

Analysis of the Effect of Wind Direction on Ozone Level

Jong-Hwa Na^{1,a}, Su-Jin Sung^b, Hye-Kyung Yu^a

^aDepartment of Information & Statistics, Chungbuk National University

^bKorean Educational Development Institute

Abstract

In this paper we analyze the effect of circular variables such as wind direction, time and month on the ozone level. In particular, we examined the effect of wind direction by exploratory data analysis methods and provide the correlation and regression analyzes in the cases including all circular explanatory variables. In the analysis, we convert time and month variables to circular variables and analyze the effect of these variables on regression analysis; in addition, we also consider circular-circular regression. We used weather condition and air pollution data collected from Dongdaemoon district of Seoul in 2007.

Keywords: Circular data, circular-linear regression, circular-circular regression, ozone level.

1. 서론

오존농도에 미치는 많은 변수 가운데 풍향(wind direction)은 $0^\circ \sim 360^\circ$ 사이의 값을 가지는 이차원의 방향자료 즉, 원형자료(circular data)로 취급될 수 있다. 일반적인 선형분석에서의 연속형 자료와는 달리 원형자료에 대해서는 특별히 고안된 통계적 방법을 통해 분석이 이루어져야 한다. 원형자료는 단위 원 상의 점들로 표현될 수 있으며, 영방향(zero direction)과 회전(시계 혹은 반시계) 방향의 설정에 따라 그 측정값이 달라질 수 있음에 유의하여야 한다. 따라서 바람의 방향과 같은 원형자료에 대해서는 선형에서의 산술평균, 표준편차 등은 무의미하며, 상관 및 회귀분석의 경우에도 특별히 고안된 방법이 적용되어야 한다. 원형자료의 분석과 관련된 문헌으로는 Mardia (1972), Batschelet (1981), Fisher (1993), Jammalamadaka와 SenGupta (2001) 등이 있다. Mardia (1972)는 방향자료에 대한 고전적 이론을 주로 다루었고, Batschelet (1981)은 생물학 분야에의 적용을, Fisher (1993)는 원형자료에 대한 이론 및 회귀 등의 응용 문제를 다루고 있다. 특히 Jammalamadaka와 SenGupta (2001)는 R패키지를 이용한 원형자료의 분석 방법을 함께 소개하고 있다.

본 연구는 풍향이 오존농도에 미치는 영향을 탐색적으로 분석하고, 원형자료를 포함하는 상관 및 회귀분석을 수행한다. 아울러 월, 시간 등의 자료를 원형자료로 변환하여, 상관 및 회귀분석 결과에 미치는 영향을 알아본다. 이러한 분석이 가지는 의미는 원형변수인 범주형으로 다루어질 수도 있는 풍향 변수 뿐 아니라 월, 시간 등의 시계열 속성의 변수들을 원형자료화하여 일반적 형태의 회귀모형에 포함 시킴으로써 보다 많은 정보를 이용한 회귀 모형을 구축하고, 그로부터 추정된 회귀식의 정도를 제고하는 데 있다고 할 수 있다. 또한 원형변수들(풍향, 시간) 간의 회귀를 통해 풍향에 미치는 시간의 효과를 모형화 하였다. 2절에서는 원형자료에 대한 상관 및 회귀분석 방법을 간략히 소개하고, 3절에서는 실제자료 분석 결과를 제시하였으며 4절에서는 결론을 제시하였다.

¹ Corresponding author: Professor, Department of Information & Statistics, Chungbuk National University, Cheong-ju, Chungbuk 361-763, Korea. E-mail: cherin@cbu.ac.kr

2. 원형자료에 대한 상관 및 회귀분석 소개

2.1. 기초분석

먼저 각(degree)으로 주어지는 n 개의 원형자료를 $\theta_1, \dots, \theta_n$ 이라 하자. 이때 원형자료의 평균방향을 나타내는 원형평균(circular mean) $\bar{\theta}$ 는

$$\begin{aligned} \bar{\theta} &= \arctan^*(S/C) \\ &= \begin{cases} \arctan(S/C), & \text{if } C > 0, S \geq 0, \\ \pi/2, & \text{if } C = 0, S > 0, \\ \arctan(S/C) + \pi, & \text{if } C < 0, \\ \arctan(S/C) + 2\pi, & \text{if } C \geq 0, S < 0, \\ \text{undefined}, & \text{if } C = 0, S = 0 \end{cases} \end{aligned} \quad (2.1)$$

로 정의된다. 위 식에서 C 와 S 는 합성벡터 r 의 성분으로 다음과 같이 정의된다.

$$r \equiv (C, S) = \left(\sum_{i=1}^n \cos \theta_i, \sum_{i=1}^n \sin \theta_i \right), \quad i = 1, \dots, n.$$

또한 원형자료의 평균방향으로부터 퍼진 정도를 나타내는 원형분산(circular variance) V_0 는

$$V_0 = 1 - \bar{r}, \quad 0 \leq V_0 \leq 1$$

으로 정의된다. 위 식에서 $\bar{r} = \|r\|/n$ 으로 평균합성벡터의 크기로 원형평균으로부터의 집중도를 나타낸다.

2.2. 원형 상관분석

2.2.1. 원형-선형 상관계수

먼저 x 를 선형변수, α 를 원형변수라 하자. n 개의 표본을 $(\alpha_1, x_1), (\alpha_2, x_2), \dots, (\alpha_n, x_n)$ 이라 하자. Mardia (1976)와 Johnson과 Wehrly (1977)는 두 변수 간의 표본상관계수를 다음과 같이 정의하였다.

$$r^2 = \frac{r_{xc}^2 + r_{xs}^2 - 2r_{xc}r_{xs}r_{cs}}{1 - r_{cs}^2},$$

여기서 r_{xc}, r_{xs}, r_{cs} 는 다음과 같이 정의된다.

$$\begin{aligned} r_{xc} &= \text{Corr}(x, \cos \alpha), \\ r_{xs} &= \text{Corr}(x, \sin \alpha), \\ r_{cs} &= \text{Corr}(\cos \alpha, \sin \alpha). \end{aligned}$$

2.2.2. 원형-원형 상관계수

이변량의 원형자료를 $(\alpha_1, \beta_1), \dots, (\alpha_n, \beta_n)$ 이라 할 때, Jammalamadaka와 Sarma (1988)는 두 원형변수들 간의 표본상관계수를 다음과 같이 정의하였다.

$$r_c = \frac{\sum_{i=1}^n \sin(\alpha_i - \bar{\alpha}) \sin(\beta_i - \bar{\beta})}{\sqrt{\sum_{i=1}^n \sin^2(\alpha_i - \bar{\alpha}) \sum_{i=1}^n \sin^2(\beta_i - \bar{\beta})}},$$

여기서 $\bar{\alpha}, \bar{\beta}$ 는 원형평균으로 식 (2.1)과 동일한 방법으로 정의된다.

원형변수들 간의 또 다른 상관계수에 대해서는 Mardia (1975), Johnson과 Wehrly (1977), Mardia와 Puri (1978), Stephens (1979), Rivest (1982)를 참고하기 바란다.

2.3. 원형 회귀분석

2.3.1. 선형-원형 회귀

반응변수가 선형이며 독립변수가 원형변수인 회귀모형을 다음과 같이 정의하자.

$$y = A_0 + A_1 \cos \omega(\alpha - \alpha_0),$$

여기서 α 는 원형변수이고, α 에 대한 주기를 T 라고 할 때, 위 모형에서는 다음과 같은 모수를 포함한다. 위 식에서 A_0 는 반응변수의 평균, A_1 은 진폭, ω 는 진동(주기) ($\omega = 2\pi/T$ 혹은 $\omega = 360^\circ/T$), α_0 는 최대 위상(acrophase)을 나타낸다. 이때 α 와 T 는 라디안(radian)또는 각(degree)으로 측정된다.

위의 모형식은 다음과 같이 표현될 수 있다.

$$Y = A_0 + A_1 \cos(\omega\alpha - \phi), \quad \phi = \omega\alpha_0,$$

여기서 $\cos(\omega\alpha - \phi) = \cos \omega\alpha \cos \alpha_0 + \sin \omega\alpha \sin \alpha_0$ 이다. 이를 Y 에 관한 식으로 다시 쓰면 다음과 같이 표현된다.

$$Y = A_0 + a \cos \omega\alpha + b \sin \omega\alpha,$$

여기서 $a = A_1 \cos \alpha_0$, $b = A_1 \sin \alpha_0$ 이다. 이 모형은 A_0, a, b, ω 의 모수를 포함하지만 ω 는 주기에 의해 주어지게 되므로, 남은 세 개의 모수만 추정하면 된다. 수식을 선형회귀로 간주하여 최소제곱법에 의해 모수 A_0, a, b 를 추정하면 된다.

2.3.2. 원형-원형 회귀

원형변수 (α, β) 가 결합확률분포 $f(\alpha, \beta)$ 를 갖는다고 하자($0 < \alpha, \beta \leq 2\pi$). 원형확률변수 α 가 주어졌을 때 β 를 예측하기 위해 다음의 조건부 기대치를 고려한다.

$$E(e^{i\beta} | \alpha) = \rho(\alpha)e^{i\mu(\alpha)} = g_1(\alpha) + ig_2(\alpha) (\equiv g(\alpha)),$$

여기서 $\mu(\alpha)$ 는 α 가 주어졌을 때 원형확률변수 β 의 조건부 평균방향이고, $\rho(\alpha)$ 는 $0 \leq \rho(\alpha) \leq 1$ 로 평균방향으로의 조건부 집중도라고 할 수 있다. 또한

$$E(\cos \beta | \alpha) = g_1(\alpha),$$

$$E(\sin \beta | \alpha) = g_2(\alpha)$$

으로부터, β 는 다음과 같이 결정된다.

$$\mu(\alpha) = \hat{\beta} = \arctan^* \frac{g_2(\alpha)}{g_1(\alpha)}.$$

위 식에서 \arctan^* 는 식 (2.1)을 참고하기 바란다. 또한, 선형회귀에서 최소제곱추정의 관점에서 볼 때, 위 추정량은 다음식

$$E \left\{ \left| e^{i\beta} - g(\alpha) \right|^2 \right\}$$

Table 1: Variables of air pollution and weather conditions

category	variable name	explanation	reference
air pollution	ozone	ozone level(ppm)	reponse variable
	sulf_acid	sulfurous acid(ppm)	
	nit_diox	nitrogen dioxide(ppm)	
	car_mono	carbon monoxide(ppm)	
	dust	fine dust ($\mu g/m^3$)	
weather condition	wind_sp	wind speed(m/s)	circular variable
	wind_dir	wind direction(degree)	
	temp	temperature($^{\circ}C$)	
	rain	rainfall(mm)	
time	month	month(1~12)	circular variable
	time	time(1~24)	circular variable

을 최소화하는 값으로 이해될 수 있다 (Jammalamadaka와 Sarma, 1993).

위 식에서 $g_1(\alpha)$ 와 $g_2(\alpha)$ 는, 주기가 2π 이고 차수가 m 이므로, 적당한 차수 m 의 삼각다항식으로 다음과 같이 근사될 수 있다.

$$g_1(\alpha) \approx \sum_{k=0}^m (A_k \cos k\alpha + B_k \sin k\alpha),$$

$$g_2(\alpha) \approx \sum_{k=0}^m (C_k \cos k\alpha + D_k \sin k\alpha).$$

따라서 다음의 일반선형모형 형태의 모형을 얻게 된다.

$$\cos \beta = \sum_{k=0}^m (A_k \cos k\alpha + B_k \sin k\alpha) + \epsilon_1,$$

$$\sin \beta = \sum_{k=0}^m (C_k \cos k\alpha + D_k \sin k\alpha) + \epsilon_2.$$

여기서 (ϵ_1, ϵ_2) 는 평균 벡터가 0이고, 분산행렬이 미지의 Σ 인 오차 벡터이다.

n 개의 관측자료 $(\alpha_1, \beta_1), (\alpha_2, \beta_2), \dots, (\alpha_n, \beta_n)$ 를 이용하여 위 모형의 계수들 (A_k, B_k, C_k, D_k) , $k = 0, 1, \dots, m$ 을 추정하는 절차와 차수 m 의 결정과정에 대해서는 Jammalamadaka와 SenGupta (2001)를 참고하기 바란다.

3. 실제자료분석

3.1. 자료 설명

분석에 사용된 자료는 2007년 1월부터 12월까지 서울시 동대문구 지역에서 한시간 단위로 측정된 총 3706개(결측치 제외)의 기상자료와 대기오염자료이다 (자료제공: Air Korea(<http://www.airkorea.or.kr>), 자동기상관측자료(AWS)).

분석에 다루어질 변수들에 대한 설명은 Table 1과 같다. 분석의 전반에서 반응변수는 오존이며, 설명변수는 오존(반응변수)을 제외한 나머지 대기오염 변수와 기상 변수들이다. 이 가운데 풍향 변수는 지금까지 많은 연구에서 범주형 변수로 다루어져 왔으나, 본 연구에서는 원형변수로 다루어 탐색적 분석을 비롯한 여러 가지 분석을 실시하고, 시간 변수(월, 시간) 역시 원형변수로의 전환을 통

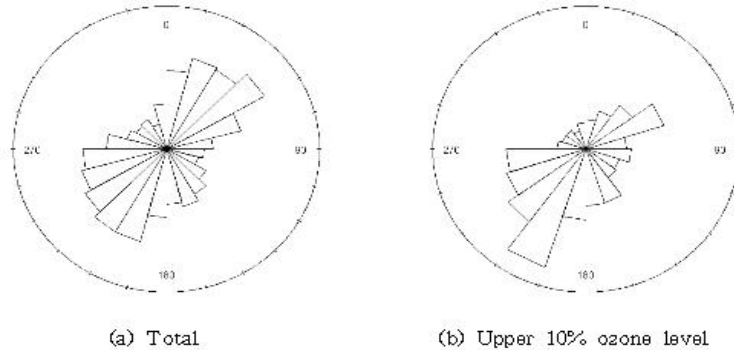


Figure 1: Rose diagram of wind direction data

해 관련(상관 및 회귀) 분석을 실시하였다. 풍향의 경우 북쪽(N)을 영방향으로하여 시계방향으로 $0\sim 2\pi$ 범위의 값으로 나타내었다. 시간변수 가운데 월(month) 변수는 $(\text{월} \times 2\pi/12)$, 시간(time) 변수는 $(\text{시간} \times 2\pi/24)$ 을 통해 원형 자료로 변환하였다. 오존농도의 추정과 관련하여 Jammalamadaka와 Lund (2006)는 Table 1에 소개된 변수 가운데 기상변수(wind_sp, temp, wind_dir)와 시간변수(month)만을 고려하여 분석한 바 있으나, 본 논문에서는 5개의 대기오염 변수를 추가하여 분석을 실시하였다. 추가된 대기오염 변수는 오존, 아황산가스, 이산화질소, 일산화탄소, 미세먼지 변수로 자료제공처에서 제시된 변수들이다.

3.2. 풍향 자료 분석

이 절에서는 풍향 자료에 대한 분석을 실시한다. Figure 1은 풍향자료를 로즈다이아그램(rose diagram)으로 나타낸 것이다.

Figure 1의 (a)로부터 이 지역에서의 풍향은 연중 북동과 남서 방향에 집중되는 현상을 보이고 있으며, 오존의 상위 10%에 해당하는 풍향은 남서 방향에 집중되는 경향이 더욱 뚜렷하며, 전체 자료(Figure 1(a))의 패턴과는 상당한 차이를 보이고 있다.

Figure 2는 풍향 변수가 오존을 비롯한 여러 가지 기상 변수들에 미치는 영향을 나타내고 있다. 그림 속에 나타난 숫자는 해당 기상변수의 평균값을 제시한 것이다.

Figure 2의 (a)와 (b)를 통해 기온과 강수량 수치는 풍향에 따라 큰 차이를 보이지 않음을 확인할 수 있다. 그러나 Figure 2의 (c)와 (d)에서는 기타 방향에 비해 남서풍의 경우의 수치가 크게 나타나는 현상을 보이고 있다. 따라서 풍속과 오존 변수는 풍향의 영향을 어느 정도 받고 있으며, 이는 동대문구의 지역적 위치를 고려하였을 때 인천과 안산지역으로부터 불어오는 바람에 의해 오존 수준이 높아지는 원인을 생각해 볼 필요가 있다.

3.3. 상관 및 회귀분석

3.3.1. 상관분석

Table 2는 변수들 간의 상관분석을 실시한 결과이다. 원형변수들(월, 시간, 풍향)과 기타 변수들 간에는 원형-선형 상관계수를, 이들 세 변수들 간에는 원형-원형 상관계수의 값을 제시하였으며, 나머지는 피어슨 상관계수를 나타낸다.

Table 2에서 오존과의 상관계수가 가장 높은 변수는 풍속(wind_sp)이며, 다음으로 온도(temp), 미세먼지(dust)의 순으로 높은 양의 상관을 보이고 있음을 알 수 있다. 따라서 풍속이 클수록, 온도나 미

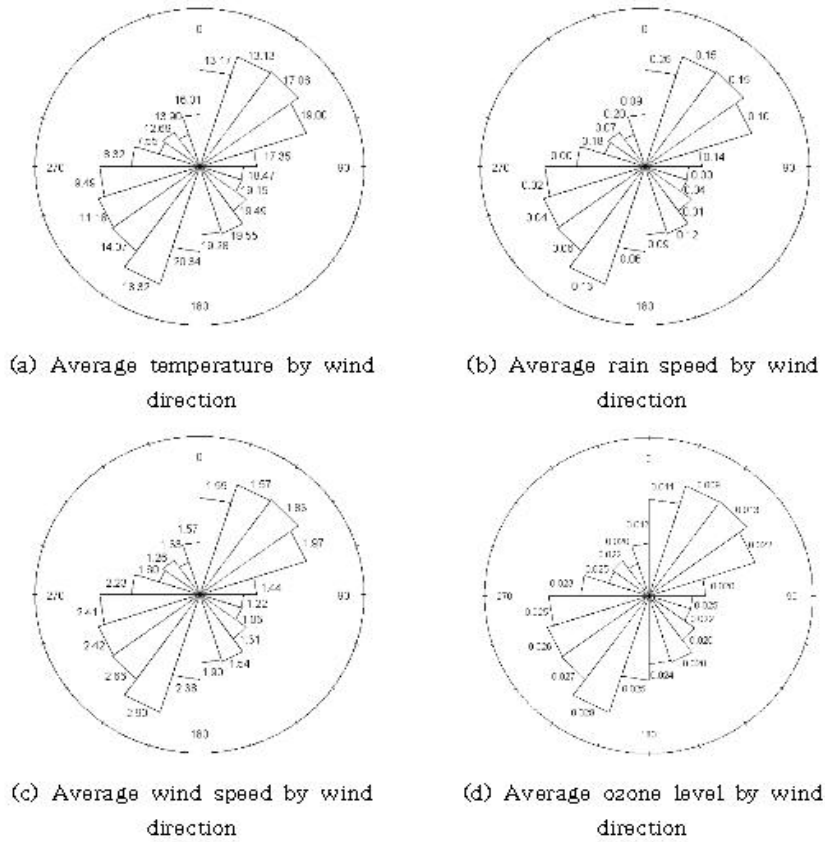


Figure 2: Summary of weather condition data by wind direction

Table 2: Correlation of variables including circular variables

	month	time	sulf_acid	nit_diox	car_mono	dust	wind_dir	wind_sp	temp	rain	ozone
month	1.0000	-0.0039 ^c	0.1855 ^b	0.0914 ^b	0.2335 ^b	0.1195 ^b	-0.0840 ^c	0.0300 ^b	0.8208 ^b	0.0115 ^b	0.1571 ^b
time	-0.0039 ^c	1.0000	0.0141 ^b	0.0180 ^b	0.0062 ^b	0.0007 ^b	0.1859 ^c	0.1070 ^b	0.2540 ^b	0.0019 ^b	0.1381 ^b
sulf_acid	0.1855 ^b	0.0141 ^b	1.0000	0.4490 ^a	0.5970 ^a	0.4130 ^a	0.0025 ^b	-0.0700 ^a	-0.3020 ^a	-0.0850 ^a	-0.0260 ^a
nit_diox	0.0914 ^b	0.0180 ^b	0.4490 ^a	1.0000	0.7240 ^a	0.3830 ^a	0.0363 ^b	-0.3430 ^a	-0.1670 ^a	-0.0420 ^a	-0.3440 ^a
car_mono	0.2335 ^b	0.0062 ^b	0.5970 ^a	0.7240 ^a	1.0000	0.4310 ^a	0.0286 ^b	-0.3150 ^a	-0.3920 ^a	-0.0580 ^a	-0.3510 ^a
dust	0.1195 ^b	0.0007 ^b	0.4130 ^a	0.3830 ^a	0.4310 ^a	1.0000	0.0291 ^b	-0.0310 ^a	-0.0940 ^a	-0.0700 ^a	0.1320 ^a
wind_dir	-0.0840 ^c	0.1859 ^c	0.0025 ^b	0.0363 ^b	0.0286 ^b	0.0291 ^b	1.0000	0.1242 ^b	0.1061 ^b	0.0029 ^b	0.0904 ^b
wind_sp	0.0300 ^b	0.1070 ^b	-0.0700 ^a	-0.3430 ^a	-0.3150 ^a	-0.0310 ^a	0.1242 ^b	1.0000	0.1060 ^a	0.0500 ^a	0.3890 ^a
temp	0.8208 ^b	0.0254 ^b	-0.3020 ^a	-0.1670 ^a	-0.3920 ^a	-0.0940 ^a	0.1061 ^b	0.1060 ^a	1.0000	0.0700 ^a	0.2880 ^a
rain	0.0115 ^b	0.0019 ^b	-0.0850 ^a	-0.0420 ^a	-0.0580 ^a	-0.0700 ^a	0.0029 ^b	0.0500 ^a	0.0700 ^a	1.0000	-0.0500 ^a
ozone	-0.1571 ^b	0.1381 ^b	-0.0260 ^a	-0.3440 ^a	-0.3510 ^a	0.1320 ^a	0.0904 ^b	0.3890 ^a	0.2880 ^a	-0.0500 ^a	1.0000

a: Pearson corr., b: linear-circular corr., c: circular-circular corr.

세면지의 수준이 높을수록 공단지역(측정지역의 남서쪽에 위치함)으로부터의 영향을 크게 받아 오존의 농도가 높아지는 것으로 추측할 수 있다. 또한 이산화질소(nit-diox)와 일산화탄소(car_mono), 강수량(rain)은 오존과 음의 상관을 가짐을 알 수 있다. 원형변수들로 전환된 풍향(wind-dir) 및 시간(month,

Table 3: Multiple regression including circular variables

	Estimate	Std.Error	t value	Pr(> t)
(Intercept)	1.50E-02	1.16E-03	12.917	< 2e-16***
sin_m	1.29E-02	4.81E-04	26.788	< 2e-16***
com_m	3.61E-03	8.53E-04	4.229	2.41e-05***
sin_t	-6.33E-03	3.73E-04	-16.984	< 2e-16***
cos_t	-3.03E-03	3.46E-04	-8.772	< 2e-16***
sin_wd	-2.04E-05	4.18E-04	-0.049	9.61e-01
cos_wd	-5.79E-04	3.58E-04	-1.619	1.05e-01
dust	5.45E-05	5.57E-06	9.792	< 2e-16***
nit_diox	-4.56E-01	1.83E-02	-24.908	< 2e-16***
car_mono	-1.53E-03	9.62E-04	-1.585	1.13e-01
sulf_acid	6.81E-01	7.63E-02	8.936	< 2e-16***
rain	-1.50E-03	3.10E-04	-4.826	1.45e-06***
temp	9.64E-04	6.30E-05	15.307	< 2e-16***
wind_sp	8.79E-04	2.76E-04	3.725	1.98e-04***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '1'
 Residual standard error: 0.0133 on 3692 degrees of freedom
 Multiple R-squared: 0.5326, Adjusted R-squared: 0.531
 F-statistic: 323.7 on 13 and 3692 DF, p-value: < 2.2e-16

time) 변수와 오존 간에는 상대적으로 약한 양의 상관성이 존재함을 알 수 있다. 다만 원형변수들과 오존 간의 상관관계에 대해서는, 원형변수는 방향의 정보만 가질 뿐 그 크기가 의미를 갖지 않으므로, 해석에 주의를 요한다. 위 표에서 음영 처리된 부분은 (오존과의) 양의 상관을, 나머지는 음의 상관을 나타낸다.

3.3.2. 원형-선형회귀분석

다음은 오존의 수준을 반응변수, 기타의 변수들을 설명변수로 하는 회귀분석을 실시하였다. 원형 변수(풍향, 월, 시간)와 기타의 설명변수가 포함된 회귀분석 결과는 Table 3과 같다. 추정된 계수를 살펴보면 온도(temp)와 풍속(wind.sp), 미세먼지(dust)는 양의 계수로, 이산화질소(nit-diox)와 일산화탄소(car_mono), 강우량(rain)은 음의 값으로 추정되었음을 알 수 있다. 이 결과는 상관분석의 결과와도 일치한다. 다만 아황산가스(sulf.acid)의 경우 매우 약한 음의 상관을 보였으나, 회귀계수는 다소 큰 양의 값을 취함을 확인할 수 있다. 상관분석의 경우와 마찬가지로 원형변수(month, time, wind.dir)에 대한 추정값은 크기에 대한 해석에 주의를 요한다.

적합된 회귀모형의 성능을 알아보기 위해, 세 종류의 모형을 추가로 적합하고 그 성능을 비교하였다. Table 4는 적합모형 간의 성능을 비교한 것이다. 모형비교에는 결정계수(R²)와 수정된 결정계수(adjusted R²), 평균절대백분율오차(MAPE; Mean Absolute Percentage Error) 및 제곱근평균오차제곱합(RMSE; Root Mean Squared Error)이 사용되었다. MAPE와 RMSE의 정의는 다음과 같다.

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{\text{관측값}_i - \text{예측값}_i}{\text{관측값}_i} \right| \times 100(\%),$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (\text{관측값}_i - \text{예측값}_i)^2}{N - 1}}.$$

Table 4로부터 세 종류의 원형변수를 모두 포함하는 (모형 4)의 성능이 가장 우수한 것으로 나타났다. 오존예측을 위한 회귀분석에서 방향과 시간변수를 원형변수로 간주하여 모형에 포함시킴으로써

Table 4: Comparison of regression models including circular variables

	model1	model2	model3	model4
R ²	0.3586	0.3721	0.3673	0.5326
Adjusted R ²	0.3573	0.3657	0.3657	0.5310
MAPE	191.5618	187.8827	188.2600	173.1953
RMSE	0.0155	0.0154	0.0154	0.0132
모수의 수	7	14	9	13

model1: linear reg. without wind direction

model2: linear reg. treating wind direction as categorical with 8 categories

model3: linear reg. treating wind direction as circular variable

model4: linear reg. including various circular variable(wind direction, time, month)

Table 5: Circular-circular regression between time and wind direction

coefficient	reponse variable	
	cos(wind_dir)	sin(wind_dir)
(Intercept)	0.3930	0.0775
First-order terms		
cos(time)	0.4061	-0.0966
sin(time)	0.3540	0.7900
<i>p</i> -value for seconde-order	0.3400	0.0102
Second-order terms		
cos(time)	-0.0027	0.2885
sin(time)	-0.1392	-0.0665
<i>p</i> -value for third-order	0.0955	0.8557

상당한 성능 개선이 가능함을 확인하였다. 아울러 (모형 2)와 (모형 3)을 비교해 볼 때, 두 모형 간의 성능에는 큰 차이가 없으나, 모형의 단순성의 측면에서는 (모형 3)이 보다 효율적이라 말할 수 있다.

3.3.3. 원형-원형 회귀분석

여기서는 원형변수 간의 회귀모형의 적합과정을 다루었다. 시간(time)변수가 풍향(wind-dir)에 미치는 영향을 분석하기 위해 앞 절에서 소개한 다음의 모형을 적합하였다.

$$\cos(\theta) = \sum_{k=0}^m (A_k \cos k\alpha + B_k \sin k\alpha) + \epsilon_1,$$

$$\sin(\theta) = \sum_{k=0}^m (C_k \cos k\alpha + D_k \sin k\alpha) + \epsilon_2.$$

분석에 사용된 자료는 2007년 1월 9일과 11일 양일간 관측된 자료만을 이용하였다. Table 5는 위 모형의 적합 결과를 정리한 것이다. 위 모형의 적합과정에서 차수 m 은 2차항까지 유의한 것으로 나타나(p -값 = 0.3400, 0.0102), $m = 2$ 로 결정하였다. Table 5로부터 차수 2($m = 2$)인 적합 모형을 기술하면 다음과 같다.

$$\begin{aligned} \cos(\text{wind_dir}) &= 0.3932 + 0.4061 \cdot \cos(\text{time}) + 0.3540 \cdot \sin(\text{time}) \\ &\quad - 0.0027 \cdot \cos 2(\text{time}) - 0.1392 \cdot \sin 2(\text{time}), \\ \sin(\text{wind_dir}) &= 0.0755 - 0.0966 \cdot \cos(\text{time}) + 0.7900 \cdot \sin(\text{time}) \\ &\quad + 0.2885 \cdot \cos 2(\text{time}) - 0.0665 \cdot \sin 2(\text{time}). \end{aligned}$$

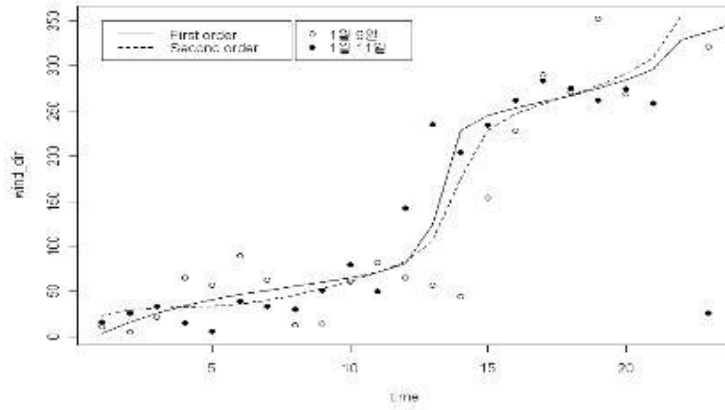


Figure 3: Fitted circular-circular regression

Table 6: Performance of circular-circular regression

	MAPE	RMSE
model 1 ($m = 1$)	108.1162	1.3670
model 2 ($m = 2$)	95.1373	1.1285

위의 결과로부터 풍향(wind_dir)에 대한 예측값은 다음의 식을 통해 추정된다.

$$\hat{\beta} = \arctan^* \frac{\cos(\text{wind_dir})}{\sin(\text{wind_dir})}.$$

Figure 3은 시간에 따른 풍향의 추정 결과를 그림으로 나타낸 것이다.

적합된 모형의 성능을 알아보기 위해, 위에서 적합된 차수 2($m = 2$)인 모형과 차수 1($m = 1$)인 모형과의 비교를 실시하였다. Table 6은 두 모형 간의 성능을 비교한 결과이다. 그 결과 차수 2의 모형이 차수 1의 모형을 상당히 개선하고 있음을 알 수 있다.

4. 결론

지금까지 오존농도의 추정과 관련된 연구에서 풍향은 주로 범주형의 형태로 분석에 이용되어 왔다. 이 경우 범주화를 하는 과정에서 정보의 손실이 발생하며, 아울러 범주의 수준을 어느 정도 세분화하느냐에 따라 모형이 매우 복잡하게 되기도 한다. 본 연구에서는 원형변수에 대한 분석 기법을 이용하여 풍향 변수가 오존농도에 미치는 영향을 분석하였다. 로즈다이어그램을 통한 탐색적 분석을 수행하고, 원형-선형 상관 및 회귀분석을 이용하여 풍향을 비롯한 시계열 관련 정보가 오존농도의 예측에 미치는 영향을 분석하였다. 또한 원형변수들 간의 회귀(풍향에 미치는 시간의 효과)에 대해서도 언급하였다. 본 논문에서 제시한 원형변수를 포함한 회귀와 기존의 회귀 및 시계열 관련 모형 간의 비교는 흥미로운 주제라 생각된다.

References

Batschelet, E. (1981). *Circular Statistics in Biology*, Academic Press, London.
 Fisher, N. I. (1993). *Statistical Analysis of Circular Data*, Cambridge University Press.

- Jammalamadaka, S. R. and Lund, U. J. (2006). The effect of wind direction on ozone levels: A case study, *Environmental Ecological Statistics*, **13**, 287–298.
- Jammalamadaka, S. R. and Sarma, Y. R. (1988). A correlation coefficients for angular variables, *Statistical Theory and Data Analysis II*, 349–364.
- Jammalamadaka, S. R. and Sarma, Y. R. (1993). Circular regression, *Statistical Science and Data Analysis*, 109–128
- Jammalamadaka, S. R. and SenGupta, A. (2001). *Topics in Circular Statistics*, World Scientific.
- Johnson, R. A. and Wehrly, T. E. (1977). Measure and models for angular correlation and angular-linear correlation, *Journal of the American Statistical Association*, **39**, 222–229.
- Mardia, K. V. (1972). *Statistical of Directional Data*, Academic Press, New York.
- Mardia, K. V. (1975). Statistics of directional data (with discussion), *Journal of the Royal Statistical Society, Series B*, **37**, 349–393.
- Mardia, K. V. (1976). Linear-circular correlation coefficients and rhythmometry, *Biometrika*, **63**, 403–405.
- Mardia, K. V. and Puri, M. L. (1978). A robust spherical correlation coefficient against scale, *Biometrika*, **65**, 391–396.
- Rivest, L. P. (1982). Some statistical methods for bivariate circular data, *Journal of the Royal Statistical Society, Series B*, **44**, 81–90.
- Stephens, M. A. (1979). Vector correlation, *Biometrika*, **66**, 41–48.

2012년 5월 22일 접수; 2012년 7월 11일 수정; 2012년 7월 12일 채택