Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Linlin Du Date

Estimating $PM_{2.5}$ Concentration and Evaluating National Ambient Air Quality Standard in Johannesburg Area, South Africa

By

Linlin Du

Master of Public Health

Gangarosa Department of Environmental Health

Yang Liu
Thesis Advisor

Estimating $PM_{2.5}$ Concentration and Evaluating National Ambient Air Quality Standard in Johannesburg Area, South Africa

By

Linlin Du

B.S.
Shandong University
2018

Thesis Advisor: Yang Liu, Ph.D.

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Gangarosa Department of Environmental Health
2020

Abstract

Estimating $PM_{2.5}$ Concentration and Evaluating National Ambient Air Quality Standard in Johannesburg Area, South Africa

By Linlin Du

It is well recognized that fine particle ($PM_{2.5}$) from multiple sources, including fuel combustion, vehicle emission and domestic burning, is strongly associated with a large burden of illness in South Africa, especially respiratory diseases, yet few studies well characterized ambient $PM_{2.5}$ concentration with high spatiotemporal resolution in South Africa. We developed a random forest model to estimate daily $PM_{2.5}$ concentration at 1 km^2 resolution in the Johannesburg Area, combining satellite AOD, meteorological factors and land-use variables, and evaluated the impact of implementation of national air quality standard on $PM_{2.5}$ concentration. Overall cross-validation R^2 was 0.67, indicating a good fit between model estimation and ground measurements. Mean $PM_{2.5}$ for ground measurements was 28.15 $\mu g/m^3$ and mean estimated $PM_{2.5}$ concentration was 28.24 $\mu g/m^3$. MAIAC AOD, total precipitation, winter, population, population, spring, summer, policy, temperature at 2-meter, relative humidity at planetary boundary layer height, and wind speed at planetary boundary layer height were the most important predictors. Estimation from the model has captured the temporal pattern for ground monitoring stations. Mpumalanga had a lower annual $PM_{2.5}$ concentration than Gauteng. The maximum annual $PM_{2.5}$ concentration appeared in the region between Pretoria and Bronkhorstspruit. By comparing $PM_{2.5}$ concentration, we concluded that the implementation of national air quality standards has not achieved the goal of reducing $PM_{2.5}$ concentration.

Estimating $PM_{2.5}$ Concentration and Evaluating National Ambient Air Quality Standard in Johannesburg Area, South Africa

By

Linlin Du

B.S.
Shandong University
2018

Thesis Advisor: Yang Liu, Ph.D.

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Gangarosa Department of Public Health
2020

Contents

1.Introduction **1**

2.Method **4**

2.1 Study Area **4**

2.2 Ground measurement **4**

2.3 Satellite AOD **5**

2.4 Meteorological data **5**

2.5 Land use **6**

2.6 Temporal Variables **6**

2.7 Modelling **7**

2.8 Policy Analysis **7**

3.Results **7**

3.1 Ground Measurements **7**

3.2 Gap Filled MAIAC AOD **8**

3.3 Random Forest Model Performance **9**

3.4 $PM_{2.5}$ Prediction **9**

3.5 $PM_{2.5}$ Concentration Changes in Implementation of New Standard **11**

4.Discussion **11**

5.Conclusion **15**

6.Reference **16**

7.Figures and Tables **19**

1.Introduction

Fine particles ($PM_{2.5}$) refer to atmospheric particulate matters with aerodynamic diameters of 2.5 μm or less, mostly released from various human-made sources, including industrial emission, power generation, vehicle combustion, agriculture incineration, and residential burning (Tucker 2000). It is widely reported that the majority of populations living in developing regions are disproportionately experiencing air quality levels exceeding WHO standard, resulting in estimated 4.2 million premature death worldwide (Cohen, Brauer et al. 2017) (WHO 2016). Numerous epidemiological studies had buttressed exposure to particle pollution is strongly associated adverse acute and chronic health outcomes, not only respiratory diseases but also cardiovascular and neurologic morbidity and mortality and even reproductive effects (Boogaard, Walker et al. 2019) (Liu, Xu et al. 2017) (Hamanaka and Mutlu 2018) (Babadjouni, Hodis et al. 2017). Despite these causes for concern, the current sparse air quality monitoring network is insufficient to quantify fine particle exposure and risk at a local level, particularly in low- and middle-income countries.

The past decade has seen the application of satellite remote sensing products, aerosol optical depth (AOD), providing global coverage and relatively high resolution, in estimating surface $PM_{2.5}$ concentration. AOD measures the extinction of a ray of light at a wavelength as it passes the atmosphere column, and generally, it is positively related to the surface $PM_{2.5}$ concentration (Liu, Koutrakis et al. 2007). Multi-Angle Implementation of Atmospheric Correction (MAIAC) algorithm used time series analysis and a combination of pixel and image-based processing for Moderate Resolution Imaging Spectroradiometer (MODIS) measurement to get a higher spatial resolution (1 km^2) and improve the accuracy

of aerosol retrievals (Lyapustin et al. 2018). Various statistical models have been developed to capture the non-linear relationship between AOD and ground $PM_{2.5}$ measurement and improve the prediction accuracy and robustness, including nested linear mixed-effects model (Ma, Liu et al. 2016) , and geographically weighted regression (GWR) (Ma, Hu et al. 2014). Unlike these parametric models, the random forest model is a non-parametric model, without restrictive assumption for independence and population distribution, could capture non-linear relationships and interaction between the variables, and measure variables importance (Breiman 2001). MAIAC AOD and random forest model have been widely employed in the estimation of ambient $PM_{2.5}$ concentration in China, the United States and Peru. Vu et al. (Vu, Sanchez et al. 2019) developed an advanced $PM_{2.5}$ exposure model in Lima, Peru, from 2011 to 2016, applying MAIAC AOD and random forest method, with the result showing overall cross-validation R^2 for the model was 0.70. Huang et al. (Huang, Xiao et al. 2018) built a random forest $PM_{2.5}$ model for North China Plain with the cross-validation R^2 for the model of 0.88, showing a good fit between MAIAC AOD and ground measurements.

In South Africa, coal is the dominant energy resource, which provided 77% of South Africa's primary energy needs and more than 90% of electricity (Department of Energy 2010). Other human activities, including mobile vehicle emission and biomass burning, could also exacerbate the generation of particulate matters, resulting in significant public health burden in South Africa (Katoto, Byamungu et al. 2019) (Wright et al. 2017). In Recent years, several studies had been conducted in South Africa to assess ambient air quality and its impact on public health. Saucy et al. (Saucy, Roosli et al. 2018) developed an annual land use regression model for outdoor $PM_{2.5}$ concentration in Western Cape

Province, with the R^2 of 0.21, indicating a lower explanatory power. Marais et al. (Marais, Silvern et al. 2019) applied the GEOS-Chem model and estimated pollutant emissions to simulate ambient $PM_{2.5}$ concentration in 2030 and calculated excess deaths in Africa. Moreover, a Benefits Mapping and Analysis Program (BenMAP) model, using data of population, mortality rate and $PM_{2.5}$ concentration, indicated 28,000 premature deaths in South Africa were attributed to current high PM pollution, which caused economy $29.1 billion (Altieri and Keen 2019).

To ensure South Africans could breathe air that is not harmful to public health, as a part of Air Quality Act (2004), the government established 136 ground ambient air quality stations across the country monitoring main air pollutants, like PM, carbon monoxide (CO), nitrogen oxide (NO_x), sulfur dioxide (SO_2), lead (Pb), hydrogen sulfide (H_2S), black carbon and meteorological factors (Gwaze and Mashele 2018). To reduce $PM_{2.5}$ concentration, in 2012 Department of Environmental Affairs issued national ambient air quality standard for $PM_{2.5}$, shown in Table 1, with the daily concentration from 65 $\mu g/m^3$ to 40 $\mu g/m^3$ effective since 2016 and annual concentration from 25 $\mu g/m^3$ to 20 $\mu g/m^3$ executed from 2016 (Department of Environmental Affairs 2012). The Department of Environmental Affairs also issued other regulations incorporating with ambient $PM_{2.5}$ concentration standards, including list of activities which might cause adverse effect to the environment and public health, as well as maximum emission standard for these activities (Department of Environmental Affairs 2013). In other countries, like China, multiple models were built to estimate $PM_{2.5}$ concentration change to assess the impact of air pollution control policies on air quality, like GEOS-Chem model(Cai, Ma et al. 2018), WRF-CMAQ model(Cai, Wang et al. 2017).

To date, there are few studies well characterizing ambient $PM_{2.5}$ concentration with high spatiotemporal resolution in South Africa. In this study, we built a 1 km^2 spatial resolution daily $PM_{2.5}$ concentration model in the Johannesburg area from the year 2014 to 2018, based on satellite AOD, meteorological fields, and land use variables, and assessed $PM_{2.5}$ concentration change in the implementation of the new air quality standard. Also, the retrieved $PM_{2.5}$ concentration might be applied in future epidemiologic studies to analyze its impact on health burden, respiratory and cardiovascular diseases.

2.Method

2.1 Study Area

Our study area is in the northeast of South Africa, covering approximately 84000 km^2, which includes Gauteng province and part of Mpumalanga province, as Figure 1 shows. Gauteng province, with a population of approximately 15 million people, contains Johannesburg, the country's largest city, and Pretoria, its administrative capital. Mpumalanga covers several distinct physiographic: west plateau, east forested mountains and plain, which leads to diverse climate types, allowing for 68% of the Mpumalanga area used to agricultural activities.

2.2 Ground measurement

There were 21 ground monitoring stations included in our study, shown in Table 2. Hourly $PM_{2.5}$ data were downloaded from South Africa Air Quality Information System (SAAQIS) and underwent quality control, with the negative value and repeating value (more than 3 consecutive identical value) removed. To improve the data completeness and make it better representative, the daily aggregation was only conducted when more than

75% of data is available for the averaging period.

2.3 Satellite AOD

MAIAC AOD at 550 nm from Terra (transiting at 10:30 am local time) and Aqua (transiting at 1:30 pm local time) satellites were downloaded. To improve the coverage of AOD, we developed a customized approach to combine Aqua and Terra observations. First, we conducted a univariate linear regression analysis between Aqua AOD and Terra AOD separated by season and used the estimated coefficients, shown in Table 3, to gap-fill the missing Aqua AOD for those grids with only Terra AOD and vice versa. Second, Aerosol Robotic Network (AERONET) L2 measurements, which is the quality assured ground-based remote sensing aerosol network (Giles), were used to validate the gap-filled AOD observations. The AERONET AOD at 550 nm within 30 minutes of MAIAC measurement was computed based on AOD at 440 nm and Angstrom exponent (α)of wavelength range 440-675 nm, as Equation 1 shows. We developed a linear mixed-effect model, including season-specific random effect, between the AERONET site AOD and matched pixel AOD and used the resulting coefficients, displayed in Table 4, to correct gap-filled AOD data. Finally, the mean of validated Aqua and Terra AOD was calculated and used as the parameter in the $PM_{2.5}$ model.

$$AOD_{550nm} = AOD_{440nm} * \left(\frac{550}{440}\right)^{-\alpha_{(440nm-675nm)}} \qquad \text{Equation (1)}$$

2.4 Meteorological data

Hourly meteorological data, including surface albedo, surface incident shortwave flux, evaporation from turbulence, total cloud fraction, total precipitation, wind speed, wind direction, planetary boundary layer (PBL) height, temperature, humidity and surface

pressure, with a spatial resolution of 0.25° latitude × 0.3125° longitude, were obtained from Goddard Earth Observing System Data Assimilation System GEOS-5 Forward Processing (GEOS-5 FP). All data were converted to a daily level and interpolated to match with the MAIAC grid using the inverse distance weighting method.

2.5 Land use

The European Space Agency Climate Change Initiative Land Cover (ESA CCI-LC) provided a global land cover map at 300m resolution. Percent for different types of materials on the earth's surface, such as agriculture, forest, grassland, wetland, and settlement, was calculated through reclassifying the value and resampling to the MAIAC grid. The Gridded Population of the World (GPW) collection version 4 estimated the worldwide population counts and densities for 2000, 2005, 2010, 2015, and 2020, at the resolution of 30 arc-seconds (CIESIN 2016). Yearly population counts were retrieved by linear interpolation and matched to $1km^2$ pixel. 30-meter elevation data were extracted from Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Global Digital Elevation Model Version 3 (GDEM 003) and were aggregated to the MAIAC grid level. The total main road and railway length within each pixel were calculated by ArcGIS.

2.6 Temporal Variables

Season variables, including summer (Dec. – Feb.), winter (Jun.- Aug.) and spring (Sep. - Nov.) were introduced in this analysis. South Africa is in the southern hemisphere, so its seasonality is opposite to the northern hemisphere. Since the national standard for ambient air $PM_{2.5}$ concentration was changed on January 1^{st}, 2016, we classified the year to a dichotomous variable, before policy (2014-2015) and after policy (2016-2018).

2.7 Modelling

Random forest model constructs a high number of decision trees (n_{tree}) at training time and randomly selects several variables from the dataset to form each decision split for the individual tree (m_{try}), which was used in estimating surface 1 km^2 daily $PM_{2.5}$ concentration. The response variable was daily mean $PM_{2.5}$ concentration for each station. The independent variables included gap filled daily MAIAC AOD, meteorological factors, percent of land use type, population, elevation, road length and a dummy variable for season and policy. 10-fold cross-validation was carried out to evaluate the performance of the random forest model, which randomly divided the dataset into 10 segments with nine used for training and one used for predicting and repeated this process 10 times to get the equal number of observations to the original dataset.

2.8 Policy Analysis

$PM_{2.5}$ estimation from the random forest model, divided into before and after the implementation of the new standard by January 1st, 2016, was applied to evaluate the implementation of the national ambient air quality standard for $PM_{2.5}$. The percentage of the study area with a $PM_{2.5}$ concentration meeting nation standard was calculated. Besides, the difference in $PM_{2.5}$ concentration between two time periods was computed and plotted.

3.Results

3.1 Ground Measurements

A total of 14927 daily average ground-based $PM_{2.5}$ measurements were included in our study, which started on January 1st, 2014 and ended on December 31st, 2018, showing a right-skewed distribution. Mean and standard deviation of daily $PM_{2.5}$ for all stations was

25.27 μg/m^3 and 20.35 μg/m^3 separately, with the range of 0.29 – 263.88 μg/m^3. The minimum station average $PM_{2.5}$ value occurred in Middelburg, 16.17 μg/m^3 while Olivenhoutbosch had the maximum station average $PM_{2.5}$, 88.42 μg/m^3. Ten daily $PM_{2.5}$ observations exceeded 200 μg/m^3 in this study, with nine happening in Olivenhoutbosch, in June 2017 and July 2017 and one in Sharpeville, in January 2017.

The time series of monthly averaged $PM_{2.5}$ concentration for each ground station was shown in Figure 3. Most stations had the same temporal pattern, which increased from January, with the mean value of 17.34 μg/m^3, reached the peak value of 37.94 μg/m^3 in June, the winter for the southern hemisphere, and then decreased to 15.27 μg/m^3 at December. The majority of monthly aggregated $PM_{2.5}$ measurements were less than 50 μg/m^3, while some stations had winter $PM_{2.5}$ observations approaching or exceeding 100 μg/m^3, such as Olivenhoutbosch in 2017 and Xanadu in 2015.

3.2 Gap Filled MAIAC AOD

The coefficient for the univariate linear model between Aqua and Terra AOD, separated by season, was presented in Table 3 and Table 4 displayed season-specific coefficients for the mixed-effect model between AERONET and satellite AOD. Except for summer, the univariate linear model and mixed-effect model can explain about 60% variance of the dependent variable, illustrating good performance for these models.

As is shown in Figure 2, the gap-filling method had increase coverage of MAIAC AOD during our study period, from 48% and 59%, average coverage for Aqua and Terra AOD, to 67%, mean coverage for gap filled MAIAC AOD. Apart from the water area, the coverage of gap filled MAIAC AOD was more than 60%.

Mean and standard deviation of gap filled AOD for all station-day was 0.153 and 0.07 $\mu g/m^3$ separately, with the range of 0.041 – 0.7. The max monthly AOD was 0.20, which occurred in September.

3.3 Random Forest Model Performance

After matching all variables to the 1 km^2 fixed pixel, our final training dataset had 10340 station-day observations and 29 variables. The R-square (R^2) and Root Mean Square Error (RMSE) for the cross-validation were 0.67 and 12.41 $\mu g/m^3$ respectively, indicating the random forest model is reliable and there is a good fit between estimated $PM_{2.5}$ value and ground-based $PM_{2.5}$ measurement. The slope and intercept of the univariate linear regression between $PM_{2.5}$ measurement and estimation were 1.06 and -1.94 $\mu g/m^3$ separately, demonstrated the random forest model might overestimate some low $PM_{2.5}$ concentration and underestimate some high $PM_{2.5}$ value, especially when $PM_{2.5}$ concentration exceeds 150 $\mu g/m^3$.

The importance rank of random forest model predictors, measuring parameters' predictive power, was provided in Figure 5. As the plot suggested, gap filled MAIAC AOD was the most important variable, followed by total precipitation, winter, population, spring, summer, policy, temperature at 2-meter, relative humidity at planetary boundary layer height, and wind speed at planetary boundary layer height.

3.4 $PM_{2.5}$ Prediction

Figure 6 exhibited the time series plot for monthly $PM_{2.5}$ ground measurement and estimation from the random forest model of each station. Our model captured main temporal trends, that is, peak in winter, but it tended to underestimate some high value and overestimate low observation, which appears in many stations. Xanadu ground

measurements had a similar yearly temporal pattern as other stations and a decrease of $PM_{2.5}$ concentration from 2014-2015 to 2016-2018, which also shown in our model estimation. However, our model underestimated the peak value, especially in June 2015, with the difference of more than 25 $\mu g/m^3$ between actual and estimated concentrations. Our model performed better in Middelburg in 2014-2016, with more estimation matching with ground measurement or only a slight difference. While, in 2017-2018, our model misestimated some $PM_{2.5}$ value, with 4 difference larger than 5 $\mu g/m^3$. In Diepkloof, the monthly $PM_{2.5}$ ground measurements had high value in winter and low observations in summer. However, compared to other stations, it had a larger difference between adjacent months. The estimation only matched with the ground monitor on this trend in 2015. The largest difference between observed and estimated concentrations occurred in February 2017 and March 2018, which were more than 10 $\mu g/m^3$. In Three Rivers, our model indicated similar trends for $PM_{2.5}$ ground measurement and estimation, with the most month the same or little difference, except October 2014, February and April 2017. In Ermelo, there is a good fit between observed $PM_{2.5}$ concentration and estimated value in most months. The largest underestimation happened in June 2018, with a value of approximately 20 $\mu g/m^3$.

Figure 7 showed the estimated annual mean $PM_{2.5}$ concentration across the study domain, ranging from 15.97 $\mu g/m^3$ to 97.62 $\mu g/m^3$, with a similar spatial pattern from 2014 to 2018. Mpumalanga had a lower $PM_{2.5}$ concentration than Gauteng, mostly less than 25 $\mu g/m^3$. In other parts of our study zone, except for the stripe Magaliesberg Mountain area centered at Pretoria, a large amount of area had $PM_{2.5}$ value not exceeding 50 $\mu g/m^3$. The maximum annual $PM_{2.5}$ concentration appeared in the region between Pretoria and Bronkhorstspruit,

with a value of more than 80 μg/m^3.

3.5 $PM_{2.5}$ Concentration Changes in Implementation of New Standard

Table 5 shows the percentage of study area meeting annual $PM_{2.5}$ concentration standards. Before the implementation of new policy (2014 – 2015), 42% of the study region met the national air quality standard at that time, with the annual $PM_{2.5}$ concentration less than 25 μg/m^3. After the new $PM_{2.5}$ concentration criteria execution (2016 - 2018), only 14% of the study area complied current level, which is 20 μg/m^3 yearly.

Figure 8 displayed the difference in $PM_{2.5}$ concentration due to the change in national air quality standard. The most eastern and southern parts didn't have a significant change in $PM_{2.5}$ concentration. There was a more than 0.5 μg/m^3 decrease in $PM_{2.5}$ concentration occurring in the northwestern part of the study domain after the new standard. However, since the new rule came into force, we observed a cluster of increase of more than 2 μg/m^3 in $PM_{2.5}$ concentration existing in the middle of Gauteng and west of Mpumalanga, especially in Embalenhle and Secunda, with an increase of more than 5 μg/m^3.

4.Discussion

It is well recognized that fine particle from multiple sources, including fuel combustion, vehicle emission and domestic burning, is associated with a large burden of illness in South Africa, especially respiratory diseases. To assess and prevent the current air pollution, the South Africa government issued Air Quality Act in 2004 and put in place various measures in the past 15 years, such as developing ground air quality monitoring station network. However, these monitoring stations mainly located in the urban or industrial area with a high density of people, which is insufficient to assess fine particle exposure and risks at a

local level. Previous researchers have conducted several modeling methods, like land-use regression model, and GEOS-Chem model, to estimate $PM_{2.5}$ concentration and their impact on public health. Nevertheless, these studies have coarse spatial resolution and low prediction power, which is limited to provide $PM_{2.5}$ measurement at fine spatiotemporal resolution for epidemiological research. Therefore, our model for estimating daily surface $PM_{2.5}$ concentration at 1 km^2 resolution in the Johannesburg Area, which implemented MAIAC AOD measurement and random forest method, is the first advanced model in Johannesburg area, improving the accuracy and robustness for the model as well as the coverage and resolution for $PM_{2.5}$ measurement.

Estimation from this model has captured the temporal pattern for each ground monitoring station. Typically, maximum $PM_{2.5}$ concentration occurred in May to August, the wintertime for the southern hemisphere, and minimum concentration was observed from November to February, the local summer. This trend can be attributed to the increasing amount of coal consumption for heating and electricity, due to a lower temperature in the winter season. However, for monthly $PM_{2.5}$ concentration, our model might misestimate some value. The variable with the highest prediction power in our model is the gap-filled MAIAC AOD, whose temporal distribution did not correspond with $PM_{2.5}$ concentration in South Africa. $PM_{2.5}$ concentration peaked in winter, mostly June and July, while maximum AOD value occurred in late winter and early spring, that is, August and September, which can be observed on both satellite AOD and AERONET AOD (Adesina, Piketh et al. 2017) (Hersey, Garland et al. 2015). AOD is a quantitative measurement for aerosol presenting in the atmosphere column. However, the composition of the aerosol column is not homogeneous vertically and only aerosol near ground could be inhaled and

cause adverse health effects. In winter and early spring, biomass burning in the tropical area could be transported to South Africa to the upper air. However, it would not affect surface atmosphere aerosol, which mostly comes from domestic pollution. This transportation causes an increase in aerosol, as well as AOD, but not in surface $PM_{2.5}$ concentration (Hersey, Garland et al. 2015) (Tyson et al., 1996). Another evidence for aerosol column inhomogeneity is the Goddard Ozone Chemistry Aerosol Radiation and Transport (GOCART) model, which estimated the composition of total column AOD. The maximum proportion of organic carbon in total column AOD was observed in August and September, consistent with the onset of biomass burning in the tropics. Therefore, different vertical composition of the aerosol column in different time causes the discrepancy of $PM_{2.5}$ and AOD and we introduced season variable to the model to increase the accuracy of prediction. However, lack of predictors characterizing aerosol property affects the prediction capability of the model.

Annual $PM_{2.5}$ estimation from our model shows the spatial pattern for fine particle pollution. Yearly average $PM_{2.5}$ concentration eliminates the influence of time-dependent variables such as meteorological factors and MAIAC AOD, mainly reflecting the difference in elevation, land use type and population density in fixed grid. The eastern part of the study domain, in other words, Mpumalanga Province, has the lowest $PM_{2.5}$ annual concentration. Since the Mpumalanga province was divided into west high-altitude grassland and east mountain, most of the land is used for agriculture, which means little $PM_{2.5}$ sources, resulting in lower $PM_{2.5}$ concentration. However, in Pretoria and its east, max $PM_{2.5}$ concentration appears, which is mostly driven by elevation. Pretoria, as administrational capital, located in a valley surrounded by Magaliesberg Mountain, has a

large density of population, well-developed road network and manufacturing industry, which means prodigious $PM_{2.5}$ sources. Due to the unique geographic location, particulate matters, generated from Pretoria and other industrial regions, quickly spreads around the surroundings, then was blocked by the mountain, causing accumulation of fine particles in this valley, forming $PM_{2.5}$ pollution hotspot.

The difference in $PM_{2.5}$ concentration, calculated from both ground monitor and our model, shows a constant trend. There was a reduction in $PM_{2.5}$ concentration observed in the northwestern part of the study area, which could be attributed to the execution of the Air Quality Act 2012 Framework. The Department of Environmental Affairs set the new standard for fine particle, listed some manufacturing activities might exert adverse effects on human health and environment, and enact maximum emission standard for these listed activities. However, due to the lack of effective supervision and penalty, $PM_{2.5}$ concentration in most study regions didn't have a significant change. Furthermore, in Embalenhle and Secunda, there is an evident increase in $PM_{2.5}$ concentration. Secunda has Sasol South Africa Ltd manufacturing operations complex, which undertaking coal mining and synthesis of related chemicals. In 2016, Secunda Synfuels Operation also submitted application for postponement of compliance timeframes for the Air Quality Act Maximum Emission Standard and gain approval in 2018 (Sasol Ltd 2016) (Department of Environmental Affairs 2018). Therefore, $PM_{2.5}$ concentration in some areas has increased since the implementation of new ambient air quality standards.

The limitation of this study is the lack of ground $PM_{2.5}$ observations for some stations. In our training dataset, half of the ground stations only had $PM_{2.5}$ concentration data available for less than half of study time, which would influence prediction ability across

the study period. For those stations only had $PM_{2.5}$ data for one or two years, they would need to borrow prediction capability from other stations to estimate $PM_{2.5}$ concentration for the entire study time. In addition, there is non-random missing value for AOD after the gap fill procedure, which might affect the performance for random forest, causing misestimation of $PM_{2.5}$ concentration. The temporal and spatial variation in $PM_{2.5}$, as well as the difference for new air quality standard, might be related to the non-random missing.

5.Conclusion

Our model is the first advanced model estimating daily $PM_{2.5}$ concentration at fine spatial resolution in the Johannesburg area, which combined satellite AOD, meteorological factors as well as land-use variables, and increased temporal and spatial coverage for $PM_{2.5}$ observations. Estimated $PM_{2.5}$ concentration from this model could be applied to future epidemiologic health study as $PM_{2.5}$ exposure data. By comparing $PM_{2.5}$ concentration before and after the execution of new ambient air quality, we concluded that the strategic objectives for national ambient air quality standards have yet to be met. Since this study use the simple linear regression and mixed-effect model to gap fill MAIAC AOD, in the future, to further improve prediction capability, more accurate and robust model would be developed. Besides, more attention should be paid to $PM_{2.5}$ composition and its adverse effects on health and environment in future study.

6.Reference

Adesina, A. J., et al. (2017). "Characteristics of columnar aerosol optical and microphysical properties retrieved from the sun photometer and its impact on radiative forcing over Skukuza (South Africa) during 1999-2010." Environ Sci Pollut Res Int 24(19): 16160-16171.

Altieri, K. E. and S. L. Keen (2019). "Public health benefits of reducing exposure to ambient fine particulate matter in South Africa." Sci Total Environ 684: 610-620.

Babadjouni, R. M., et al. (2017). "Clinical effects of air pollution on the central nervous system; a review." J Clin Neurosci 43: 16-24.

Boogaard, H., et al. (2019). "Air pollution: the emergence of a major global health risk factor." Int Health 11(6): 417-421.

Breiman, L. (2001). "Random Forests". Machine Learning 45: 5-32.

Cai, S., et al. (2018). "Impact of air pollution control policies on future PM2.5 concentrations and their source contributions in China." J Environ Manage 227: 124-133.

Cai, S., et al. (2017). "The impact of the "Air Pollution Prevention and Control Action Plan" on PM2.5 concentrations in Jing-Jin-Ji region during 2012-2020." Sci Total Environ 580: 197-209.

Center for International Earth Science Information Network - CIESIN - Colubia University. 2016. Gridded Population of the World, Version 4 (GPWv4): Administrative Unit Center Points with Population Estimates.

Cohen, A. J., et al. (2017). "Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015." The Lancet 389(10082): 1907-1918.

Department of Energy (2010). South Africa National Energy Balance (2000-2010).

Department of Environmental Affairs (2012). National Ambient Air Quality Standard for Particulate Matter with Aerodynamic Diameter less than 2.5 Micro Meters (PM2.5).

Department of Environmental Affairs (2013). List of Activities Which Result in Atmosphere Emissions Which Have or May Have A Significant Detrimental Effect on the Environment, Including Health, Social Conditions, Economic Conditios, Eological Conditions or Cultural Heritage.

Department of Environemtnal Affairs (2018). Ref: FS/SSO-FDDM/20170419.

Giles, D.M. AERONET: AEROSOL ROBOTIC NETWORK. Available online: https://aeronet.gsfc.nasa.gov/.

Gwaze, P. and S. H. Mashele (2018). "South African Air Quality Information System (SAAQIS) mobile application tool: bringing real time state of air quality to South Africans." Clean Air Journal 28(1): 3-4.

Hamanaka, R. B. and G. M. Mutlu (2018). "Particulate Matter Air Pollution: Effects on the Cardiovascular System." Front Endocrinol (Lausanne) 9: 680.

Hersey, S. P., et al. (2015). "An overview of regional and local characteristics of aerosols in South Africa using satellite, ground, and modeling data." Atmos Chem Phys 15: 4259-4278.

Huang, K., et al. (2018). "Predicting monthly high-resolution PM2.5 concentrations with random forest model in the North China Plain." Environ Pollut 242(Pt A): 675-683.

Katoto, P., et al. (2019). "Ambient air pollution and health in Sub-Saharan Africa: Current evidence, perspectives and a call to action." Environ Res 173: 174-188.

Liu, Q., et al. (2017). "Effect of exposure to ambient PM2.5 pollution on the risk of respiratory tract diseases: a meta-analysis of cohort studies." J Biomed Res 31(2): 130-142.

Liu, Y., et al. (2007). "Estimating fine particulate matter component concentrations and size distributions using satellite-retrieved fractional aerosol optical depth: part 1--method development." J Air Waste Manag Assoc 57(11): 1351-1359.

Lyapustin, A, et al. (2018). MODIS Collection 6 MAIAC Algorithm. ATMOSPHERIC MEASUREMENT TECHNIQUES, 11(10), 5741-5765.

Ma, Z., et al. (2014). "Estimating ground-level PM2.5 in China using satellite remote sensing." Environ Sci Technol 48(13): 7436-7444.

Ma, Z., et al. (2016). "Satellite-derived high resolution PM2.5 concentrations in Yangtze River Delta Region of China using improved linear mixed effects model." Atmospheric Environment 133: 156-164.

Marais, E. A., et al. (2019). "Air Quality and Health Impact of Future Fossil Fuel Use for Electricity Generation and Transport in Africa." Environ Sci Technol 53(22): 13524-13534.

Sasol South Africa Ltd (2016). Applications for Postponement of Certain Requirements of National Environmental Management: Air Quality Act - Minimum Emission Standards, for Sasol South Africa (Pty) Ltd Operations in Secunda.

Saucy, A., et al. (2018). "Land Use Regression Modelling of Outdoor NO(2) and PM2.5 Concentrations in Three Low Income Areas in the Western Cape Province, South Africa." Int J Environ Res Public Health 15(7).

Tucker, W. (2000). An Overview of PM2.5 Sources and Control Strategies. Fuel Processing Technology, 65-66, 379–392.

Tyson, P.D., et al. (1996) An Air Transport Climatology for Subtropical Southern Africa. Int J Climatol. 16:265–291.

Vu, B. N., et al. (2019). "Developing an Advanced PM2.5 Exposure Model in Lima, Peru." Remote Sens (Basel) 11(6).

WHO (World Health Organization). Ambient Air Pollution: a global assessment of

exposure and burden of disease. 2016.

Wright, C., et al. (2017). Air Quality and Human Health Impacts in Southern Africa.

7.Figures and Tables

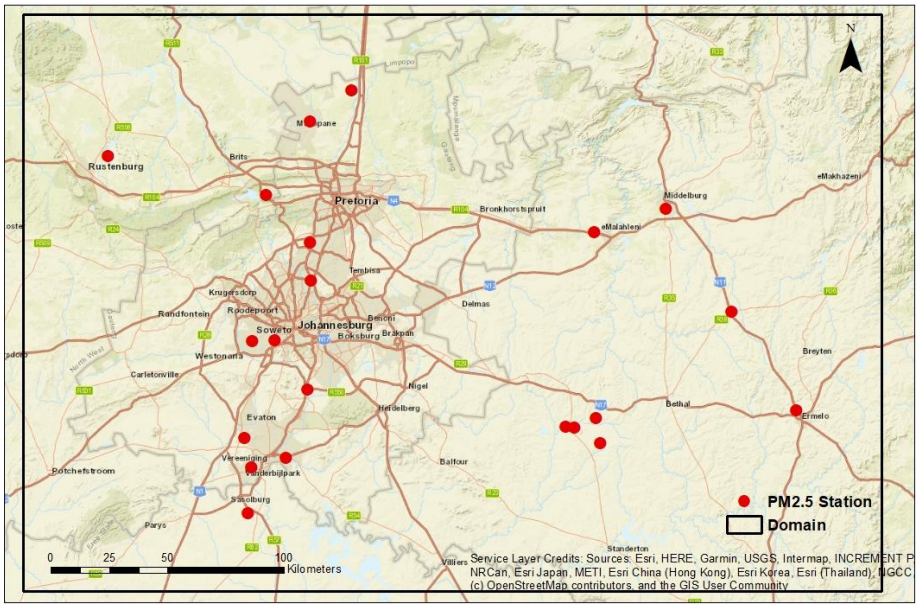

Figure 1. Study Domain and Ground Monitoring Stations

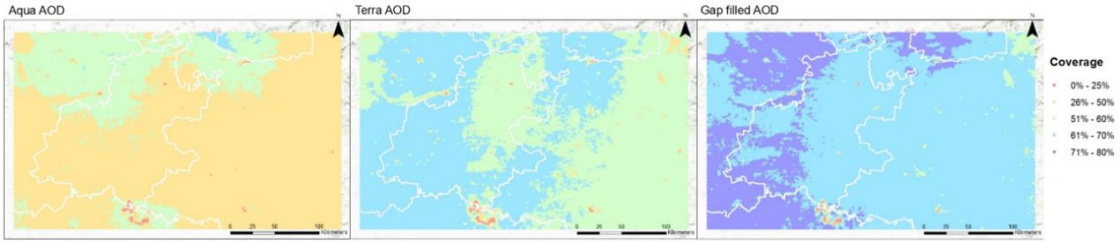

Figure 2. Coverage of AOD

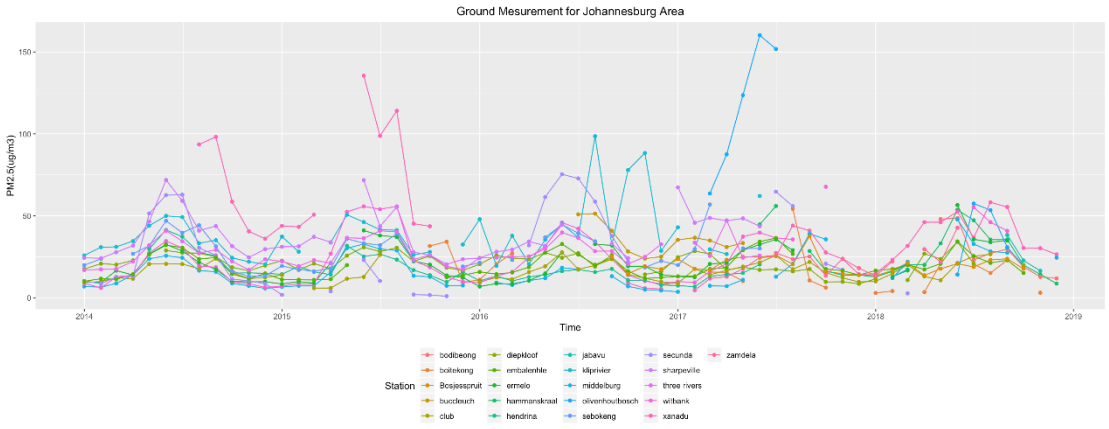

Figure 3. Time Series Plot for Monthly $PM_{2.5}$ Ground Measurements

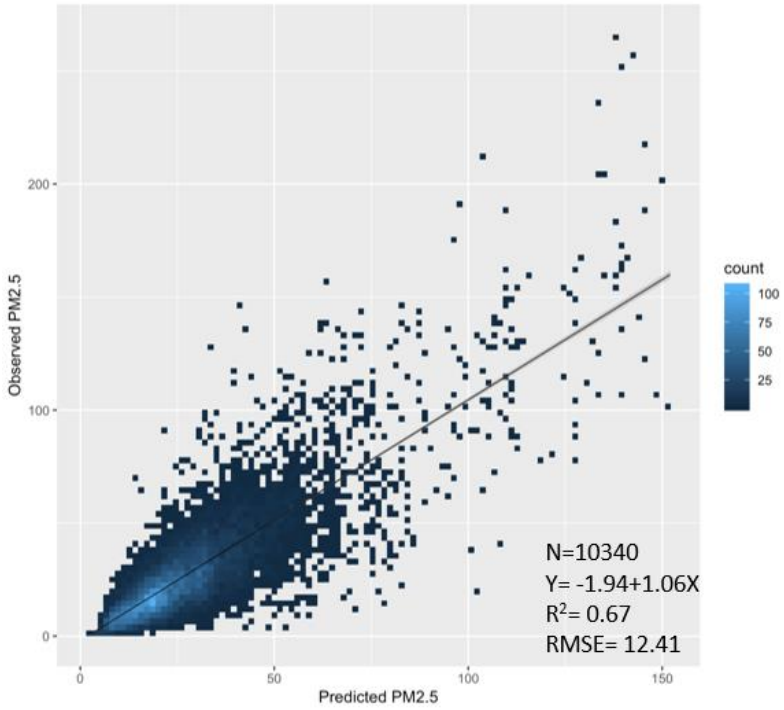

Figure 4. Scatter Plot for 10-fold Cross-validation of Daily $PM_{2.5}$

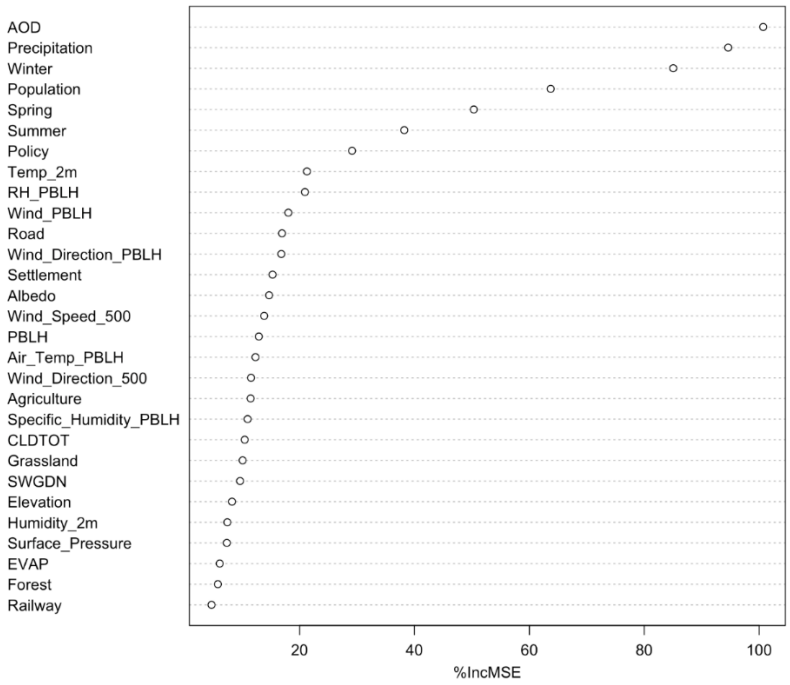

Figure 5. Importance Rank Plot

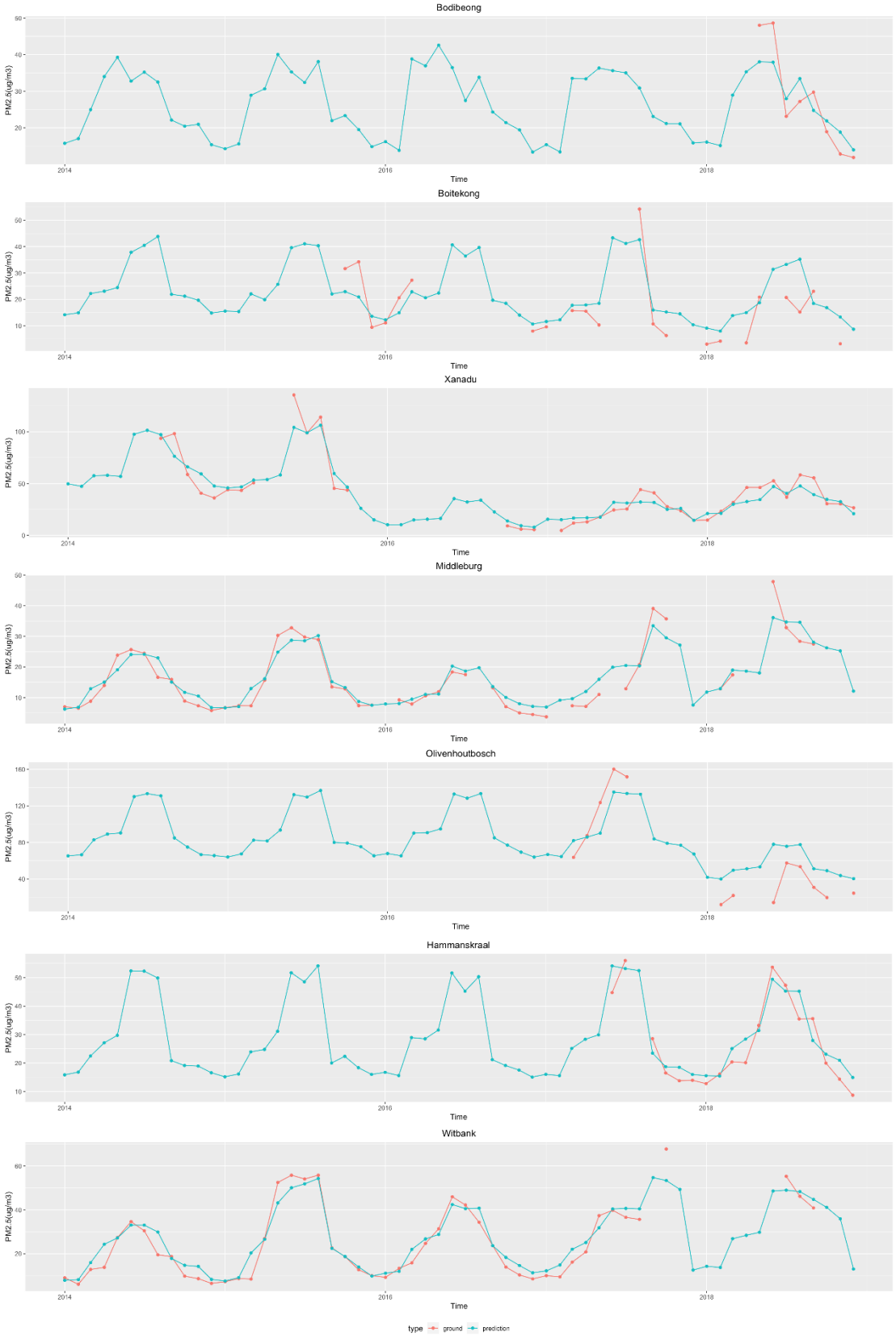

Figure 6. *Cont.*

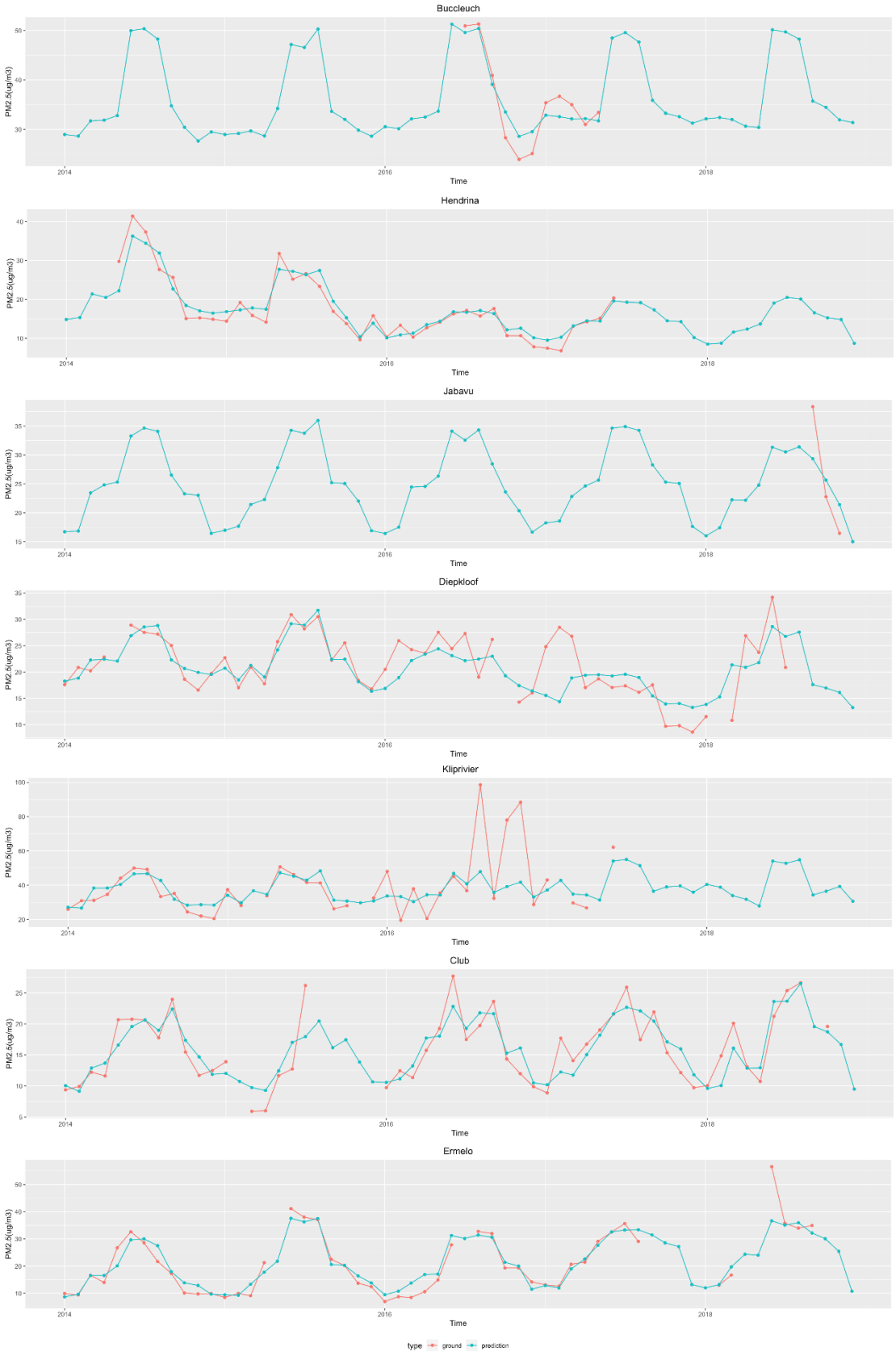

Figure 6. *Cont.*

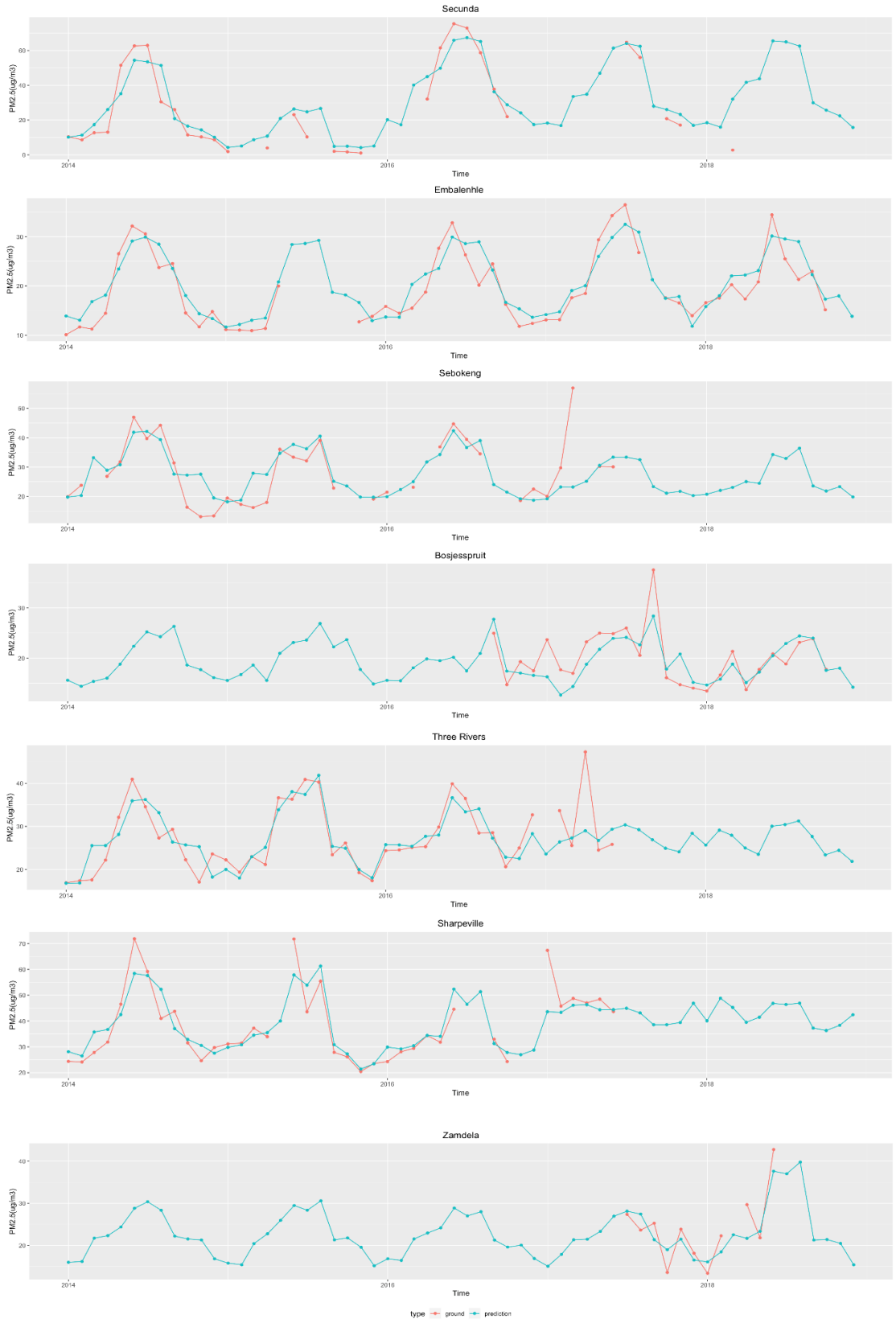

Figure 6. Observed and Estimated Monthly $PM_{2.5}$ Concentration for Each Station

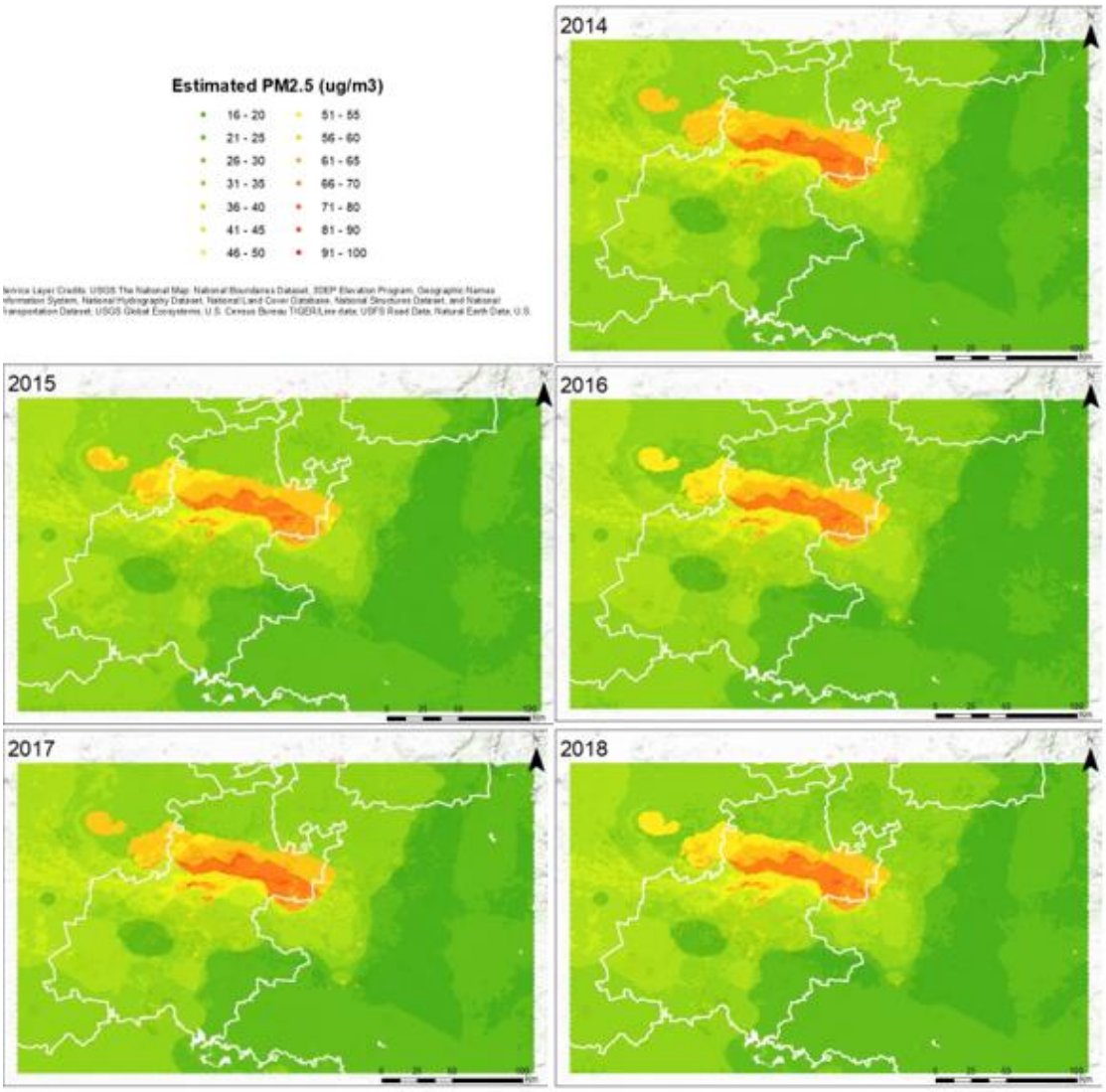

Figure 7. Annual Estimated $PM_{2.5}$ Concentration Map

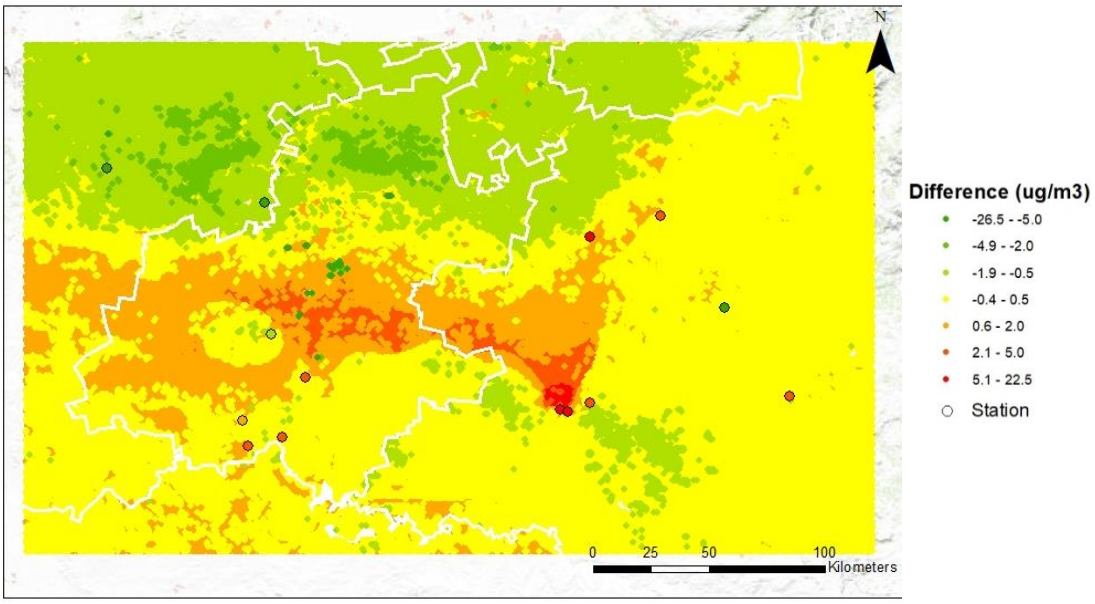

Figure 8. Difference in Annual $PM_{2.5}$ Concentration for New Standard

Table 1. National Ambient Air Quality Standard for $PM_{2.5}$

Averaging Period	**Concentration**	**Frequency of Exceedance**	**Compliance Date**
24 hours	65 µg/m^3	4	Immediate–31 December 2015
24 hours	40 µg/m^3	4	1 January 2016–31 December 2029
24 hours	25 µg/m^3	4	1 January 2030
1 year	25 µg/m^3	0	Immediate–31 December 2015
1 year	20 µg/m^3	0	1 January 2016–31 December 2029
1 year	15 µg/m^3	0	1 January 2030

Table 2. Number of Observations and Mean $PM_{2.5}$ for Ground Stations

Station Name	Number of Observations	Mean $PM_{2.5}$ ($\mu g/m^3$)
Bodibeong	212	25.21
Boitekong	72	17.75
Bosjesspruit	618	20.67
Buccleuch	190	34.55
Club	1316	16.18
Diepkloof	1043	21.80
Embalenhle	1386	19.64
Ermelo	1168	20.31
Hammanskraal	385	26.63
Hendrina	1020	17.65
Jabavu	78	24.15
Kliprivier	648	36.62
Middleburg	1281	16.17
Olivenhoutbosch	132	88.42
Sebokeng	662	29.88
Secunda	677	30.46
Sharpeville	927	37.86
Three Rivers	953	27.64
Witbank	1196	25.42
Xanadu	838	41.56
Zamdela	125	23.15

Table 3. Linear Regression between Aqua and Terra AOD

	$AOD_{Aqua} = \alpha_1 + \beta_1 * AOD_{Terra}$			$AOD_{Terra} = \alpha_2 + \beta_2 * AOD_{Aqua}$		
	α_1	β_1	R^2	α_2	β_2	R^2
Spring	0.034	0.848	0.57	0.029	0.676	0.57
Summer	0.053	0.682	0.31	0.049	0.451	0.31
Fall	0.021	0.892	0.57	0.024	0.644	0.57
Winter	0.019	0.823	0.59	0.024	0.718	0.59

Table 4. Mixed-effect Model for AERONET and Satellite AOD

	$Aqua_{AERONET} = \alpha_1 + \beta_1 * Aqua_{Satellite}$			$Terra_{AERONET} = \alpha_2 + \beta_2 * Terra_{Satellite}$		
	α_1	β_1	R^2	α_2	β_2	R^2
Spring	0.043	0.936	0.63	0.021	1.121	0.74
Summer	0.042	0.590		0.021	0.743	
Fall	0.042	0.711		0.021	0.909	
Winter	0.043	0.975		0.021	1.115	

Table 5. Percent of Area Meeting Annual $PM_{2.5}$ Standard

	< 20 µg/m^3	**20-25 µg/m^3**
Before	16%	26%
After	14%	26%