Thesis for the Degree of Master of Engineering

Estimation of surface NO_2 volume mixing ratio using regression models with OMI data

by

Dae Won Kim

Department of Spatial Information Engineering

The Graduate School

Pukyong National University

August 2017

Estimation of surface NO_2 volume mixing ratio using regression models with OMI data

(OMI 자료와 회귀모델들을 이용한 지표 이산화질소 혼합비 추정)

Advisor: Prof. Han Lim Lee

by

Dae Won Kim

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Engineering

in Department of Spatial Information Engineering,
The Graduate School, Pukyong National University

August 2017

Estimation of surface NO_2 volume mixing ratio using regression models with OMI data

A dissertation

by

Dae Won Kim

Approved by:

(Chairman) Kyung Soo Han

(Member) Jae Jin Kim

(Member) Han Lim Lee

July 2017

Contents

List of Figures

List of Tables

OMI 자료와 회귀모델들을 이용한 지표 이산화질소 혼합비 추정

김 대 원

부경대학교 대학원 공간정보시스템공학과

요약

본 연구에서는 처음으로 OMI (Ozone Monitoring Instrument) 자료를 활용하여 세 가지 회귀모델로 13시 45분에서의 지표 이산화질소 혼합비와 월 평균 지표 이산화질소 혼합비를 대한민국 대도시인 서울, 경기, 대전, 광주에서 추정하였다. 지점 장비로 측정된 지표 이산화질소 혼합비와 OMI 센서로 부터 획득한 대류권 이산화질소 수직 칼럼 농도 사이의 관계를 통한 회귀모델과 행성 경계층 높이, 지표면 압력, 지표면 온도, 지표면 이슬점 온도, 지표면 풍향, 지표면 풍속 자료가 고려된 회귀모델을 사용하였다. 본 연구의 회귀모델의 회귀 계수를 결정하기 위한 훈련 기간은 2007년부터 2013년까지이며, 회귀모델의 성능은 2006년, 2014년의 지점 장비로 측정된 지표 이산화질소 혼합비와 비교를 통하여 평가하였다. 세 가지 회귀모델 중에서 다중 회귀모델이 13시 45분의 지표 이산화질소 혼합비와 월 평균 지표 이산화질소 혼합비 추정에 가장 좋은 성능을 보였다. 검증년도에서 다중 회귀모델로 추정된 13시 45분의 지표 이산화질소 혼합비와 지점 측정장비로 측정된 지표 이산화질소 혼합비사이의 평균 R (correlation coefficient), 기울기, MB (mean bias), MAE (mean absolute error), RMSE (root mean square error), Percent difference는 각각 0.66, 0.41, -1.36 ppbv, 6.89 ppbv, 8.98 ppbv, 31.50% 이다. 반면에 다른 두 회귀모델로 추정된 13시 45분의 지표 이산화질소 혼합비와 지점 측정 장비로 측정된 지표 이산화질소 혼합비 사이의 평균 R, 기울기, MB, MAE, RMSE, Percent difference는 각각 0.75, 0.41, -1.40 ppbv, 3.59 ppbv, 4.72 ppbv,

and 16.59% 이다. 월 평균 지표 이산화질소 혼합비 추정에 있어서는 다중 회귀모델과 다른 두 회귀모델이 비슷한 성능을 보였다. 세 가지 회귀모델로 추정된 월 평균 지표 이산화질소 혼합비와 지점 측정 장비로 측정된 지표 이산화질소 혼합비 사이의 평균 R, 기울기, MB, MAE, RMSE, Percent difference는 각각 0.74, 0.49, -1.90 ppbv, 3.93 ppbv, 5.05 ppbv, 18.76% 이다.

1. Introduction

A main anthropogenic sources of nitrogen dioxide (NO_2) is fossil fuel combustion while natural sources of NO_2 are lightning, forest fire, and soil emissions [IPCC, 2007; Van der et al., 2008]. In particular, since NO_2 is emitted in large quantities in automobile exhaust gas, NO_2 is often used as an indicator for traffic-related air pollution in urban areas [Kharol et al., 2015]. In terms of its effect on human health, Long-term NO_2 exposure can lead to pulmonary depression and respiratory illness [Ackermann-Liebrich et al., 1997; Schindler et al., 1998; Gauderman et al., 2000; Panella et al., 2000; Smith et al., 2000]. In addition, it is precursors of aerosol nitrate, tropospheric ozone, and the hydroxyl radical (OH), the main atmospheric oxidant [Boersma et al., 2009]. Therefore, NO_2 is measured by various methods and chemiluminescence is one of the well known methods for measuring surface NO_2 volume mixing ratio (VMR) [Demerjian, 2000]. In-situ measurement such as the chemiluminescence measurement method is, in general, more accurate than remote sensing techniques, but it requires a number of in-situ instruments to provide the spatial distribution of the NO_2 VMR in high resolution. In recent years, the NO_2 vertical column density (VCD) measurements via satellites that can monitor NO_2 of global scale in a short time has been actively conducted. The following various satellite sensors have been utilized to measure these regional or global NO_2 distributions. Space-born sensors that observed global distributions of NO_2 are Global Ozone Monitoring Experiment (GOME) (1995–2003) aboard

European Remote Sensing-2 (ERS-2), Scanning Imaging Absorption Spectrometer for Atmospheric Chartography/Chemistry (SCIAMACHY) aboard Environmental Satellite (Envisat) (2002~), Ozone Monitoring Instrument (OMI) aboard EOS-AURA (2004~), and GOME-2 aboard Meteorological operational satellite (MetOp)-A platform (2007~) and MetOp-B platform (2012~) [Leue et al., 2001; Richter and Burrows, 2002; Martin et al., 2002; Boersma et al., 2004; Boersma et al., 2007; Bucsela et al., 2006]. In many countries, the NO_2 VCD obtained from satellites is can't be directly used for air quality regulation because the surface NO_2 VMR is used for air quality regulation. In recent years, studies have been conducted to investigate the feasibility of estimating the surface NO_2 VMRs using the NO_2 VCD obtained from satellite measurements and the correlation between the NO_2 VCD obtained from satellite measurements and the surface NO_2 VMRs.

ORDÓNEZ et al., (2006) reported the correlation between the troposphere NO_2 VCD and the NO_2 VCD measured by GOME and ground based in-situ device in Milan. Kharol et al., (2015) estimated the annual average ground-level NO_2 concentrations in North America using chemical transport model (GEOS-Chem) data and OMI NO_2 columns. It also reported the annual trend of the estimated ground-level NO_2 concentrations. However, no studies have attempted to estimate the surface NO_2 VMR in higher temporal resolution such as hourly and monthly using the NO_2 VCD measured by satellites.

In this present study, we for the first time estimated the surface NO_2 VMR at a specific time (13:45 Local time; LT) (NO_2 $VMR_{ST\text{-}estimate}$) and monthly mean surface NO_2 VMR (NO_2 $VMR_{M\text{-}estimate}$) using two linear regression models and a multiple regression model with the troposphere NO_2 VCD obtained from OMI (Trop NO_2 VCD_{OMI}) in five metropolitan cities. In addition, performances of each regression method were evaluated by comparing the estimated surface NO_2 VMRs with those obtained from in-situ measurement (NO_2 $VMR_{In\text{-}situ}$).

2. Studay area & period

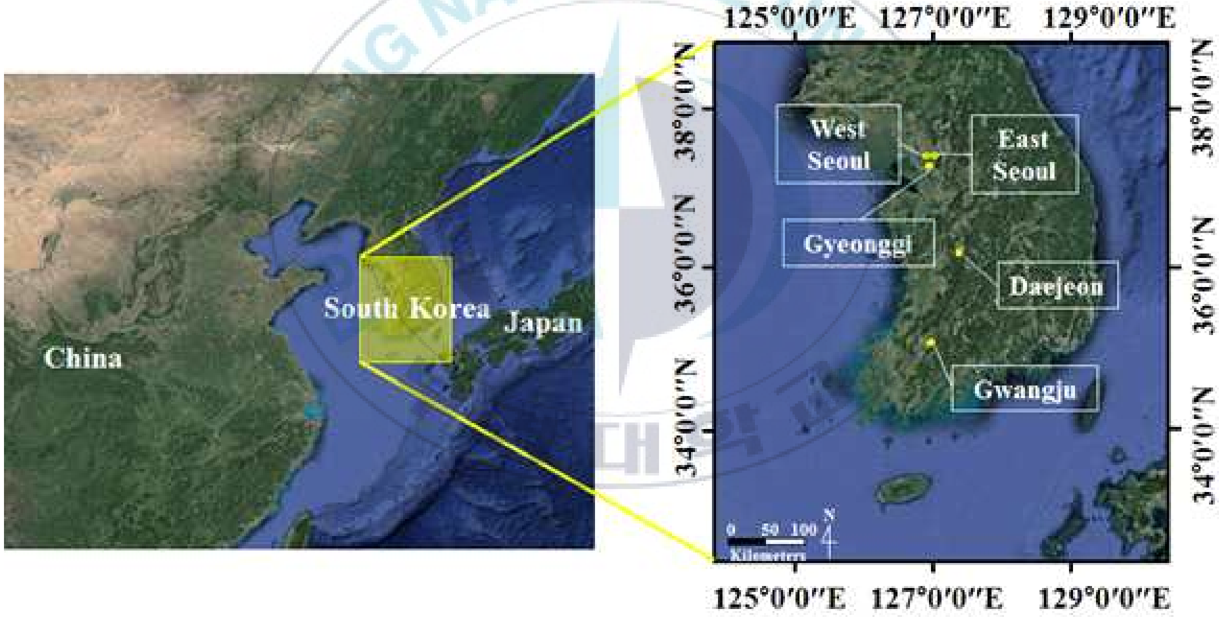


Figure 1. Study areas are located in South Korea.

The study areas were selected where the surface NO_2 VMR was continuously measured in Korean metropolitan cities. Metropolitan cities such as Busan and Incheon where the OMI pixel covers both the sea and cities are excluded since there is no surface NO_2 data available over the sea. Therefore, the selected areas include Seoul, Gyeonggi, Daejeon, and Gwangju. Among the study areas, Seoul, where four OMI pixels exist, is divided into eastern and western areas (West Seoul and East Seoul). The study period is nine years from 2006 to 2014. Seven years (2007 - 2013) are the training period to determine the regression coefficients of the regression models used in this study, whereas two years (2006, 2014) the validation period when for the surface NO_2 VMRs estimated from three regression models with the determined regression coefficients are evaluated via comparison with the in-situ data. The three regression models used in this study are described in detail in Section 3.

2.1. Data

The data used in this study include Trop NO_2 VCD_{OMI}, boundary layer height obtained from Atmospheric Infrared Sounder (AIRS) (BLH_{AIRS}), atmospheric temperature obtained from AIRS ($Temp_{AIRS}$), pressure obtained from AIRS ($Press_{AIRS}$), NO_2 $VMR_{In\text{-}situ}$, surface temperature obtained from in-situ measurement ($Temp_{In\text{-}situ}$), surface pressure from in-situ measurement ($Press_{In\text{-}situ}$), surface dew point from in-situ measurement ($Dewpoint_{In\text{-}situ}$), surface wind speed from in-situ measurement ($WS_{In\text{-}situ}$), and surface wind direction from in-situ measurement ($WD_{In\text{-}situ}$). The detailed information of the data are summarized in Table 1.

Table 1. Satellite and In-situ data used in this study.

	Data		Time
Satellite	Trop. NO_2 VCD	OMI Level3 NO_2 Daily data (OMNO2d)	13:45
	BLH, Temperature, Pressure	AIRS/Aqua L3 Daily Support Product (AIRS + AMSU) V006 (AIRX3SPD)	13:30
In-situ	Surface NO_2 VMR	Air Korea	13:00 and 14:00
	Surface Temperature, Surface Pressure, Surface Dewpoint, Surface Wind Speed, Surface Wind Data	AWS (Automatic Weather System)	

2.1.1. Ozone Monitoring Instrument (OMI) data

The OMI sensor was launched in July 2004. It measures hyperspectral radiance in ultraviolet and visible wavelength range via push-broom. Since the OMI sensor utilizes the hyperspectral feature, it can improve retrieval accuracy of air pollutants and enable the OMI sensor to do precise radiometric and wavelength calibration for a long time. The OMI sensor has two ultraviolet channels including UV-1 (270 nm - 314 nm) and UV-2 (310 nm - 365 nm). The spectral resolutions of UV-1 and UV-2 are 0.42 nm and 0.45 nm, respectively. The wavelength range and spectral range of visible channel are from 350 nm to 500 nm and 0.63 nm, respectively.

The OMI sensor continues the heritage of the TOMS dataset. The OMI sensor onboard Aura satellite is known as a main sensor to monitor ozone hole and observes key air pollutants including nitrogen dioxide, sulfur dioxide, and aerosols. The Aura is a polar orbiting satellite with an overpass time of 13:45LT.

The Trop NO_2 VCD_{OMIs} were obtained from OMI Level3 NO_2 Daily Data (OMNO2d) provided by NASA Goddard Earth Sciences Data and Information Services Center (http://disc.sci.gsfc.nasa.gov/Aura/data-holdings/OMI). Cloud-screened NO_2 data (Level-3 OMI NO_2 Cloud-Screened Total and Tropospheric Column NO_2 (V003)) are used in this present study (Cloud Fraction < 30 %).

2.1.2. Atmospheric Infrared Sounder (AIRS) data

The BLH_{AIRS}, $Temp_{AIRS}$, and $Press_{AIRS}$ used in this study were obtained from the AIRS / Aqua L3 Daily Support Product (AIRS + AMSU) 1 degree x 1 degree V006 (AIRX3SPD.00) of NASA Goddard Earth Sciences Data and Information Services Center (http://disc.sci.gsfc.nasa.gov/uui/datasets/AIRX3SPD_V006/summary?keywords=%22AIRS%22). The AIRS / Advanced Microwave Sounding Unit (AMSU) is a sounding suite that aboard the Aqua launched in May 2002 [Aumann et al., 2003; Chahine et al., 2006]. The Aqua is a polar orbiting satellite with an overpass time of 13:30 local time (day and night time) and a spatial resolution of 40 km horizontal at nadir.

2.1.3. In situ NO_2 data

The NO_2 $VMR_{In\text{-}situ}$ was obtained from Air Korea (http://www.airkorea.or.kr/last_amb_hour_data). Since NO_2 $VMR_{In\text{-}situ}$ is available hourly, the average value of 13:00 LT and 14:00 LT is used to be closer the OMI overpassing time. The Previous study [ORDÓNEZ et al., 2006], the in-situ measurement sited was grouped into five different NO_2 levels, clean, slightly polluted, average polluted, polluted, and heavily polluted account for many stations which are located close to streets and are exposed to emissions. In addition, the NO_2 data obtained from many in-situ measurement stations in the GOME pixels were averaged, since in-situ measurements are only representative of a small fraction of the satellite ground scene. In this present study, the NO_2 $VMR_{In\text{-}situs}$ obtained from in-situ measurements located close to streets were excluded in this study. We used the average of three or more the NO_2 $VMR_{In\text{-}situ}$ located at least 2 km distance away from each other.

2.1.4 In situ meteorological data

The $Temp_{In\text{-}situ}$, $Press_{In\text{-}situ}$, $Dewpoint_{In\text{-}situ}$, $WS_{In\text{-}situ}$, and $WD_{In\text{-}situ}$ used in this study are Automatic Weather System (AWS) data provided by Korea Meteorological Administration (http://sts.kma.go.kr/jsp/home/contents/statistics/newStatisticsSearch.do?menu=SFC&MNU=MNU). Since meteorological data are available hourly, the mean values of the data at 13:00 LT and 14:00 LT are used.

3. Methodology

In this study, NO_2 $VMR_{ST\text{-}estimate}$ and NO_2 $VMR_{M\text{-}estimate}$ were estimated using three regression models with Trop NO_2 VCD_{OMI} Table 2 shows the summary of the three models used to estimate NO_2 $VMR_{ST\text{-}estimate}$ and NO_2 $VMR_{M\text{-}estimate}$.

Table 2. The regression models used for surface NO_2 VMR estimation in this study.

Model		Equation
M1	13:45LT & Monthly	$NO_2\,VMR_{Insitu} = a^{*}\,Trop\,NO_2\,VCD_{OMI} + b$
M2	13:45LT & Monthly	$NO_2\,VMR_{Insitu} = a^{**}\,BLH\,NO_2\,VMR_{OMI} + b$
M3	13:45LT	Section 3, Multiple regression equation (1)
M4	Monthly	

Notes: *NO_2 tropospheric vertical column density obtained from OMI.

$$^{**}BLH\,NO_2\,VMR_{OMI} = \frac{Trop\,NO_2\,VCD_{OMI} \times ^{***}Gas\,constant\,R \times ^{****}Temp_{AIRS} \times 10^{13}}{^{*****}Avogadro\,constant\,NA \times ^{******}BLH_{AIRS} \times ^{*******}\mathrm{Press}_{AIRS}}$$

$$^{***}Gas\,constant = 8.314472\,m^{3}\,pa\,k^{-1}\,mol^{-1}$$

****Boundary layer mean temperature (K) obtained from AIRS.

***** $Avogadro\,constant = 6.022 \times 10^{23} mol^{-1}$

******Boundary layer height (m) obtained from AIRS.

*******Boundary layer mean pressure (pa) obtained from AIRS.

3.1 M1

M1 is the liner regression equation where Trop NO_2 VCD_{OMI} used as an independent variable. Figure 2 shows the linear regression between Trop NO_2 VCD_{OMI} and NO_2 $VMR_{In\text{-}situ}$ at 13:45LT during the training period. In figure 2, R^2 (coefficient of determination), slope, and intercept are 0.47, 0.80 and 11.47, respectively. Figure 3 shows the linear regression between monthly mean Trop NO_2 VCD_{OMI} and monthly mean NO_2 $VMR_{In\text{-}situ}$ during the training period. In figure 3, R^2, slope, and intercept are 0.62, 0.77, and 10.95, respectively. The final form of the equation M1 to estimate NO_2 $VMR_{ST\text{-}estimates}$ is shown Table 4. Whereas, the final form of the equation M1 to estimate NO_2 $VMR_{M\text{-}estimates}$ is shown Table 5.

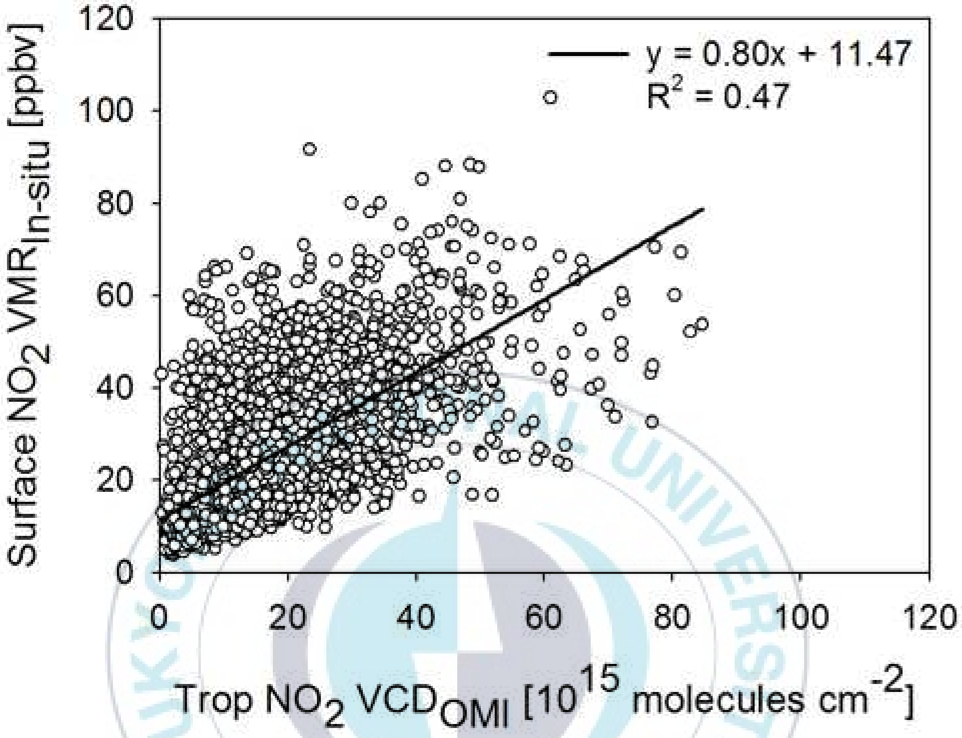


Figure 2. Correlation between Trop NO_2 VCD_{OMI} and NO_2 $VMR_{In-situ}$ to determine regression coefficient for the equation, M1 for the training period between 2007 and 2013.

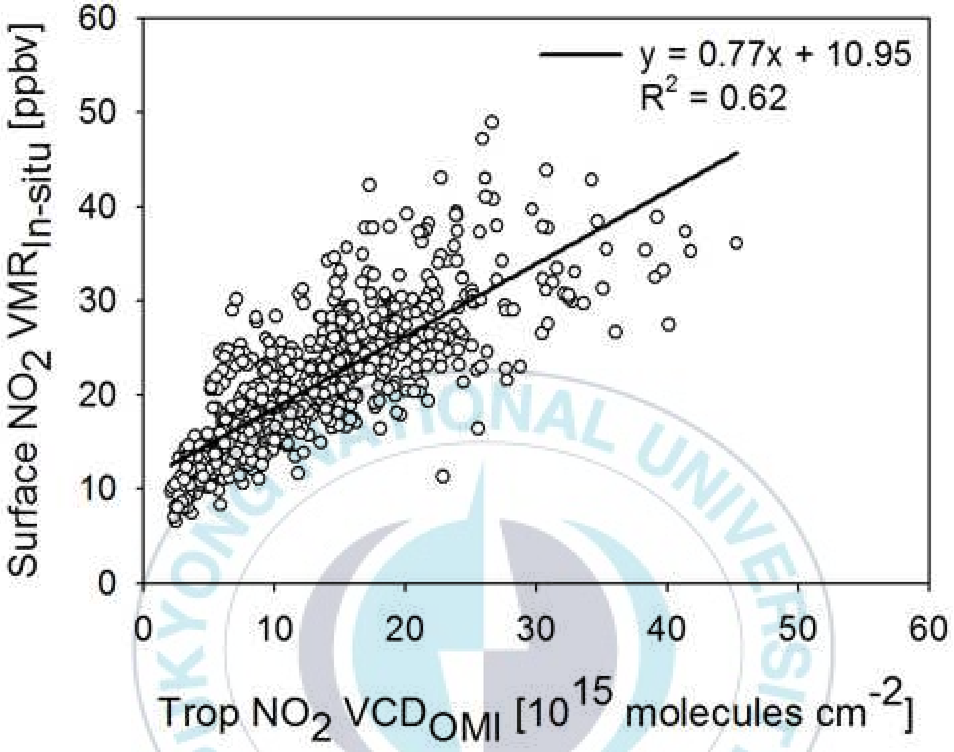


Figure 3. Correlation between monthly mean Trop NO_2 VCD_{OMI} and monthly mean NO_2 $VMR_{In-situ}$ to determine regression coefficient for the equation, M1 for the training period between 2007 and 2013.

3.2 M2

Assuming that Trop NO_2 VCD_{OMI} is mostly present within the planetary boundary layer (PBL), the relationship between Trop NO_2 VCD_{OMI} and the surface NO_2 VMR may change as the planetary boundary layer varies. To reflect the BLH in the regression equation, Trop NO_2 VCD_{OMI} is first divided by BLH_{AIRS} to calculate the NO_2 concentration in the PBL and then converted to the NO_2 mixing ratio in PBL (BLH NO_2 VMR_{OMI}) using $Temp_{AIRS}$ and $Press_{AIRS}$ (see Table 2). Figure 4 shows the linear regression between BLH NO_2 VMR_{OMI} and NO_2 $VMR_{In\text{-}situ}$ at 13:45LT during the training period. In figure 4, R^2, slope and intercept are 0.38, 1.58, and 14.30, respectively. Figure 5 shows the linear regression between monthly mean BLH NO_2 VMR_{OMI} and monthly mean NO_2 $VMR_{In\text{-}situ}$ during the training period. In figure 5, R^2, slope and intercept are 0.59, 1.71, and 12.75, respectively. The final form of the equation M2 to estimate NO_2 $VMR_{ST\text{-}estimates}$ is shown Table 4. Whereas, the final form of the equation M2 to estimate NO_2 $VMR_{M\text{-}estimates}$ is shown Table 5.

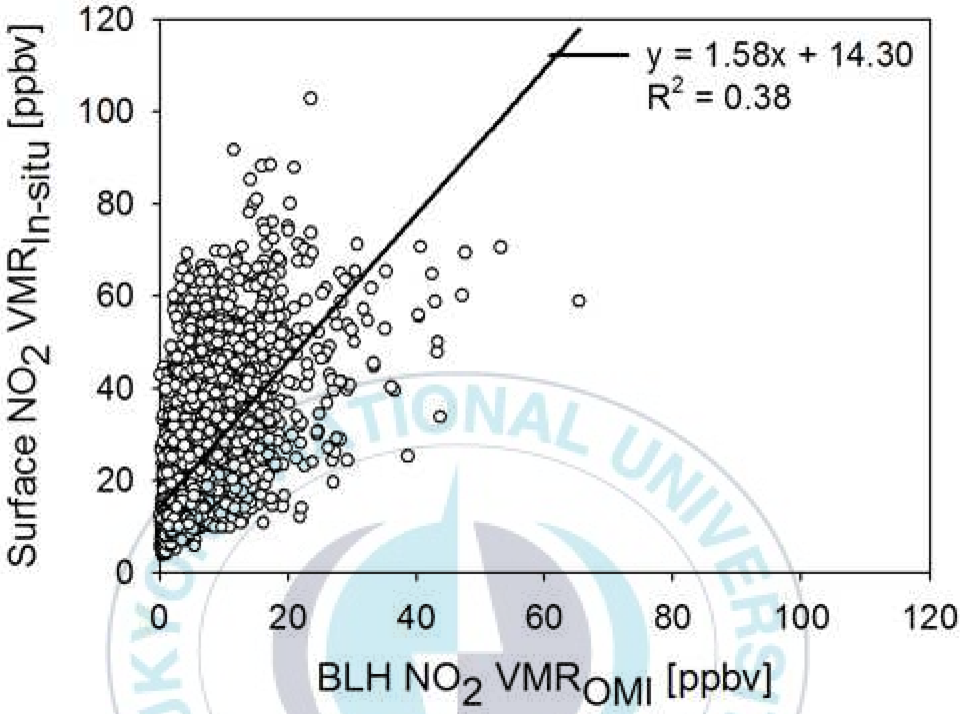


Figure 4. Correlation between BLH NO_2 VMR_{OMI} at specific time(13:45LT) and NO_2 $VMR_{In-situ}$ to determine regression coefficient for the equation, M2 for the training period between 2007 and 2013.

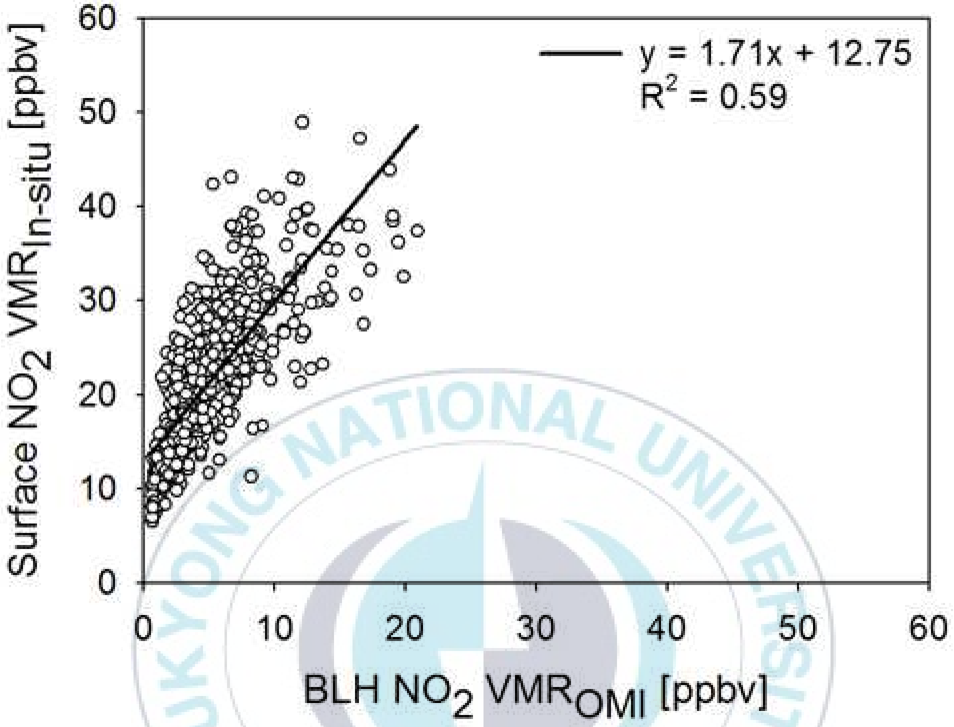


Figure 5. Correlation between monthly mean BLH NO_2 VMR_{OMI} and monthly mean NO_2 $VMR_{In\text{-}situ}$ to determine regression coefficient for the equation, M2 for the training period between 2007 and 2013.

3.3 M3 & M4

M3 and M4 are multiple regression equations to estimate NO_2 $VMR_{ST\text{-}estimates}$ and NO_2 $VMR_{M\text{-}estimates}$. Multiple regression equation consists of a dependent variable, independent variables, and their regression coefficients. For the independent variable candidates, in addition to Trop NO_2 VCD_{OMI} and BLH_{AIRS}, meteorological factors (surface temperature, surface dew point, atmospheric pressure, wind direction, and wind speed) are used as independent variable candidates for the multiple regression equation in this present study. In previous study [Xue and Yin, 2014], these meteorological factors were also used for the multiple regression equation as independent variable candidates to estimate surface SO_2 concentration in Shanghai, China.

The multiple regression equation can be defined as the following equations:

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \epsilon \tag{1}$$

where $\hat{y}$ and β_0 are dependent variable (NO_2 $VMR_{In\text{-}situ}$) and regression coefficient, x_1, x_2,...x_n are the candidates of independent variables (Trop NO_2 VCD_{OMI}, $Dewpoint_{In\text{-}situ}$, $Press_{In\text{-}situ}$, $Temp_{In\text{-}situ}$, BLH_{AIRS}, $WS_{In\text{-}situ}$, $WD_{In\text{-}situ}$), β_1, β_2, …, β_n are the regression coefficients of the independent variables, and ε is the difference between observations (NO_2 $VMR_{In\text{-}situ}$) and estimates values (NO_2 $VMR_{estimates}$). The regression coefficients can be estimates by the least square fitting (Equation 2).

$$\sum_{j=1}^{m} \epsilon_j^2 = \sum_{j=1}^{m} (y_j - \hat{y_j})^2 \qquad (2)$$

Where y_j is observed value with m numbers of data points. By minimizing the sum of ε^2, regression coefficients can be derived. To determine independent variable (x_n) and regression coefficients (β_n) included in the final form of the equations M3 and M4, we considered variation inflation factor (VIF) and p-value to ensure statistical significance of those variables and their coefficient. First, we examined the VIF that explains the multicollinearity of an independent variable candidate with regard to other independent variable candidates. The VIF of the j-th independent variable is expressed as:

$$VIF(x_j) = \frac{1}{1 - R_j^2} \qquad (3)$$

Where R^2_j is the coefficient of determination for the regression of x_j against the other (a regression that does not involve the dependent variable j). The VIF indicates how much x_j is correlated with the variables. A candidate for independent variables with a very high VIF can be considered redundant and should be removed from the multiple regression equations. The candidates for independent variables that do not satisfy the criterion VIF < 10 [Kutner et al., 2004], were excluded from the independent variables. p-value was also used to select independent variables. The highest, still statistically significant p-level was shown by Sellke et al. (2001) to be 5%. Among the independent variables that satisfy the VIF criterion, those that also satisfy the p-value less than 0.05 ($p\text{-value} < 0.05$) are selected as final independent variables in the multiple regression equations. The lists of independent variables selected for the equations M3 and M4 are shown in Table 3. The final form of the equation M3 to estimate NO_2 $VMR_{ST\text{-}estimates}$ is shown Table 4. Whereas, the final form of the equation M4 to estimate NO_2 $VMR_{M\text{-}estimates}$ is shown Table 5.

Table 3. Final independent variables included in multiple regression equations (M3 and M4).

	Final included independent variables	p-value	VIF
M3	Trop NO_2 VCD_{OMI}	0	1.26
	*****$Temp_{In\text{-}situ}$	0.000032	7.02
	*$Dewpoint_{In\text{-}situ}$	0.000306	7.16
	**$Press_{In\text{-}situ}$	0.009981	3.14
	BLH_{AIRS}	1.73×10^{-15}	1.12
	***$WS_{In\text{-}situ}$	3.86×10^{-133}	1.33
	****$WD_{In\text{-}situ}$	1.7493×10^{-38}	1.07
M4	Trop NO_2 VCD_{OMI}	2.4832×10^{-89}	1.64
	*$Dewpoint_{In\text{-}situ}$	0.000421	6.47
	**$Press_{In\text{-}situ}$	0.034582	6.65
	BLH_{AIRS}	0.000834	2.32
	***$WS_{In\text{-}situ}$	3.86×10^{-133}	1.59
	****$WD_{In\text{-}situ}$	1.699×10^{-7}	1.25

Notes: *Surface dewpoint

**Surface pressure

***Surface wind speed

****Surface wind direction

*****Surface temperature

Table 4. Final form of the regression models used for surface NO_2 VMR at specific time estimations and R^2 obtained from the regression between NO_2 $VMR_{In-situ}$ and each model's independent variable for the training period.

		Equation	R^2
13:45LT	M1	$NO_2\,VMR_{ST-estimate} = 1.71 \times Trop\,NO_2\,VCD_{OMI} - 0.68$	0.47
	M2	$NO_2\,VMR_{ST-estimate} = 4.19 \times BLH\,NO_2\,VMR_{OMI} + 1.57$	0.38
	M3	$NO_2\,VMR_{ST-estimate} = 0.000602 \times Trop\,NO_2\,VCD_{OMI}$ $- 0.000107 \times Temp_{Insitu} - 0.000083 \times Dewpoint_{Insitu}$ $+ 0.000061 \times \mathrm{Press}_{Insitu} - 0.000002 \times BLH_{AIRS}$ $- 0.002435 \times WS_{Insitu} + 0.001190 \times WD_{Insitu} - 0.039996$	0.47

Table 5. Final form of the regression models used for monthly mean surface NO_2 VMR estimations and R^2 obtained from the regression between NO_2 $VMR_{In-situ}$ and each model's independent variable for the training period.

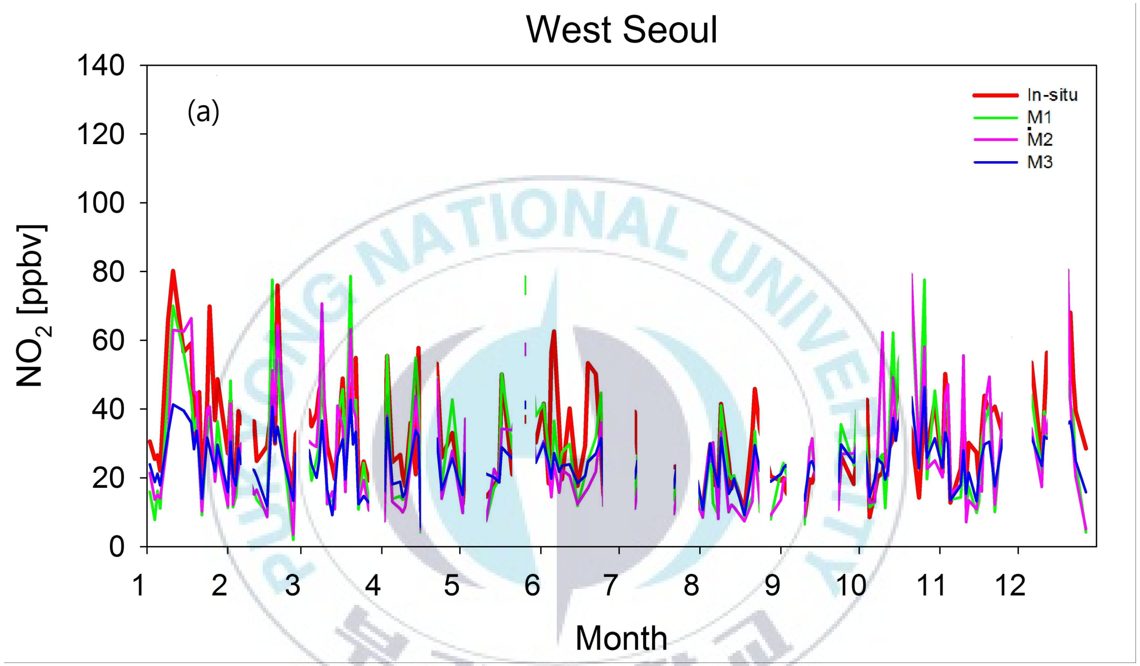
		Equation
Monthly mean	M1	$NO_2\,VMR_{M-estimate} = 1.23 \times Trop\,NO_2\,VCD_{OMI} + 4.74$
	M2	$NO_2\,VMR_{M-estimate} = 2.92 \times BLH\,NO_2\,VMR_{OMI} + 6.74$
	M4	$NO_2\,VMR_{M-estimate} = 0.657241 \times Trop\,NO_2\,VCD_{OMI} - 0.137334 \times Dewpoint_{Insitu} + 0.136096 \times \mathrm{Pr}ess_{Insitu} - 0.004331 \times BLH_{AIRS} - 0.770356 \times WS_{Insitu} + 2.370956 \times WD(west)_{Insitu} + 157.361668$

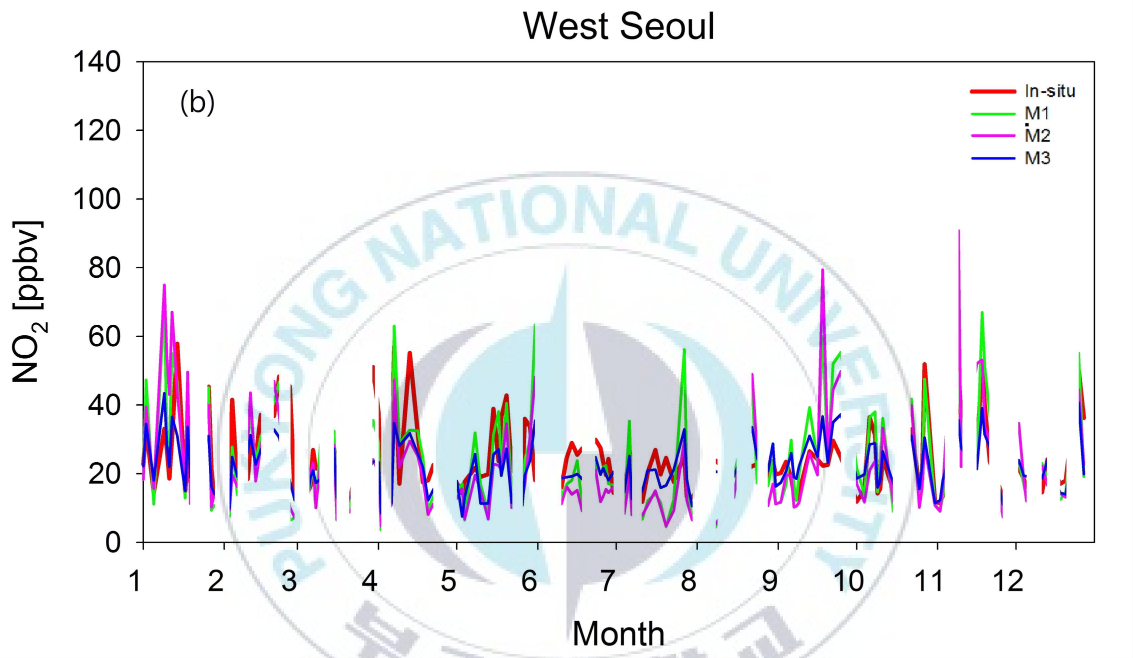
Table 4 and Table 5 show equations of M1, M2, M3, and M4, which reflect the regression coefficients determined from the training period.

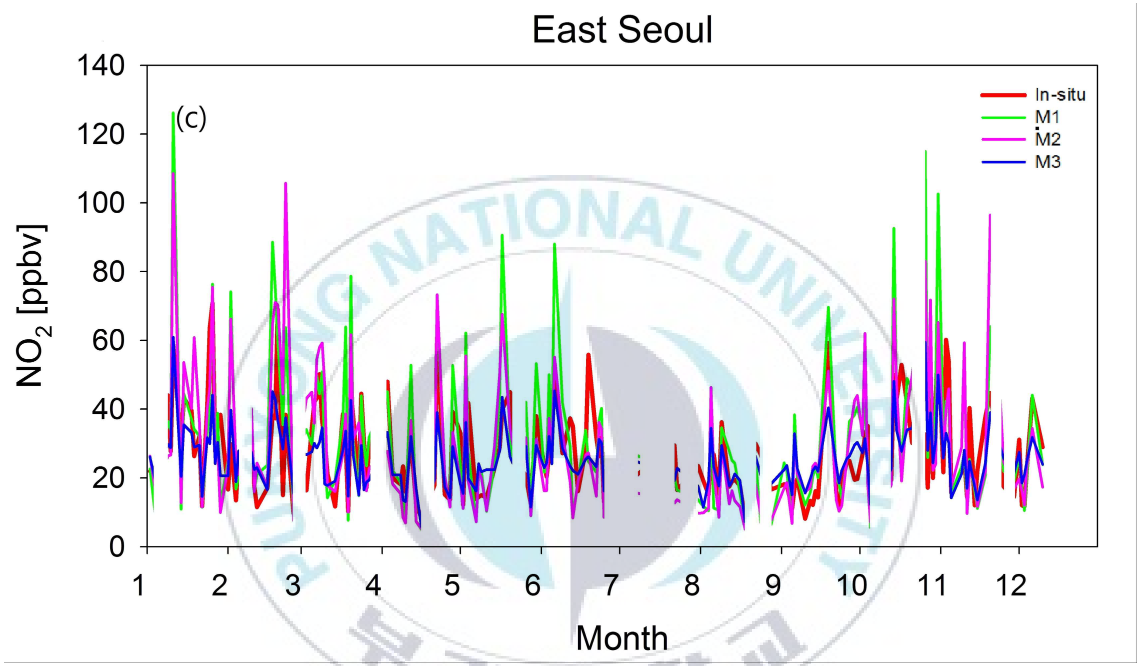
4. Results

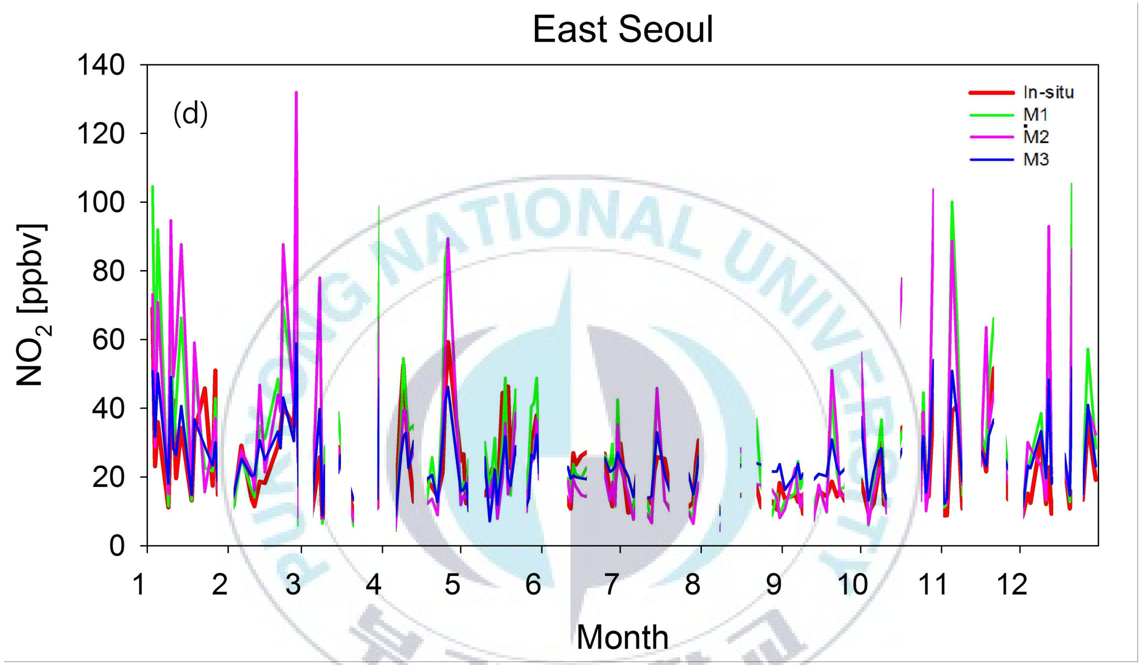
4.1 Daily estimation

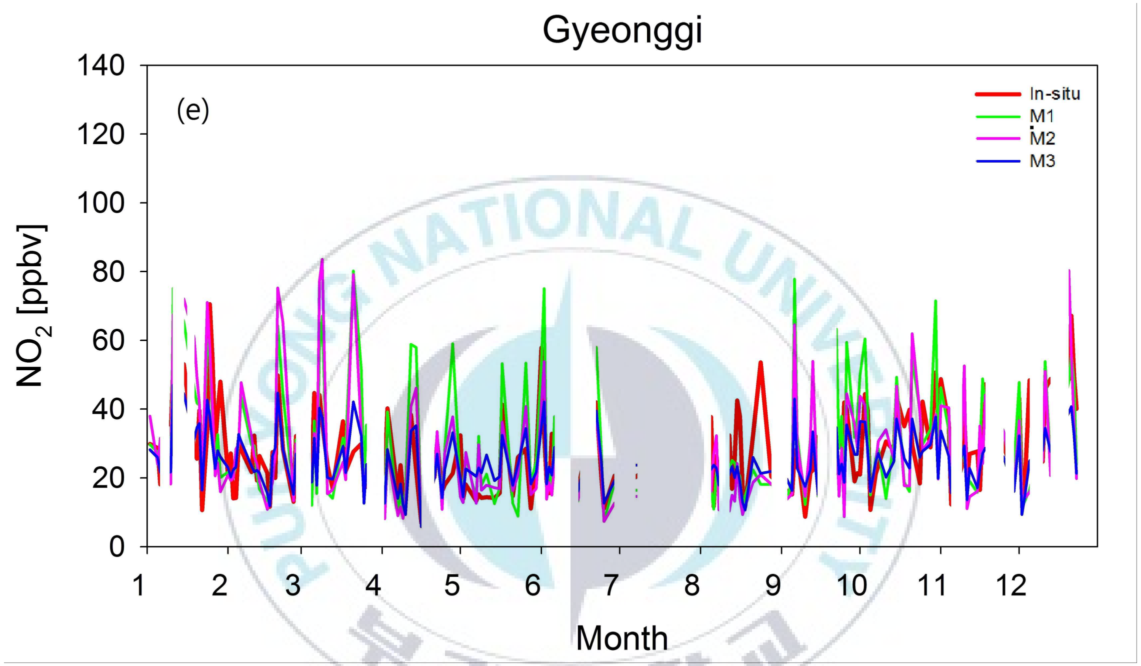
Figure 6 shows daily variations of NO_2 $VMR_{In\text{-}situ}$ and NO_2 $VMR_{ST\text{-}estimates}$ estimated at 13:45LT using the M1, M2 and M3 of Table 4 in from West Seoul and East Seoul. The NO_2 $VMR_{ST\text{-}estimates}$ are estimated using M1, M2, and M3 in Table 4 with the inputs of independent variable data for the years 2006 and 2014, the validation period. A slightly large difference in magnitude is found between NO_2 $VMR_{In\text{-}situ}$ and NO_2 $VMR_{ST\text{-}estimates}$ obtained from M3 compared to those between NO_2 $VMR_{In\text{-}situ}$ and NO_2 $VMR_{ST\text{-}estimates}$ obtained from M1 and M2. However, NO_2 $VMR_{ST\text{-}estimates}$ obtained from M3 showed good agreement with NO_2 $VMR_{In\text{-}situ}$ in hourly pattern. The graph showing the same as figure 6.

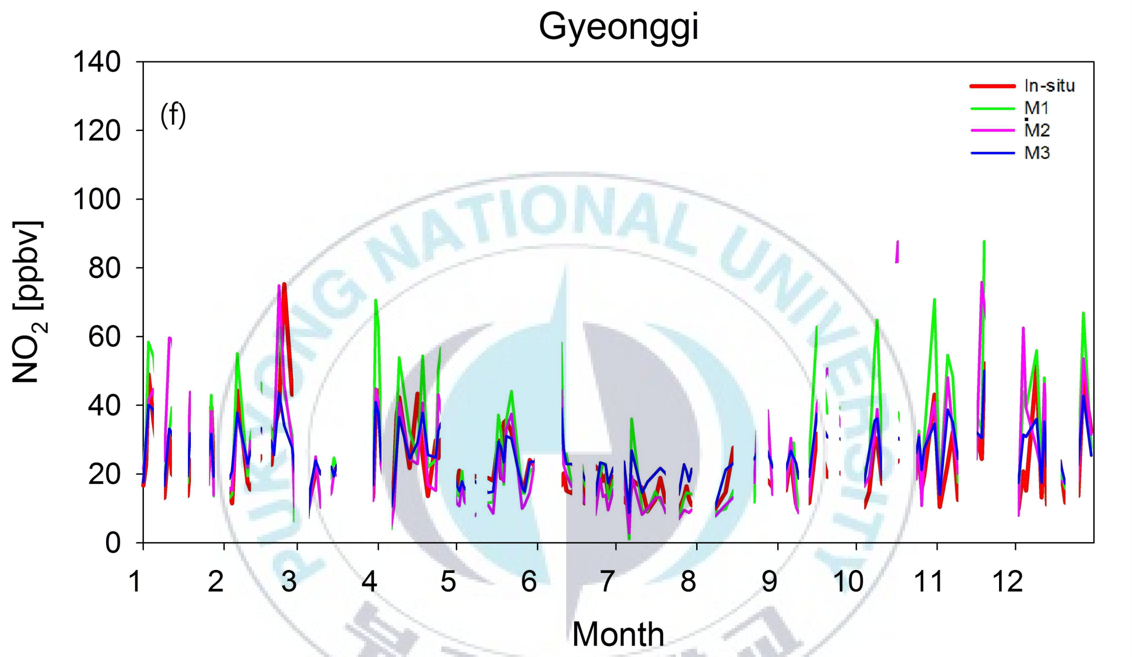

West Seoul
(a)
In-situ
M1
M2
M3
NO_2 [ppbv]
140
120
100
80
60
40
20
0
1
2
3
4
5
6
7
8
9
10
11
12
Month

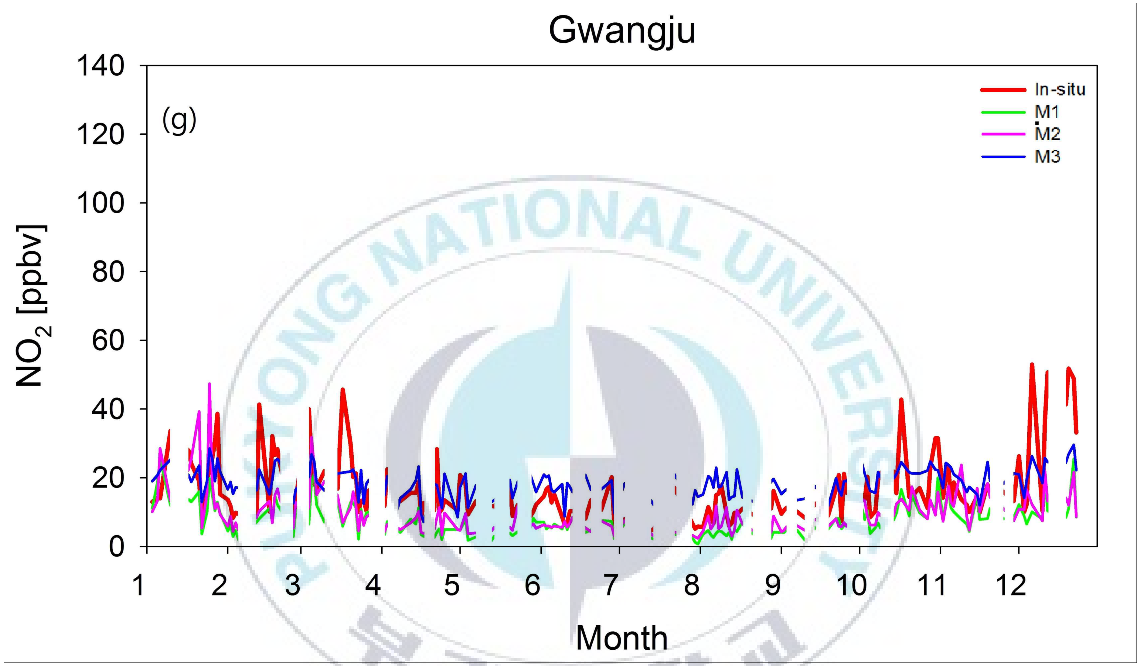

West Seoul
(b)
In-situ
M1
M2
M3
NO2 [ppbv]
0
20
40
60
80
100
120
140
1
2
3
4
5
6
7
8
9
10
11
12
Month


East Seoul
140
120
100
80
60
40
20
0
(c)
In-situ
M1
M2
M3
NO2 [ppbv]
1
2
3
4
5
6
7
8
9
10
11
12
Month

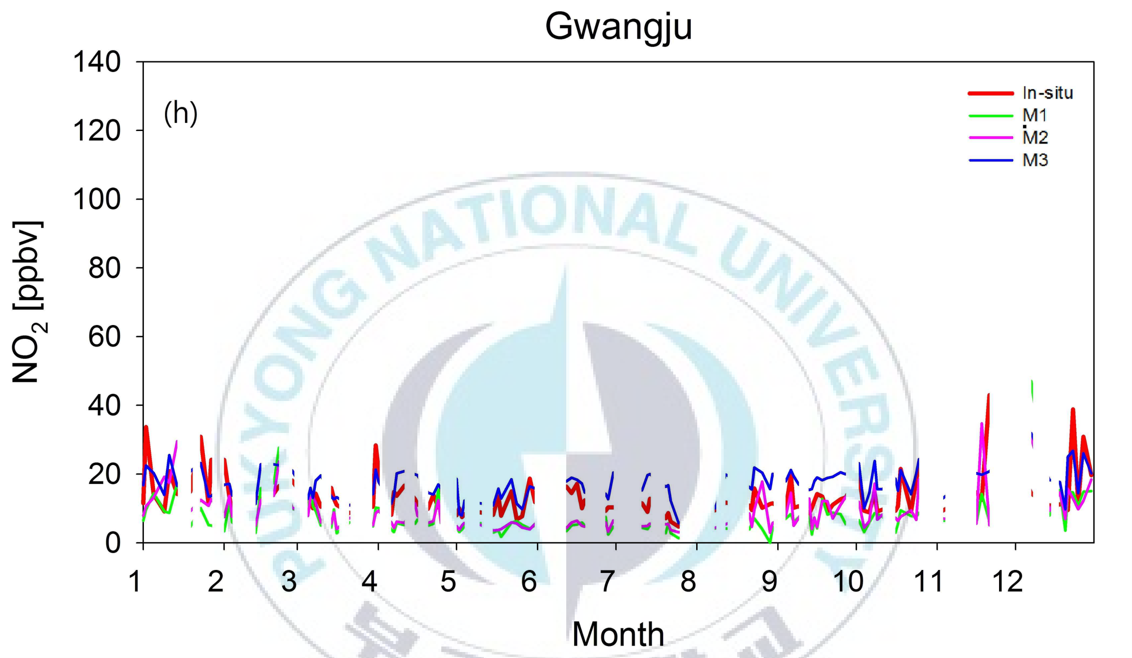

East Seoul
(d)
In-situ
M1
M2
M3
NO_2 [ppbv]
140
120
100
80
60
40
20
0
1
2
3
4
5
6
7
8
9
10
11
12
Month

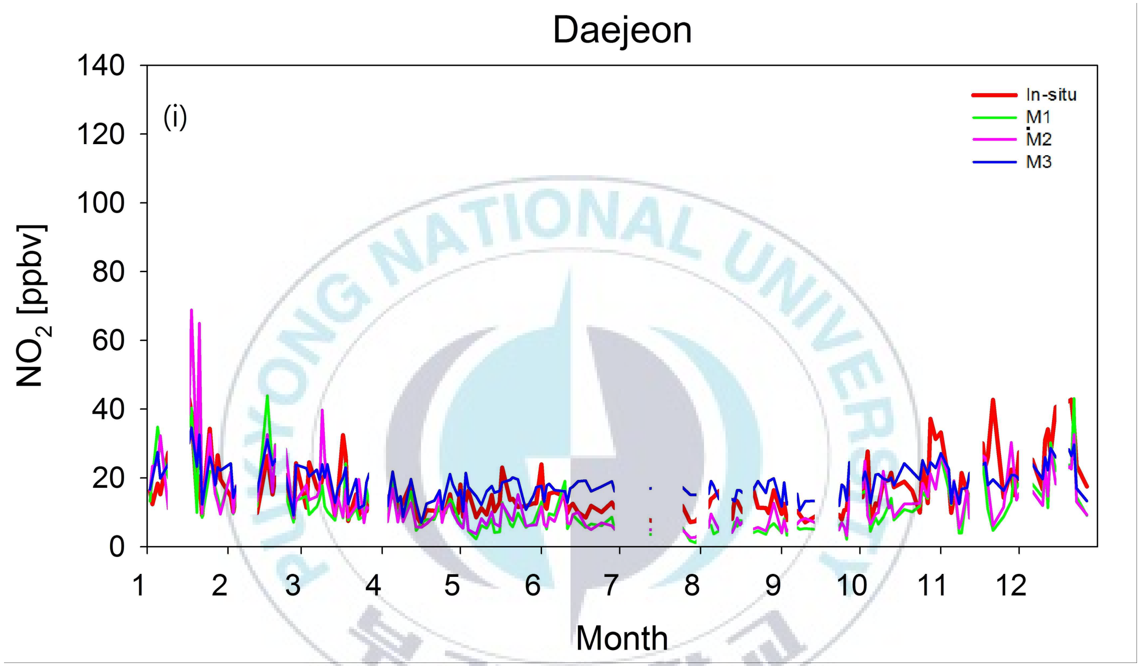

Gyeonggi
(e)
In-situ
M1
M2
M3
NO_2 [ppbv]
0
20
40
60
80
100
120
140
1
2
3
4
5
6
7
8
9
10
11
12
Month


Gyeonggi
(f)
In-situ
M1
M2
M3
NO2 [ppbv]
0
20
40
60
80
100
120
140
1
2
3
4
5
6
7
8
9
10
11
12
Month



Gwangju
(g)
In-situ
M1
M2
M3
NO2 [ppbv]
140
120
100
80
60
40
20
0
1
2
3
4
5
6
7
8
9
10
11
12
Month


Gwangju
(h)
In-situ
M1
M2
M3
NO2 [ppbv]
140
120
100
80
60
40
20
0
1
2
3
4
5
6
7
8
9
10
11
12
Month


Daejeon
(i)
In-situ
M1
M2
M3
NO2 [ppbv]
0
20
40
60
80
100
120
140
1
2
3
4
5
6
7
8
9
10
11
12
Month

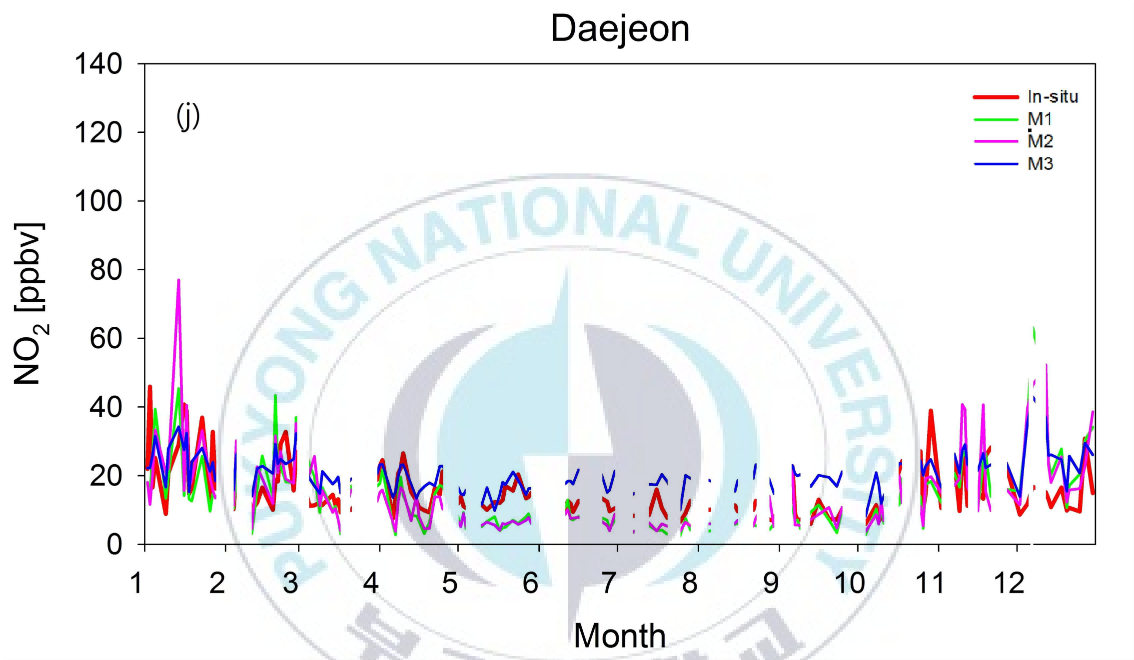


Figure 6. Time series of NO_2 $VMR_{In-situ}$ and NO_2 $VMR_{ST-estimates}$ estimates by M1, M2 and M3 in study areas for the period 2006 ((a), (c), (e), (g), and (i)), and 2014 ((b), (d), (f), (h), and (j)).

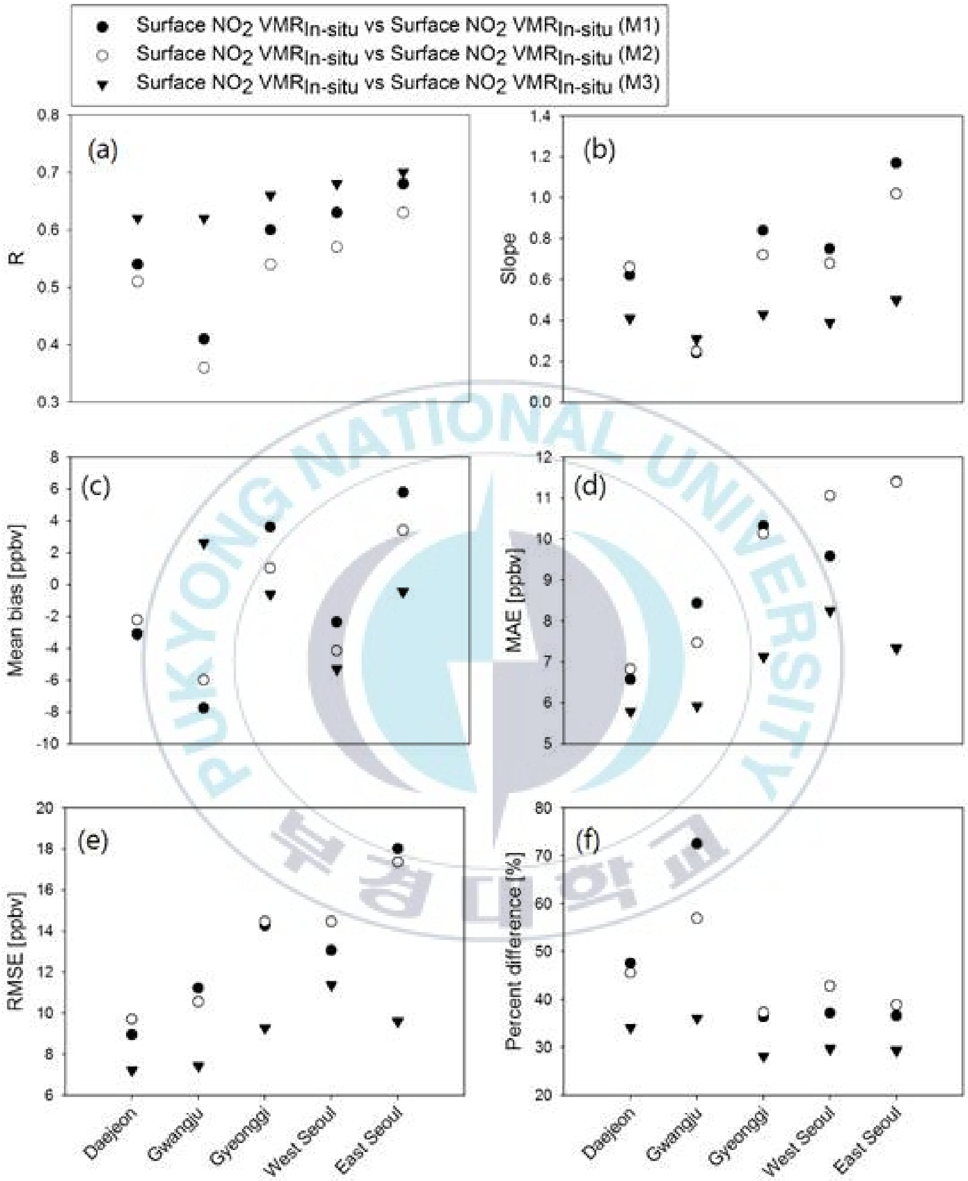


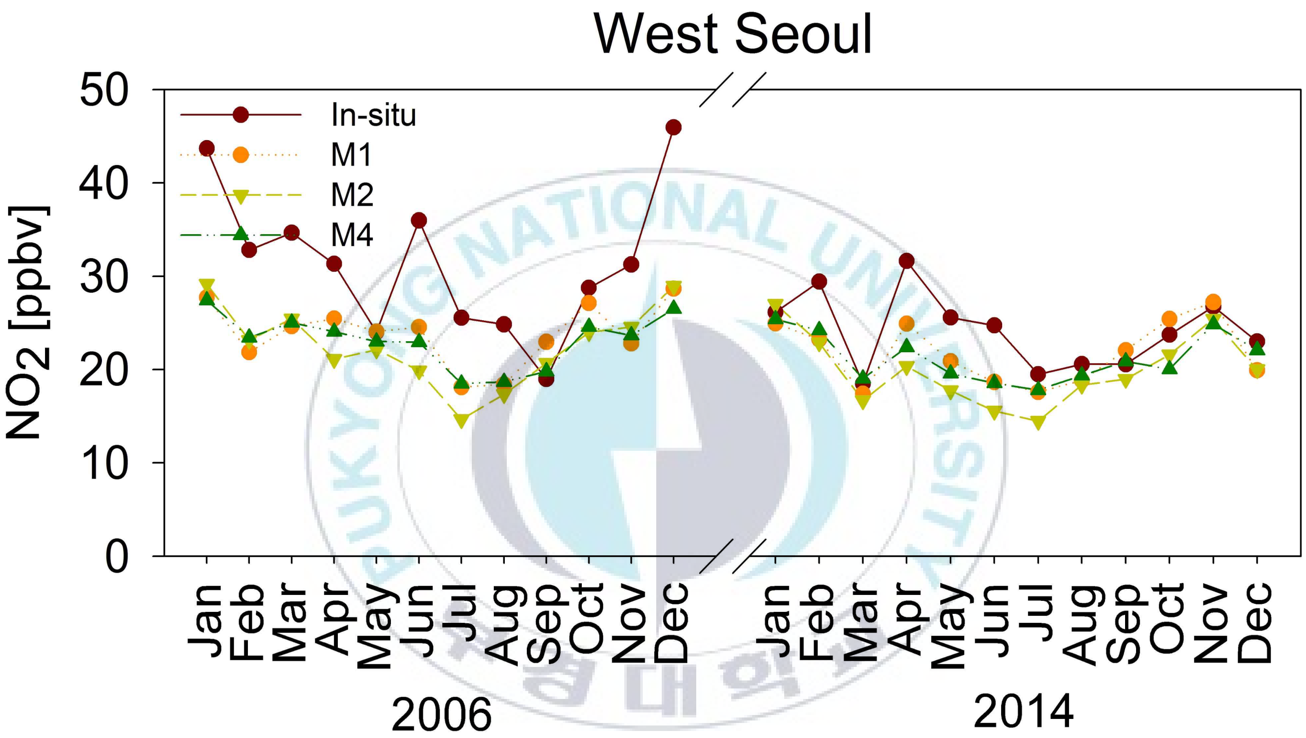
Figure 7. (a) R, (b) slope, (c) MB, (d) MAE, (e) RMSE, and (f) percent difference between NO_2 $VMR_{ST-estimates}$ against NO_2 $VMR_{In-situ}$ in 2006 and 2014.

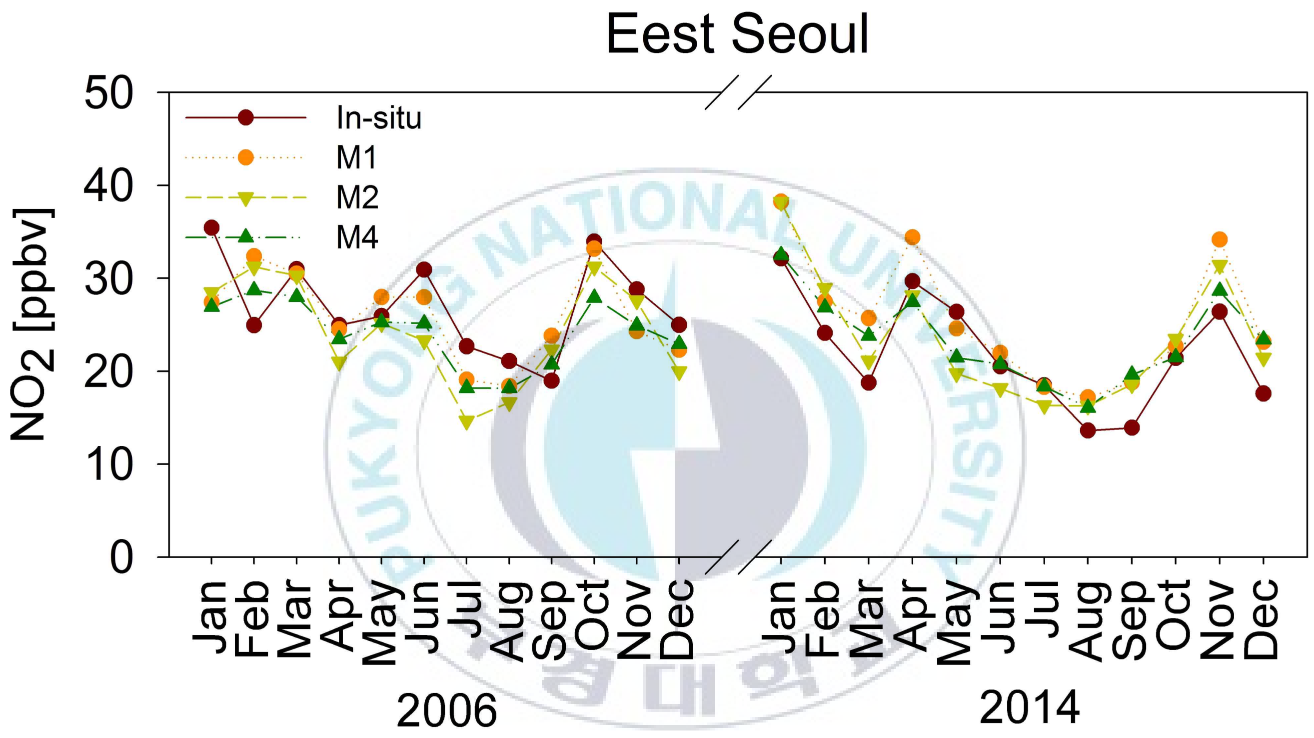
Figure 7 shows the (a) R (correlation coefficient), (b) slope, (c) mean bias (MB), (d) mean absolute error (MAE), (e) root mean square error (RMSE) and (f) percent difference between NO_2 $VMR_{ST\text{-}estimates}$ and NO_2 $VMR_{In\text{-}situ}$ in validation period (2006 and 2014). The R between NO_2 $VMR_{ST\text{-}estimates}$ obtained from M1 and NO_2 $VMR_{In\text{-}situ}$ ranges from 0.49 to 0.71, showing better agreement than that of NO_2 $VMR_{ST\text{-}estimates}$ from M2 ($0.47 < R < 0.65$). M3 showed the best correlation with NO_2 $VMR_{In\text{-}situ}$ ($0.67 < R < 0.90$) among the three methods. The slopes between NO_2 $VMR_{In\text{-}situ}$ and NO_2 $VMR_{ST\text{-}estimates}$ from both M1 and M2 are close 1 in East Seoul, whereas they are lower the other cities. The MB between NO_2 $VMR_{In\text{-}situ}$ and NO_2 $VMR_{ST\text{-}estimates}$ from M1, M2, and M3 range from -7.74 to 5.80 ppbv. In all study areas, MAE (5.79 ppbv < MAE < 8.25 ppbv) of M3 is lower than those (6.58 ppbv < MAE < 11.41 ppbv) of M1 and M2, which means that NO_2 $VMR_{ST\text{-}estimates}$ estimated from M3 show good agreement with NO_2 $VMR_{In\text{-}situ}$ in the magnitude. RMSE between NO_2 $VMR_{In\text{-}situ}$ and NO_2 $VMR_{ST\text{-}estimates}$ from M3 is found to be lower than those between NO_2 $VMR_{In\text{-}situ}$ and NO_2 $VMR_{ST\text{-}estimates}$ from M1 and M2. The NO_2 $VMR_{ST\text{-}estimates}$ by M3 showed the lowest RMSE in all study areas (7.21 ppbv < RMSE < 11.37 ppbv). In addition, percent differences between NO_2 $VMR_{ST\text{-}estimates}$ estimated from M3 and NO_2 $VMR_{In\text{-}situ}$ were lower in all study areas than estimated values by M1 and M2. In estimating NO_2 $VMR_{ST\text{-}estimates}$, M3, which is a multiple regression method with inputs of various independent variables, generally showed good statistical performance except for MB.

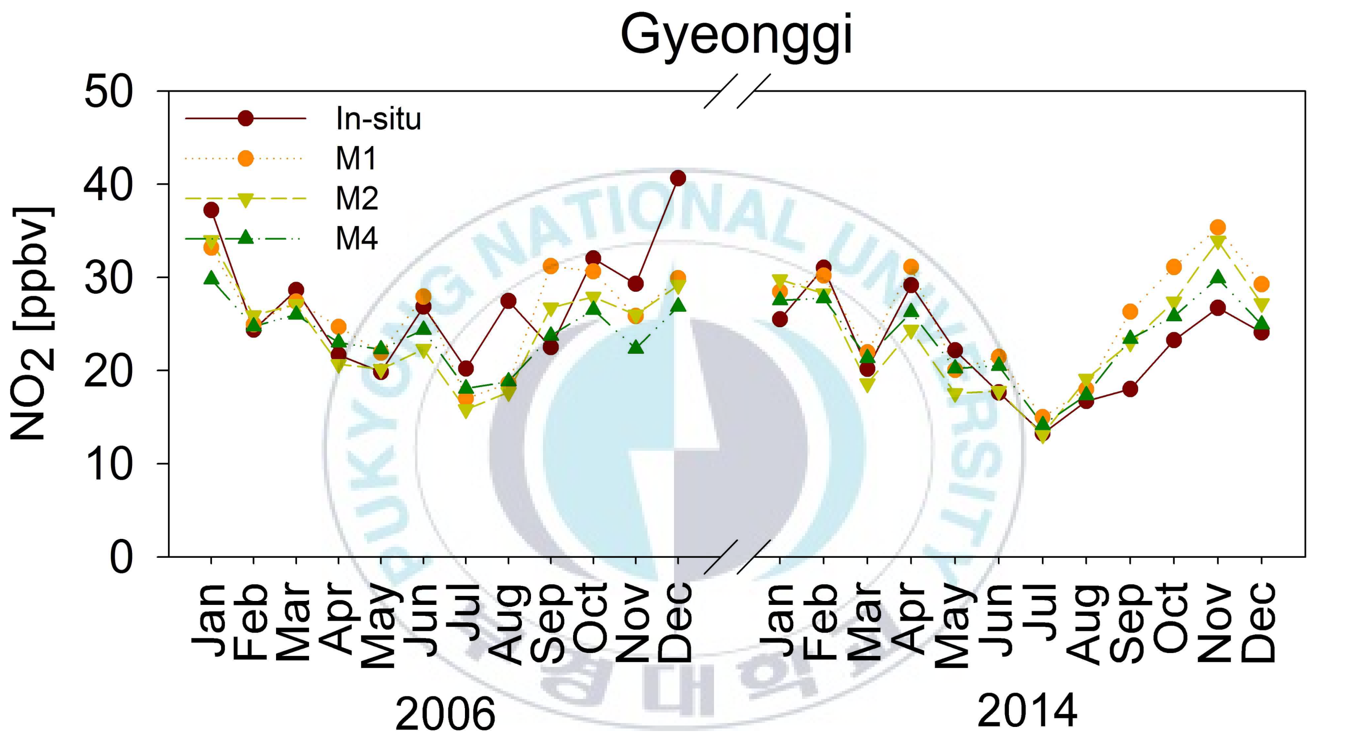
4.2 Monthly estimation

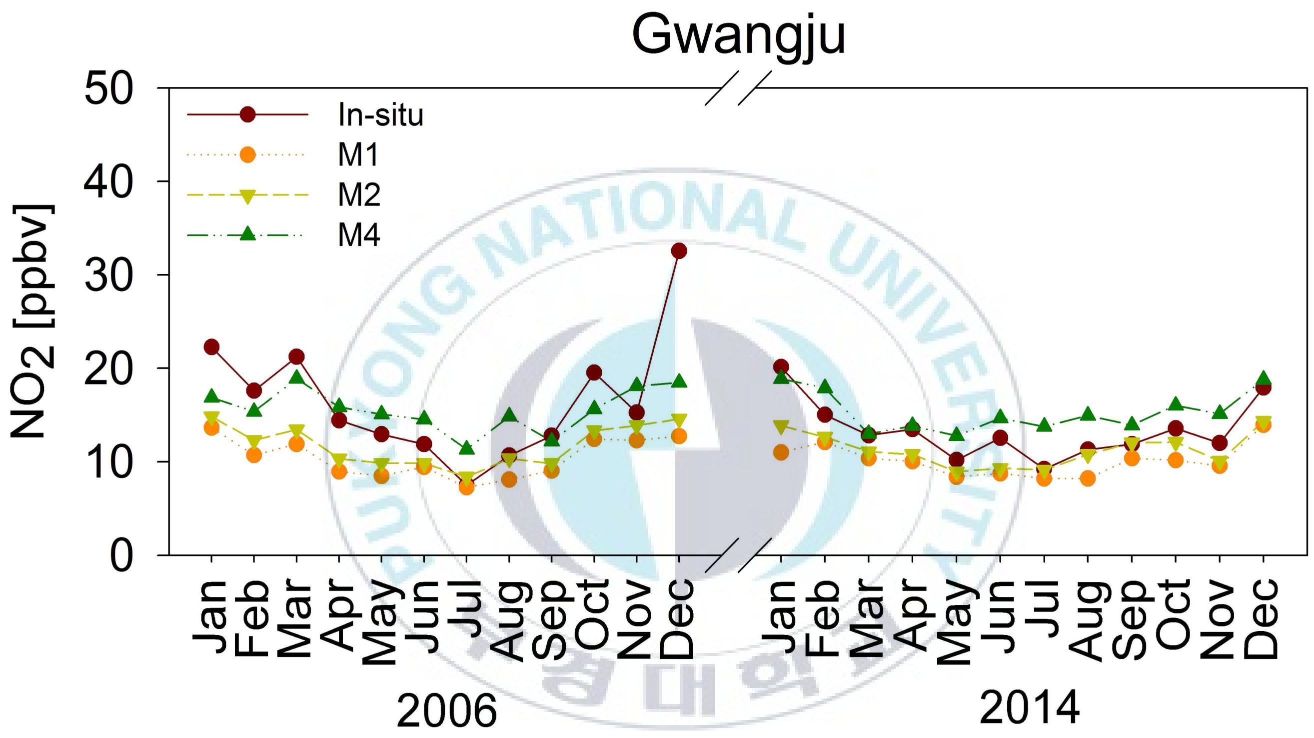
Figure 8 shows temporal variations of NO_2 $VMR_{M\text{-estimates}}$ and monthly mean NO_2 $VMR_{In\text{-situ}}$ estimated using the M1, M2 and M4 of Table 5 in from West Seoul and East Seoul. The NO_2 $VMR_{M\text{-estimates}}$ are obtained from M1, M2 and M4 in Table 5 with the inputs of monthly mean independent variables during the validation period (see the detailed input data in Section 2.1). Figure 8 shows a good agreement between the estimated NO_2 $VMR_{M\text{-estimates}}$ and monthly mean NO_2 $VMR_{In\text{-situ}}$ in the temporal pattern. However, we found a large difference between NO_2 $VMR_{In\text{-situ}}$ and NO_2 $VMR_{M\text{-estimates}}$ in the period when there was a change difference in NO_2 $VMR_{M\text{-estimates}}$ between the previous and the following months. For example, there was no models that calculated NO_2 $VMR_{M\text{-estimates}}$ that were similar to NO_2 $VMR_{In\text{-situ}}$ in December 2006 which changes largely compared to that in November 2006. NO_2 $VMR_{In\text{-situ}}$ (NO_2 $VMR_{M\text{-estimates}}$ from M1, M2, and M4) in November and December in 2006 are 19.32 ppbv (15.94, 17.96, and 17.62 ppbv) and 30.30 ppbv (15.94, 17.96, and 17.62 ppbv) in Daejeon, 15.26 ppbv (12.29, 13.87, and 18.09 ppbv) and 32.55 ppbv (12.73, 14.57, and 18.46 ppbv) in Gwangju, 29.31 ppbv (25.86, 25.97, and 22.35 ppbv) and 40.64 ppbv (29.91, 29.15, and 26.85 ppbv) in Gyeonggi, and 31.25 ppbv (22.80, 24.55, and 23.64 ppbv) and 45.93 ppbv (28.65, 28.92, and 26.49 ppbv) in West Seoul. Especially in West Seoul, There are several periods when NO_2 $VMR_{In\text{-situ}}$ changes rapidly compared to the previous month. The NO_2 $VMR_{M\text{-estimates}}$ obtained from the three models are in poor agreement with NO_2 $VMR_{In\text{-situ}}$ in monthly pattern. As described in Section

2, despite the use of NO_2 $VMR_{In\text{-}situ}$ located away from the streets, the in-situ measurement sites in West Seoul are located closer to the streets than the in-situ measurement sites in Daejeon and Gwangju. For this reason, there thought to be more periods when NO_2 $VMR_{In\text{-}situ}$ changes rapidly compared to the previous month. It is difficult to estimate the rapid change of NO_2 VMR near NO_2 source by regression models that reflect the relationship between the in-situ measurements and the OMI sensor covering both source and non-source areas in a single pixel.


West Seoul
NO2 [ppbv]
50
40
30
20
10
0
In-situ
M1
M2
M4
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
2006
2014


Eest Seoul
NO2 [ppbv]
50
40
30
20
10
0
In-situ
M1
M2
M4
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
2006
2014


Gyeongggi
In-situ
M1
M2
M4
NO_2 [ppbv]
0
10
20
30
40
50
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
2006
2014



Gwangju
50
40
30
20
10
0
NO2 [ppbv]
In-situ
M1
M2
M4
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
2006
2014

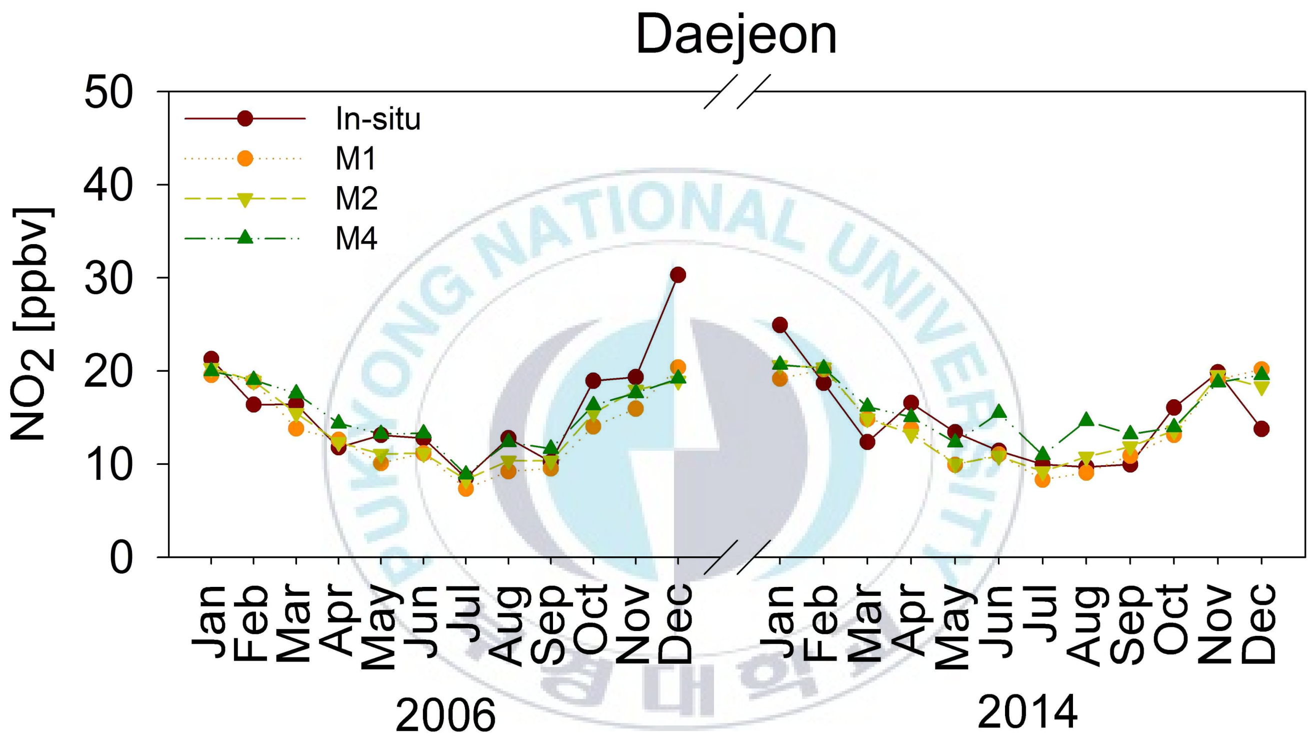


Figure 8. Time series of NO_2 $VMR_{In-situ}$ and NO_2 $VMR_{M-estimates}$ estimates by M1, M2, and M4 for the period 2006 and 2014.

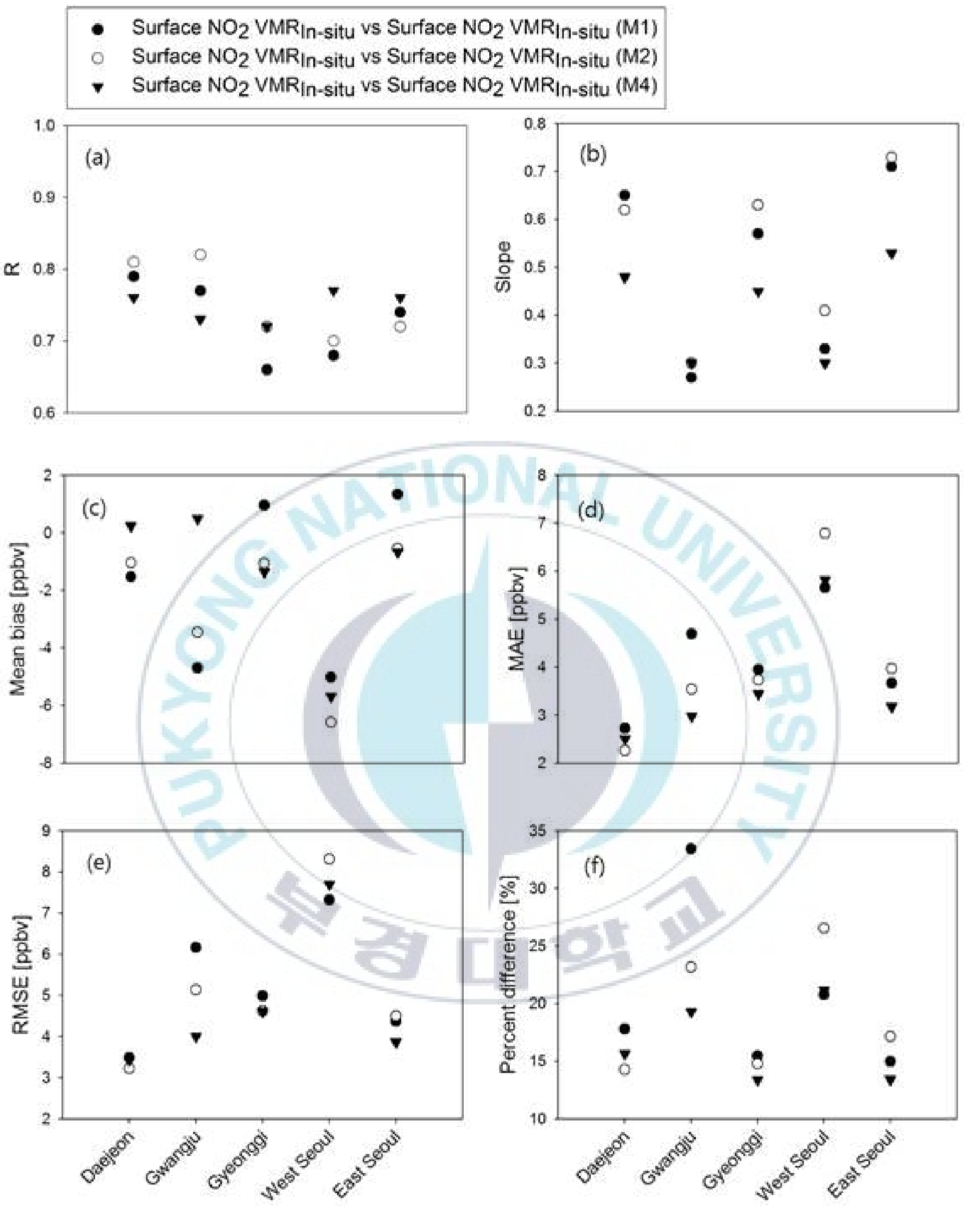


Figure 9. (a) R, (b) slope, (c) MB, (d) MAE, (e) RMSE, and (f) percent difference between NO_2 $VMR_{M\text{-}estimates}$ against monthly mean NO_2 $VMR_{In\text{-}situ}$ in 2006 and 2014.

Figure 9 shows the (a) R, (b) slope, (c) MB, (d) MAE, (e) RMSE and (f) percent difference between NO_2 $VMR_{M\text{-estimates}}$ and monthly mean NO_2 $VMR_{In\text{-situ}}$ in 2006 and 2014. In general, NO_2 $VMR_{M\text{-estimates}}$ showed better agreement with NO_2 $VMR_{In\text{-situ}}$ than the NO_2 $VMR_{ST\text{-estimates}}$. R between NO_2 $VMR_{M\text{-estimates}}$ obtained from M1, M2 and M4 and monthly mean NO_2 $VMR_{In\text{-situ}}$ areas range from 0.68 to 0.82 in all areas. MB was close to 0 in most study areas. MAE was less than 5 ppbv in Daejeon, Gwangju, Gyeonggi, and East Seoul where there are good agreements between NO_2 $VMR_{M\text{-estimates}}$ from M1, M2, and M4 and monthly mean NO_2 $VMR_{In\text{-situ}}$, whereas MAEs in West Seoul ranges from 5.66 to 6.79. RMSEs between NO_2 $VMR_{In\text{-situ}}$ and NO_2 $VMR_{M\text{-estimates}}$ from M1, M2, and M3 are found to be lower than 7 ppbv in study areas except for West seoul. In addition, the three models showed percent difference of less than 30% except for estimated value from M1 in Gwangju.

5. Discussion

In previous study [ORDÓNEZ et al., 2006], tropospheric NO_2 VCDs obtained from GOME were compared with tropospheric NO_2 VCDs calculated using both NO_2 concentrations obtained from in-situ measurements and Model of Ozone and Related Tracers 2 (MOZART-2). There are also several previous studies estimating surface NO_2 VMR using satellite data [Lamsal et al., 2008; Kharol et al., 2015]. Among them, Kharol et al. (2006) estimated the annual variation of ground-level NO_2 concentrations using both GEOS-Chem data and OMI data. However, in this present study, NO_2 $VMR_{ST\text{-}estimates}$ and NO_2 $VMR_{M\text{-}estimates}$ were estimated for the first time in higher temporal resolution using the three regression models with the inputs of Trop NO_2 VCD_{OMI}.

- Estimation of surface NO_2 VMRs of a specific time (13:45LT).
  - ✓ Among the three regression models, the multiple regression model M3, showed the best performance in estimating NO_2 $VMR_{ST\text{-}estimatie}$. In estimation of NO_2 $VMR_{ST\text{-}estimatie}$, the linear regression model (M2), in which BLH is used also as an independent variable in addition to Trop NO_2 VCD_{OMI}, shows performance comparable with that of the model (M1) which uses Trop NO_2 VCD_{OMI} as the only independent variable. The BLH varies with latitude [Zeng et al., 2004]. The BLH is not well reflected by the latitude change since the spatial resolution of the AIRS used in this study is larger than the spatial resolution of OMI. We expected better results by using BLH data obtained from

Lidar or by correcting spatial resolution difference between OMI and AIRS.

- ✓ A slightly large differences in magnitude were found between NO_2 $VMR_{In\text{-}situ}$ and NO_2 $VMR_{ST\text{-}estimates}$ obtained from M1, M2 and M3, while there were good agreements in daily patterns of NO_2 between NO_2 $VMR_{In\text{-}situ}$ and NO_2 $VMR_{ST\text{-}estimates}$ obtained from M1, M2 and M3 (Figure 4).
- ✓ In terms of statistical evaluation via comparison between NO_2 $VMR_{ST\text{-}estimates}$ and NO_2 $VMR_{In\text{-}situ}$, M3 showed in general the best performance.

- Estimation of monthly mean surface NO_2 VMRs of a specific time (13:45LT).
  - ✓ We found a good agreement between the estimated NO_2 $VMR_{M\text{-}estimates}$ and monthly mean NO_2 $VMR_{In\text{-}situ}$ in the temporal pattern (Figure 8). However, it was shown that a large difference between NO_2 $VMR_{In\text{-}situ}$ and NO_2 $VMR_{M\text{-}estimates}$ in the period when there was a change difference in NO_2 $VMR_{M\text{-}estimates}$ between the previous and the following months. Despite the use of NO_2 $VMR_{In\text{-}situ}$ located away from the streets, the in-situ measurement sites in West Seoul are located closer to the streets than the in-situ measurement sites in Daejeon and Gwangju. For this reason, there thought to be more periods when NO_2 $VMR_{In\text{-}situ}$ changes rapidly compared to the previous month. It is difficult to estimate the rapid change of NO_2 VMR near NO_2 source by regression models that reflect the

relationship between the in-situ measurements and the OMI sensor covering both source and non-source areas in a single pixel.

- ✓ In terms of statistical evaluations, performances in estimating NO_2 $VMR_{M\text{-}estimates}$ using three regression models (M1, M2, and M4) were found to be similar (Figure 9).
- ✓ NO_2 $VMR_{M\text{-}estimates}$ shows better agreement with the NO_2 $VMR_{In\text{-}situ}$ than NO_2 $VMR_{ST\text{-}estimates}$. The reason for the better performances in the monthly mean estimation could be attributed to the reduced errors in monthly mean OMI data [OMI Team, 2009] and also fewer occasions with sudden monthly changes in NO_2 $VMR_{In\text{-}situ}$ than rapid daily changes in NO_2 $VMR_{In\text{-}situ}$.

In this study, it is expected that the regression methods used to estimate the surface NO_2 VMR using Trop NO_2 VCD_{OMI} will be useful in providing information on surface NO_2 VMR in metropolitan cites. For future researches, the estimation of surface NO_2 VMR can be attempted in higher time resolution with geostationary satellite sensors (e.g., geostationary environmental monitoring spectrometer (GEMS), tropospheric emissions: monitoring of pollution (TEMPO), and sentinel-4).

6. Conclusions

In this study, NO_2 $VMR_{ST\text{-}estimates}$ and NO_2 $VMR_{M\text{-}estimates}$ were estimated for the first time using three regression models in four metropolitan cities for the two years period 2006 and 2014. Multiple regression model (M3) is found to show the best performance in estimating NO_2 $VMR_{ST\text{-}estimates}$ in all cities. For the surface NO_2 estimates at the specific time (13:45LT), there are generally better R, MAE, RMSE, and percent difference between NO_2 $VMR_{ST\text{-}estimates}$ from M3 and NO_2 $VMR_{In\text{-}situ}$ than those between NO_2 $VMR_{ST\text{-}estimates}$ from other two models (M1 and M2) and NO_2 $VMR_{In\text{-}situ}$. In comparison of the performances between monthly surface NO_2 VMR estimates and those at specific time, agreement between NO_2 $VMR_{In\text{-}situ}$ and NO_2 $VMR_{M\text{-}estimates}$ found to be better than that between NO_2 $VMR_{In\text{-}situ}$ and NO_2 $VMR_{ST\text{-}estimates}$. In estimating NO_2 $VMR_{M\text{-}estimates}$, three regression models (M1, M2, and M4) showed similar performances. In estimating daily surface NO_2 VMR variation and monthly surface NO_2 VMR variation, when surface NO_2 VMR rapidly change, difference between surface NO_2 VMR estimated from all models and NO_2 $VMR_{In\text{-}situ}$ is found to be large. For the future studies, using higher spatial resolution satellites is expected to improve the relationship with in-situ measurements. In addition, independent variables that can estimate the rapid change of surface NO_2 VMR should be investigated.

7. Reference

Ackermann-Liebrich, U., Leuenberger, P., Schwartz, J., Schindler, C., Monn, C., Bolognini, G., Bongard, J.P., Brändli, O., Domenighetti, G., Elsasser, S. and Grize, L., 1997. Lung function and long term exposure to air pollutants in Switzerland. Study on Air Pollution and Lung Diseases in Adults (SAPALDIA) Team. American journal of respiratory and critical care medicine, 155(1), pp.122-129.

Aumann, H.H., Chahine, M.T., Gautier, C., Goldberg, M.D., Kalnay, E., McMillin, L.M., Revercomb, H., Rosenkranz, P.W., Smith, W.L., Staelin, D.H. and Strow, L.L., 2003. AIRS/AMSU/HSB on the Aqua mission: Design, science objectives, data products, and processing systems. IEEE Transactions on Geoscience and Remote Sensing, 41(2), pp.253-264.

Boersma, K.F., Eskes, H.J. and Brinksma, E.J., 2004. Error analysis for tropospheric NO2 retrieval from space. Journal of Geophysical Research: Atmospheres, 109(D4).

Boersma, K.F., Eskes, H.J., Veefkind, J.P., Brinksma, E.J., Van Der A, R.J., Sneep, M., Van Der Oord, G.H.J., Levelt, P.F., Stammes, P., Gleason, J.F. and Bucsela, E.J., 2006. Near-real time retrieval of tropospheric NO 2 from OMI. Atmospheric Chemistry and Physics Discussions, 6(6), pp.12301-12345.

Boersma, K.F., Jacob, D.J., Trainic, M., Rudich, Y., DeSmedt, I., Dirksen, R. and Eskes, H.J., 2009. Validation of urban NO2 concentrations and their diurnal and seasonal variations observed from the SCIAMACHY and OMI sensors using in situ surface measurements in Israeli cities. Atmospheric Chemistry and Physics, 9(12), pp.3867-3879.

Bucsela, E.J., Celarier, E.A., Wenig, M.O., Gleason, J.F., Veefkind, J.P., Boersma, K.F. and Brinksma, E.J., 2006. Algorithm for NO2 vertical column retrieval from the Ozone Monitoring Instrument, IEEE T. Geosci. Remote, 44, 1245–1258.

Chahine, M.T., Pagano, T.S., Aumann, H.H., Atlas, R., Barnet, C., Blaisdell, J., Chen, L., Divakarla, M., Fetzer, E.J., Goldberg, M. and Gautier, C., 2006. AIRS: Improving weather forecasting and providing new data on greenhouse gases. Bulletin of the American Meteorological Society, 87(7), pp.911-926.

Demerjian, K.L., 2000. A review of national monitoring networks in North America. Atmospheric Environment, 34(12), pp.1861-1884.

IPCC, 2007: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M., and Miller, H. L., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 996 pp., 2007.

James Gauderman, W., McConnell, R.O.B., Gilliland, F., London, S., Thomas, D., Avol, E., Vora, H., Berhane, K., Rappaport, E.B., Lurmann, F. and Margolis, H.G., 2000. Association between air pollution and lung function growth in southern California children. American journal of respiratory and critical care medicine, 162(4), pp.1383-1390.

Kutner, M.H., Nachtsheim, C. and Neter, J., 2004. Applied linear regression models. McGraw-Hill/Irwin.

Kharol, S.K., Martin, R.V., Philip, S., Boys, B., Lamsal, L.N., Jerrett, M., Brauer, M., Crouse, D.L., McLinden, C. and Burnett, R.T., 2015. Assessment of the magnitude and recent trends in satellite-derived ground-level nitrogen dioxide over North America. Atmospheric Environment, 118, pp.236-245.

Leue, C., Wenig, M., Wagner, T., Klimm, O., Platt, U. and Jähne, B., 2001. Quantitative analysis of NOx emissions from Global Ozone Monitoring Experiment satellite image sequences. Journal of Geophysical Research: Atmospheres, 106(D6), pp.5493-5505.

Levelt, P.F., Hilsenrath, E., Leppelmeier, G.W., van den Oord, G.H., Bhartia, P.K., Tamminen, J., de Haan, J.F. and Veefkind, J.P., 2006. Science objectives of the ozone monitoring instrument. IEEE Transactions on Geoscience and Remote Sensing, 44(5), pp.1199-1208.

Martin, R.V., Chance, K., Jacob, D.J., Kurosu, T.P., Spurr, R.J., Bucsela, E., Gleason, J.F., Palmer, P.I., Bey, I., Fiore, A.M. and Li, Q., 2002. An improved retrieval of tropospheric nitrogen dioxide from GOME. Journal of Geophysical Research: Atmospheres, 107(D20).

Ordónez, C., Richter, A., Steinbacher, M., Zellweger, C., Nüß, H., Burrows, J.P. and Prévôt, A.S.H., 2006. Comparison of 7 years of satellite‐borne and ground‐based tropospheric NO2 measurements around Milan, Italy. Journal of Geophysical Research: Atmospheres, 111(D5).

Panella, M., Tommasini, V., Binotti, M., Palin, L. and Bona, G., 2000.

Monitoring nitrogen dioxide and its effects on asthmatic patients: Two different strategies compared. Environmental monitoring and assessment, 63(3), pp.447-458.

Richter, A. and Burrows, J.P., 2002. Tropospheric NO2 from GOME measurements. Advances in Space Research, 29(11), pp.1673-1683.

Schindler, C., Ackermann-Liebrich, U., Leuenberger, P., Monn, C., Rapp, R., Bolognini, G., Bongard, J.P., Brändli, O., Domenighetti, G., Karrer, W. and Keller, R., 1998. Associations between Lung Function and Estimated Average Exposure to NO2 in Eight Areas of Switzerland. Epidemiology, 9(4), pp.405-411.

Sellke, T., Bayarri, M.J. and Berger, J.O., 2001. Calibration of ρ values for testing precise null hypotheses. The American Statistician, 55(1), pp.62-71.

Smith, B.J., Nitschke, M., Pilotto, L.S., Ruffin, R.E., Pisaniello, D.L. and Willson, K.J., 2000. Health effects of daily indoor nitrogen dioxide exposure in people with asthma. European Respiratory Journal, 16(5), pp.879-885.

van der A, R.J., Eskes, H.J., Boersma, K.F., Van Noije, T.P.C., Van Roozendael, M., De Smedt, I., Peters, D.H.M.U. and Meijer, E.W., 2008. Trends, seasonal variability and dominant NOx source derived from a ten year record of NO2 measured from space. Journal of Geophysical Research: Atmospheres, 113(D4).

Xue, D. and Yin, J., 2014. Meteorological influence on predicting surface SO2 concentration from satellite remote sensing in Shanghai, China. Environmental monitoring and assessment, 186(5), pp.2895-2906.

Zeng, X., Brunke, M.A., Zhou, M., Fairall, C., Bond, N.A. and Lenschow, D.H., 2004. Marine atmospheric boundary layer height over the eastern Pacific: Data analysis and model evaluation. Journal of Climate, 17(21), pp.4159-4170.