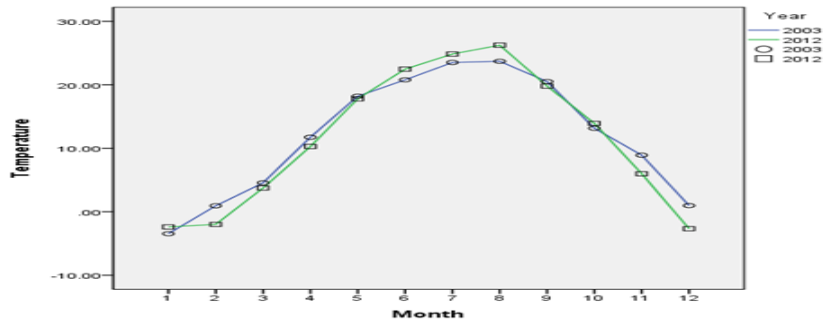


Figure 4.1 The line plots of monthly temperature in Seosan City

ARE 모형의 모수 추정에는 YW (Yule-Walker), ULS (unconditional least square), MLE (maximum likelihood estimate) 등의 방법이 있지만, 본 연구에서는 기온모형과 같이 과거시점의 종속변수 (lagged dependent variable)인  $Y_{t-1}$ 가 설명변수로 사용되는 경우에 적합한 MLE 방법을 사용하였다. ARE모형의 적합성 검정으로는 기온 모형과 같이 과거시점의 종속변수가 설명변수로 사용되는 경우에 적합한 Durbin-h 검정을 사용해 검정하였고 또한 잔차의 ACF (auto-correlation function)와 PACF (partial auto-correlation Function)의 결과를 검정하였다 (Cho와 Lee, 1997 참고).

Table 4.1 The monthly ARE models for temperature

Month	ARE Model	R <sup>2</sup>	RMSE
January	$Y_t = -18.19 + 0.66Y_{t-1} - 0.14W_t - 0.13R_{t-1} + 0.08H_t + 0.03E_t + 86.87S_t + \epsilon_t$ , $\epsilon_t = 0.24\epsilon_{t-9} + e_t$	0.62	1.96
February	$Y_t = -94.52 + 0.42Y_{t-1} - 0.29W_t - 0.14C_{t-1} + 0.03H_t + 0.08E_t + 0.07J_t + 0.14F_t + 50.90S_t + 20.00N_t + \epsilon_t$ , $\epsilon_t = 0.43\epsilon_{t-1} + e_t$	0.45	2.17
March	$Y_t = -64.18 + 0.46Y_{t-1} - 0.14W_{t-1} - 0.10R_{t-1} + 0.05D_{t-1} + 0.10E_t + 0.09F_t + 0.004P_t + 18.97N_t + 39.42O_t + \epsilon_t$ , $\epsilon_t = 0.38\epsilon_{t-1} - 0.13\epsilon_{t-10} + e_t$	0.39	2.17
April	$Y_t = 1.09 + 0.65Y_{t-1} - 0.08R_{t-1} + 51.90O_t + \epsilon_t$ , $\epsilon_t = 0.16\epsilon_{t-9} + 0.16\epsilon_{t-11} + e_t$	0.54	2.17
May	$Y_t = -7.22 + 0.69Y_{t-1} + 0.06D_t + 0.01B_t + 15.61N_t + 18.09O_t + \epsilon_t$ , $\epsilon_t = -0.12\epsilon_{t-8} + 0.16\epsilon_{t-9} - 0.19\epsilon_{t-18} + e_t$	0.62	1.50
June	$Y_t = -3.21 + 0.66Y_{t-1} + 0.06D_t + 0.005B_t + \epsilon_t$ , $\epsilon_t = -0.14\epsilon_{t-13} - 0.14\epsilon_{t-22} + e_t$	0.49	1.33
July	$Y_t = -8.68 + 0.64Y_{t-1} - 0.01R_{t-1} + 0.06D_t + 0.03E_t + 0.01J_t + 0.01F_t + \epsilon_t$ , $\epsilon_t = 0.14\epsilon_{t-4} + e_t$	0.61	1.07
August	$Y_t = -13.93 + 0.47Y_{t-1} + 0.14D_t + 0.07E_t - 25.73N_t + \epsilon_t$ , $\epsilon_t = 0.42\epsilon_{t-1} + 0.13\epsilon_{t-8} + e_t$	0.56	1.01
September	$Y_t = -8.40 + 0.50Y_{t-1} - 0.01R_t + 0.15D_t + 0.07H_t + 0.05F_t + 0.01P_t - 36.90N_t + \epsilon_t$ , $\epsilon_t = 0.29\epsilon_{t-1} - 0.16\epsilon_{t-8} - 0.18\epsilon_{t-17} + e_t$	0.48	1.42
October	$Y_t = -9.80 + 0.60Y_{t-1} - 0.17W_t + 0.18D_t + 0.06H_t + 0.03F_t + 0.01P_t + 74.11S_t + 18.63O_t + \epsilon_t$ , $\epsilon_t = 0.21\epsilon_{t-1} + e_t$	0.59	1.67
November	$Y_t = -21.60 + 0.70Y_{t-1} - 0.07R_t + 0.08H_t + 0.06F_t + 20.79N_t + 34.65O_t + \epsilon_t$ , $\epsilon_t = 0.19\epsilon_{t-9} + e_t$	0.63	2.52
December	$Y_t = -53.03 + 0.25Y_{t-1} - 0.48W_{t-1} - 0.14C_{t-1} + 0.09H_t + 0.07J_t + 0.10F_t + 29.90N_t + 90.33O_t + \epsilon_t$ , $\epsilon_t = 0.62\epsilon_{t-1} + 0.17\epsilon_{t-7} + e_t$	0.41	2.33

Table 4.1의 월별 ARE 모형을 보면, 기온을 설명하는 기상변수, 온실가스변수, 대기변수가 1월에서 12월까지 월 마다 조금씩 다르게 나타났다. 모든 달에 전일 기온 ( $Y_{t-1}$ )가 포함되었고, 겨울인 1월, 2월, 12월에는 기상변수로 풍속 ( $W_t$ ), 습도 ( $H_t$ )가 공통적으로 포함되어 있어 기온에 영향을 주는 것으로 나타났다. 온실가스는 1월, 2월에는 이산화탄소 ( $E_t$ )와 이산화황 ( $S_t$ )가 포함되었고, 12월에는 아산화질소 ( $J_t$ ), 염화불화탄소 ( $F_t$ ) 등이 포함되었다. 특히 일사량과 강수량이 포함된 달이 많아 기온에 영향을 주는 것으로 나타났다. 3월, 5월-10월에는 일사량 ( $D_t$  혹은  $D_{t-1}$ )이 모두 포함되어 있고 1월, 3월, 4월, 7월 9월 11월에는 강수량 ( $R_t$  혹은  $R_{t-1}$ )이 포함되어 있다. 일사량의 계수는 양수로 강수량의 계수는 음수로 나타나, 잘 알려진 바와 같이, 일사량이 많을수록, 강수량이 적을수록 기온은 올라가는 것으로 나타났다. 온실가스 변수로는 이산화탄소 ( $E_t$ )가 1월, 2월, 3월, 7월, 8월에 포함되어 있고, 메탄 ( $B_t$ )는 5월, 6월에 염화불화탄소 ( $F_t$ )가 3월, 7월, 9월, 10월, 11월에 영향을 주는 것으로 나타났다. 대기변수는 이산화황 ( $S_t$ )이 1월, 4월, 5월, 10월에, 이산화질소 ( $N_t$ )가 3월, 5월, 8월, 9월, 11월에 오존 ( $O_t$ )은 3월, 4월, 5월, 10월, 11월인 봄, 가을에 영향을 주는 것으로 분석되었다. 특이한 사항은 8월, 9월의 이산화질소의 계수만 음수이고, 다른 모든 온실가스와 대기변수들의 계수는 양수인 것으로 나타났다.

적합도 척도인 결정계수  $R^2$ 는 0.39-0.63으로 나타났다. 3월, 12월이 각각 0.39, 0.41로 조금 낮게 나타났다고, 1월은 0.62, 11월은 0.63으로 비교적 높게 나타났다. 1월의 예를 들면, 기온  $Y_t$ 는  $Y_{t-1}$ , 풍속 ( $W_t$ ), 강수량 ( $R_t$ ), 습도 ( $H_t$ ), 이산화탄소 ( $E_t$ ), 이산화황 ( $S_t$ )에 의해 62% 설명되는 것으로 나타났다. 여기서  $Y_{t-1}$ ,  $H_t$ ,  $E_t$ ,  $S_t$ 는 양수로 나타났고, 반면  $W_t$ ,  $R_t$ 는 음수로 나타나 습도, 이산화탄소, 이산화황이 높을수록 기온은 올라가고, 풍속과 강수량이 높을수록 기온은 낮은 것으로 나타났다.

적합도 척도로 실제값과 예측값의 차이를 나타내는 RMSE 값은 1.01에서 2.52로 나타났다. 본 연구에서 사용한 기상변수, 온실가스변수와 대기변수는 각각 기상청, 세계온실가스자료센터, 국립환경과학원에서 쉽게 구할 수 있는 자료이므로, Table 4.1을 이용하면, 월별 기온을 예측하는데 도움이 될 것으로 생각된다.

#### 4.2. 계절별 기온에 관한 ARE 모형

서산온도를 봄, 여름 가을, 겨울 계절별 ARE 모형으로 분석한 결과를 Table 4.2에 나타내었다. 계절 분석에서, 봄은 3월-5월, 여름은 6월-8월, 가을은 9월~11월, 겨울은 12월-2월의 자료를 사용하였다. 계절별 ARE 모형의 결정계수  $R^2$ 는 겨울기온의 경우 0.47로 비교적 낮게 나타났으나, 봄, 여름, 가을기온은 각각 0.80, 0.71, 0.84로 월별모형보다 매우 높게 나타났다. 계절별로 영향을 주는 기상, 온실가스, 대기변수들이 조금씩 다르게 나타났다.

봄 기온은 전일기온, 강수량, 일사량, 운량, 습도, 염화불화탄소, 이산화질소, 오존에 의해 80% 설명된다. 여름기온은 전일기온, 강수량, 일사량, 습도의 기상변수에 의해서만 71% 설명되는데 특이한 사실은 여름기온은 온실가스와 대기변수의 영향보다는 기상변수에 의한 영향이 큰 것으로 나타났다는 사실이다. 가을기온은 전일기온, 일사량, 운량, 습도, 메탄, 염화불화탄소의 기상변수와 온실가스변수에 의해 84% 설명되고 겨울기온은 전일온도, 풍속, 일사량, 메탄, 염화불화탄소, 이산화황, 오존, 이산화질소에 의해 47% 설명되는 것으로 나타났다.

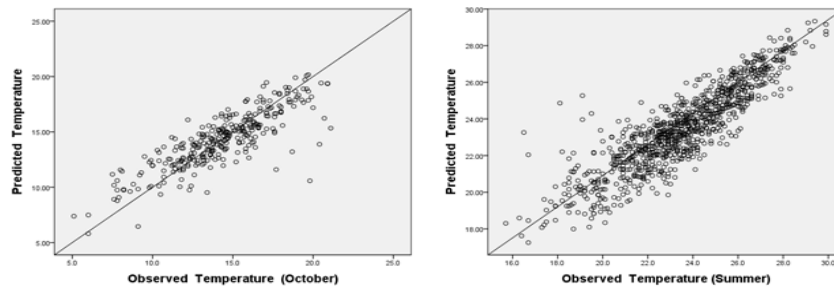
대체적으로 계절별 모형에 사용된 설명변수들이 월별모형에 포함된 설명변수들과는 조금씩 다르게 나타났다고, 특히 일사량  $D_t$ 는 봄, 여름, 가을, 겨울 4계절 모두 포함되어 있어 영향을 제일 많이 주는 기상변수로 나타났다. RMSE 값은 1.19에서 2.45로 월별 모형의 RMSE 값과 비슷하게 나타났다.

**Table 4.2** The seasonal ARE models for temperature

Season	ARE Model	$R^2$	RMSE
Spring	$Y_t = -10.77 + 0.80Y_{t-1} - 0.02R_{t-1} + 0.09D_t - 0.11C_{t-1} + 0.05H_t + 0.02F_t + 36.62N_t + 48.19O_t + \epsilon_t$ , $\epsilon_t = 0.10\epsilon_{t-1} + e_t$	0.80	2.39
Summer	$Y_t = 1.12 + 0.82Y_{t-1} - 0.01R_t + 0.08D_t + 0.02H_t + \epsilon_t$ , $\epsilon_t = 0.12\epsilon_{t-4} + 0.10\epsilon_{t-10} + 0.10\epsilon_{t-12} + e_t$	0.71	1.19
Fall	$Y_t = -23.48 + 0.78Y_{t-1} + 0.18D_t - 0.10C_{t-1} + 0.11H_t + 0.01B_t + 0.04F_t + \epsilon_t$ , $\epsilon_t = 0.09\epsilon_{t-9} + 0.09\epsilon_{t-14} + e_t$	0.84	2.45
Winter	$Y_t = -42.77 + 0.49Y_{t-1} - 0.48W_{t-1} + 0.09D_{t-1} + 0.01B_t + 0.09F_t + 33.91S_t + 28.90O_t + 20.28N_t + \epsilon_t$ , $\epsilon_t = 0.36\epsilon_{t-1} + 0.08\epsilon_{t-2} + 0.17\epsilon_{t-8} + 0.05\epsilon_{t-9} + 0.05\epsilon_{t-12} + 0.10\epsilon_{t-14} + e_t$	0.47	2.01

### 4.3. ARE 모형의 실제값과 추정치의 산점도

서산기온 실제값과 ARE 모형 추정치의 산점도를 월별에서는 10월, 계절별에서는 여름을 Figure 4.2에 나타내었다. 10월 모형의 산점도를 보면, 기온이 10°C 이하인 경우에는 약간의 과대 추정된 경향이 있고, 10° ~ 18°C 사이는 추정이 잘되었고 18°C 이상인 경우에는 과소 추정되는 값들이 있는 것으로 나타났다. 반면에 여름 모형은 기온이 20°C 이하인 경우 과대 추정된 값이 약간 있으나, 20°C 이상인 경우에는 추정이 비교적 잘 된 것으로 나타났다. ARE 모형은 기온에 영향을 주는 설명변수를 수식으로 표현하는 장점이 있으므로, 모형에 기온에 영향을 주는 기상변수, 온실가스변수와 대기변수를 좀 더 포함시켜 개발하면, 실제값에 좀 더 가까이 접근하는 좋은 결과를 가질 수 있을 것으로 기대한다.



**Figure 4.2** The scatter plots (monthly and seasonal) between actual value and ARE estimated value of temperature

## 5. 결론

충청남도 지역인 서산의 기온을 월별과 계절별로 시계열 모형인 ARE 방법으로 분석하였다. ARE 모형은 설명변수를 분석에 사용할 수 있는 모형으로 시계열자료를 적합 시킬 때 적절한 모형이다. 본 연구에서는, 기온에 영향을 주면서 쉽게 자료를 구할 수 있는 5종류의 기상자료, 4종류의 온실가스자료, 5종류의 대기자료를 설명변수로 사용하였다. 5종류의 기상자료는 서산시 기상청 자료로 풍속, 강수량, 일사량, 운량, 습도이며 4종류의 온실가스자료는 충청남도 태안군 안면도에서 측정된 자료인  $CO_2$ ,  $CH_4$ ,  $N_2O$ ,  $CFC_{11}$ 를 사용하였고 5종류의 대기자료는 국립환경과학원의 서산시  $PM_{10}$ ,  $SO_2$ ,  $NO_2$ ,  $CO$ ,  $O_3$ 이다.

연구기간인 2003년부터 2012년의 1월 1일에서 12월 31일까지의 10년 자료를 사용하여 월별과 계절별 ARE 모형으로 분석하였다. 월별 기온을 분석한 결과 기상변수로는 일사량과 강수량이 영향을 많이 주는 것으로 나타났다. 3월, 5월~10월에는 일사량이 모두 포함되어 있고 1월, 3월, 4월, 7월, 9월, 11월에는 강수량이 포함되어 있다. 온실가스변수는  $CO_2$ ,  $CFC_{11}$ 이 영향을 많이 주며 대기변수  $SO_2$ ,

$NO_2$ ,  $CO$ ,  $O_3$ 는 월별로 각각 다르게 영향을 주는 것으로 나타났다. 결정계수  $R^2$ 는 0.39-0.63으로 3월이 0.39로 낮게 나타났고 1월, 5월, 7월, 11월에는 각각 0.62, 0.62, 0.61, 0.63으로 설명력이 비교적 높게 나타났다. RMSE 값은 1.01에서 2.52로 비교적 낮게 나타났다.

계절별 기온분석 결과는 일사량이 봄, 여름, 가을, 겨울 4계절 모두 포함되어 있어, 영향을 제일 많이 주는 기상변수로 나타났다. 또한 봄, 여름에는 강수량이 가을에는 운량, 겨울에는 풍속이 영향을 주는 변수로 나타났다. 특이한 사실은 여름기온은 온실가스변수나 대기변수는 영향을 주지 않고 강수량, 일사량, 습도의 기상변수에 의한 영향이 매우 큰 것으로 나타났다. 온실가스변수로는 메탄, 염화불화탄소가 봄, 가을, 겨울에 영향을 주는 변수로 나타났다. 계절별 ARE 모형의 결정계수  $R^2$ 는 겨울기온은 0.47로 비교적 낮게 나타났으나, 봄, 여름, 가을기온은 각각 0.80, 0.71, 0.84로 월별모형보다 매우 높게 나타났다. RMSE 값은 1.19에서 2.45로 월별 모형의 RMSE 값과 비슷하게 나타났다.

ARE 모형은 Table 4.1, Table 4.2과 같이 기온에 어떤 설명변수들이 영향을 주는지 등의 수식으로 간단히 표현할 수 있는 장점이 있으므로, 본 연구에서 사용하지 못한 설명변수들을 좀 더 첨가하여 분석하면 더 좋은 모형이 되리라 기대한다. 앞으로, 기온에 사용되는 설명변수에 관한 다양한 연구와 함께 다양한 모형에 관한 연구가 지역별로 활발히 이루어져야 된다고 생각된다.

## References

- Bae, K., Park, J., Kim, J. and Lee, Y. (2013). Analysis of the abstracts of research articles in food related to climate change using a text-mining algorithm. *Journal of the Korean Data & Information Science Society*, **24**, 1429-1437.
- Cho, S. and Lee, J. (1997). *Analysis of economic time series analysis using SAS/ETS*, Freedom Academy, Seoul.
- Kim, H., Do, H. Y. and Kim, Y. (2013). A modeling of daily temperature in Seoul using GLM weather generator. *The Korean Journal of Applied Statistics*, **26**, 413-420.
- Kim, H. C., Choi, S. K. and Yun, B. R. (2011). A statistical analysis on temperature change and climate variability in Korea. *Communications of the Korean Statistical Society*, **18**, 1-12.
- Kim, H. K. and Lee, Y. (2013). A study on the density analysis of climatological stations using the correlation integral method in the fractal dimension. *Journal of the Korean Data & Information Science Society*, **24**, 53-62.
- Kim, J. S., Lee, J. B. and Lee, J. S. (2008). The status and strategy of climate change watch. *Korean Journal of Nature Conversation*, **2**, 24-31.
- Ko, W. K. (2007). Estimation for change of daily maxima temperature. *The Korean Journal of Applied Statistics*, **20**, 1-9.
- Lee, H. (2012). Analysis of time series models for consumer price index. *Journal of the Korean Data & Information Science Society*, **23**, 535-542.
- Song, C. K., Lee, S., and Yoon, J. S. (2011). A review of the integrated strategy for climate change and air pollution management. *Journal of Korean Society for Atmospheric Environment*, **27**, 805-818.
- Yoo, H. C. and Kang, H. G. (2010). Comparative analysis of temperature change trend and standard meteorological data in Korea over 40 years. *Journal of KIAEBS*, **4**, 97-103.

## Analysis of statistical models on temperature at the Seosan city in Korea

Hoonja Lee<sup>1</sup>

<sup>1</sup>Department of Information Statistics, Pyeongtaek University

Received 1 September 2014, revised 22 September 2014, accepted 2 October 2014

### Abstract

The temperature data influences on various policies of the country. In this article, the autoregressive error (ARE) model has been considered for analyzing the monthly and seasonal temperature data at the northern part of the Chungcheong Namdo, Seosan monitoring site in Korea. In the ARE model, five meteorological variables, four greenhouse gas variables and five pollution variables are used as the explanatory variables for the temperature data set. The five meteorological variables are wind speed, rainfall, radiation, amount of cloud, and relative humidity. The four greenhouse gas variables are carbon dioxide ( $CO_2$ ), methane ( $CH_4$ ), nitrous oxide ( $N_2O$ ), and chlorofluorocarbon ( $CFC_{11}$ ). And the five air pollution explanatory variables are particulate matter ( $PM_{10}$ ), sulfur dioxide ( $SO_2$ ), nitrogen dioxide ( $NO_2$ ), ozone ( $O_3$ ), and carbon monoxide ( $CO$ ). The result showed that the monthly ARE model explained about 39-63% for describing the temperature. However, the ARE model will be expected better when we add the more explanatory variables in the model.

*Keywords:* Autoregressive error model, explanatory variable, greenhouse gas variables, meteorological variables, pollution variable, temperature.

---

<sup>1</sup> Professor, Department of Information Statistics, Pyeongtaek University, Pyeongtaek 450-701, Korea.  
E-mail: esther@ptu.ac.kr