Air pollution scenario over Pakistan: characterization and ranking of extremely polluted cities using long-term concentrations of aerosols and trace gases

Article (Accepted Version)
Air Pollution Scenario over Pakistan: Characterization and Ranking of Extremely Polluted Cities using Long-Term Atmospheric Pollutants and Trace Gases

Muhammad Bilal¹, Alaa Mhawish¹, Janet E. Nichol², Zhongfeng Qiu¹, Majid Nazeer³, Md. Arfan Ali¹, Gerrit de Leeuw⁴,⁵,⁶,⁷, Yu Wang¹, Yang Chen⁸, Lunche Wang⁹, Yuan Shi¹⁰, Max P. Bleiweiss¹¹, Luqman Atique¹², Usman Mazhar¹³, and Song Ke¹⁴

¹ Lab of Environmental Remote Sensing (LERS), School of Marine Sciences, Nanjing University of Information Science and Technology, Nanjing, 210044, China.
² Department of Geography, School of Global Studies, University of Sussex, Brighton BN19RH, UK.
³ Key Laboratory of Digital Land and Resources, East China University of Technology, Nanchang 330013, China.
⁵ School of Atmospheric Physics, Nanjing University of Information Science and Technology, Nanjing, 210044, China.
⁶ Aerospace Information Research Institute, Chinese Academy of Sciences (AirCAS), No.20 Datun Road, Chaoyang District, Beijing 100101, China
⁷ School of Environment Science and Spatial Informatics, University of Mining and Technology, Xuzhou, Jiangsu 221116, China
⁸ State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China.
⁹ Laboratory of Critical Zone Evolution, School of Earth Sciences, China University of Geosciences, Wuhan 430074, China.
¹⁰ Institute of Future Cities, The Chinese University of Hong Kong, Hong Kong SAR, China.
¹¹ Department of Entomology, Plant Pathology and Weed Science, New Mexico State University, Las Cruces, NM 88003, USA.
¹² School of Earth Sciences, Zhejiang University, Hangzhou 310027, China.
¹³ School of Remote Sensing & Geomatics Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China.
¹⁴ Geological Survey of Jiangsu Province, Nanjing, 210018, China.

*Corresponding author: Zhongfeng Qiu (zhongfeng.qiu@nuist.edu.cn)
Abstract

Pakistan ranks third in the world in terms of mortality attributable to air pollution, having aerosol pollutant levels consistently well above WHO (World Health Organization) air quality guidelines (AQG). However, regulation is dependent on a sparse network of air quality monitoring stations and insufficient ground data. This study utilizes long-term atmospheric pollutants and trace gases to characterize and rank the air pollution scenarios and pollution characteristics of 80 selected cities in Pakistan. Datasets used include (1) the Aqua and Terra (AquaTerra) MODIS (Moderate Resolution Imaging Spectroradiometer) Level 2 Collection 6.1 merged Dark Target and Deep Blue (DTB) aerosol optical depth (AOD) retrievals; (2) the CAMS (Copernicus Atmosphere Monitoring Service) reanalysis PM$_1$, PM$_{2.5}$, and PM$_{10}$ data; (3) the MERRA-2 (Modern-Era Retrospective analysis for Research and Applications, Version 2) reanalysis PM$_{2.5}$ data, (4) the OMI (Ozone Monitoring Instrument) tropospheric vertical column density (TVCD) of nitrogen dioxide (NO$_2$), and VCD of sulfur dioxide (SO$_2$) in the Planetary Boundary Layer (PBL), (5) the VIIRS (Visible Infrared Imaging Radiometer Suite) Nighttime Lights data, (6) MODIS Collection 6 Version 2 global monthly fire location data (MCD14ML), (7) population density, (8) MODIS Level 3 Collection 6 land cover types, (9) AERONET (AErosol RObotic NETwork) Version 3 Level 2.0 data, and (10) ground-based PM$_{2.5}$ concentrations from air quality monitoring stations. Potential Source Contribution Function (PSCF) analyses were performed by integrating with ground-based PM$_{2.5}$ concentrations and the NOAA (National Oceanic and Atmospheric Administration) HYSPLIT (Hybrid Single-Particle Lagrangian Integrated Trajectory) air parcel back trajectories to identify potential pollution source areas which are responsible for extreme air pollution in Pakistan.
Results show that the ranking of the top polluted cities varies with respect to atmospheric pollutants and trace gases. For example, Jhang, Multan, and Vehari were characterized as the top three polluted cities in Pakistan for the AquaTerra DTB AOD retrievals; Lahore, Gujranwala, and Okara for PM\(_1\), PM\(_{2.5}\), and PM\(_{10}\); Lahore, Rawalpindi, and Islamabad for NO\(_2\); and Lahore, Mirpur, and Gujranwala for SO\(_2\). The results demonstrate that the whole of Pakistan is exposed to long-term PM\(_{2.5}\) concentrations (54.7 µg/m\(^3\); mean annual value for Pakistan between 2003 and 2020) exceeding Pakistan’s National Environmental Quality Standards (Pak-NEQS, i.e., <15 µg/m\(^3\) annual mean) for ambient air defined by the Pakistan Environmental Protection Agency (Pak-EPA) as well as the WHO Interim Target-1 (i.e., mean annual PM\(_{2.5}\) <35 µg/m\(^3\)). The spatial analyses of atmospheric pollutants and trace gases with the support of population density, nighttime lights, land cover types, and fire location data, and the PSCF analysis show that Pakistan’s air quality is mainly affected by local anthropogenic sources rather than regional sources. Statistically significant positive trends in PM\(_1\), PM\(_{2.5}\), PM\(_{10}\), NO\(_2\), and SO\(_2\) concentrations were observed in ~89%, ~67%, ~48%, 91%, and ~88% of the cities of Pakistan, respectively. This comprehensive examination of aerosol pollutant sources and characteristics in spatio-temporal domains and their trends over Pakistan is the first of its kind. Results will be helpful to the Ministry of Climate Change (Government of Pakistan), Pak-EPA, SUPARCO (Pakistan Space and Upper Atmosphere Research Commission), policymakers, and the local research community to mitigate air pollution and its effects on human health.

**Keywords:** MODIS; AOD; CAMS; MERRA-2; PM\(_1\); PM\(_{2.5}\); PM\(_{10}\); OMI; NO\(_2\); SO\(_2\); PSCF; Pakistan
1. Introduction

With the rapid increase in population and overexploitation of natural resources, air pollution is of serious global environmental concern. According to the World Health Organization (WHO 2018a), air pollution levels are dangerously high worldwide as 9 out of 10 people breathe polluted air, and 7 million deaths each year are caused by outdoor and indoor aerosol pollutants. Outdoor (ambient) air pollution is a mixture of different pollutants including airborne particulate matter (PM), ozone (O₃), nitrogen dioxide (NO₂), volatile organic compounds (VOC), carbon monoxide (CO), and sulfur dioxide (SO₂), which have adverse health effects (Mannucci and Franchini 2017). Although air pollution is a global problem, the latest WHO air quality database reveals that 97% of affected cities are in low- and middle-income countries with more than 100,000 inhabitants (WHO 2018b). Air pollution is endemic to Pakistan, being listed among low- and middle-income countries as well as being the most urbanized of its South Asian counterparts (77.42 million or 36.37 % of the urban population, with 2.52 % annual growth rate) (UNDP 2019). Purohit et al. (2013) predicted that under current emission control standards, air pollution would decrease life...
expectancy by more than 100 months by 2030. The Health Effects Institute (2019) reported that since 1990, Pakistan’s entire population has been exposed to PM$_{2.5}$ (fine suspended particulate matter with an aerodynamic diameter less than 2.5 µm) levels exceeding WHO Interim Target-1 (i.e., <35 µg/m$^3$), with 58 µg/m$^3$ in 2017. Pakistan ranks third in the world in terms of mortality attributable to air pollution, with an annual loss of 128,000 lives (Government of Pakistan 2019). Recently, on October 30, 2019, the Air Quality Index (AQI) was 484 in Lahore, the second-largest city with the highest urbanization rate of 6.12 percent per annum, well above the threshold of 300 for “hazardous” level (Amnesty International 2019). The winter of 2019-2020 witnessed a spate of smog, fueled by the buildup of anthropogenic aerosols having 65% of sources within Pakistan, compelling authorities in Punjab to close schools for an extended period. The principal cause for smog formation is NOx, derived 58% from Pakistan’s 23.6 million transport vehicles, followed by industry and power, which generate another 34% (Amnesty International 2019; Government of Pakistan 2019; UNDP 2019). According to the Pakistan Air Quality Initiative (PAQI), Lahore, Peshawar, Islamabad, and Karachi are the most polluted cities where air quality does not meet WHO air quality guidelines during autumn and winter (PAQI 2018). Air pollution monitoring throughout Pakistan is challenging, due to sparse air quality monitoring stations and the unavailability of sufficient ground data, though several remote sensing studies have been conducted.

Gupta et al. (2013) analyzed MODIS (Moderate Resolution Imaging Spectroradiometer) AOD (Aerosol Optical Depth) retrievals over Lahore and Karachi from 2001 to 2010 and reported higher aerosol loadings near the city center than outside the city. Tariq et al. (2016) analyzed ground-based and satellite-based aerosol optical properties over Lahore during intense haze.
events in October 2013 and reported crop residue burning and urban-industrial emissions as the main sources of high AOD levels. Bilal et al. (2016) evaluated the performance of the Aqua-MODIS (MYD04) level 2 aerosol products over Lahore and Karachi from 2007 to 2013, and recommended the use of Dark Target (DT) and Deep Blue (DB) algorithms over Karachi and Lahore respectively, for regional air quality applications, as these cities have different land cover characteristics and aerosol types. Similarly, Awais et al. (2018) evaluated MODIS aerosol products against handheld Sunphotometer measurements over Pakistan’s twin cities, Islamabad and Rawalpindi, and reported good performance of the MODIS aerosol product. Other remote sensing studies have been conducted of atmospheric trace gases, such as ozone ($O_3$), nitrogen dioxide ($NO_2$), sulfur dioxide ($SO_2$), and carbon dioxide $CO_2$, as well as their trends over time (Khokhar et al. 2016; Khokhar et al. 2015; Tariq and Ali 2015; ul-Haq et al. 2017; Ul-Haq et al. 2014; Ul_Haq et al. 2015).

Zhang et al. (2020) conducted the first study of the vertical distribution of aerosol optical properties over Pakistan using CALIPSO (Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation) data.

Cities are high activity areas, and every city is a huge source of local anthropogenic aerosols from industrial and human activities, which can impact air quality, visibility, and alter the physico-chemical properties of the atmosphere at local, regional, and global scales. Although several studies of AOD and atmospheric trace gases have been conducted over Pakistan, no study has been encompassing different particulate types ($PM_x$, $x = 1, 2.5, \text{and} 10$) at the national scale, i.e., ultrafine PM with an aerodynamic diameter less than 1 μm ($PM_1$), fine suspended PM with an aerodynamic diameter less than 2.5 μm ($PM_{2.5}$), and respirable suspended coarse PM with an aerodynamic diameter less than 10 μm ($PM_{10}$). It is of great importance to identify the cities most
affected by $\text{PM}_{x}$, as they have different health and secondary effects and this is the first study to do so. Moreover, very few studies have investigated the long-term trend in air pollutants at the city level, which can provide additional insight into the linkage between air pollutant trends and the changes in emissions. Furthermore, previous studies are not comprehensive enough to answer the difficult questions; which are the most and least polluted cities of Pakistan, and what are the likely pollution sources? Therefore, this study aims (1) to extensively characterize and rank the extremely polluted cities of Pakistan, considering multiple sources and aerosol pollutant types, for 80 carefully selected cities, representing almost all major urban centers of Pakistan, and (2) to identify the likely pollutant sources by performing PSCF (Potential Source Contribution Function) analysis with the integration of HYSPLIT (Hybrid Single-Particle Lagrangian Integrated Trajectory) back trajectory and ground-based $\text{PM}_{2.5}$ concentrations. This study is based on long-term combined Aqua and Terra (AquaTerra) MODIS data from 2003 to 2017, OMI (Ozone Monitoring Instrument) data ($\text{NO}_2$ and $\text{SO}_2$) from 2004 to 2019, CAMS (Copernicus Atmosphere Monitoring Service) reanalysis $\text{PM}_{1}$, $\text{PM}_{2.5}$, and $\text{PM}_{10}$ data from 2003 to 2019, MERRA-2 (Modern-Era Retrospective analysis for Research and Applications, Version 2) $\text{PM}_{2.5}$ data from 2003 to 2020, VIIRS (Visible Infrared Imaging Radiometer Suite) Nighttime Lights from 2012 to 2019, LandScan global population density for 2019, MODIS land cover type for 2019, MODIS global monthly fire location data from 2003 to 2020, ground-based $\text{PM}_{2.5}$ concentrations from 2018 to 2020, and AERONET (AErosol RObotic NETwork) AOD measurements from 2006 to 2017.
2. Study Area

Pakistan, with a population of 212.82 million, is the sixth most populous country in the world. It lies between 23°35’ to 37°05’ North and 60°50’ to 77°50’ East, having a diverse geographical landscape bordered by China, the Himalayas, India, Afghanistan, Iran, and the Arabian Sea. Geographically, Pakistan falls into three major regions: the northern highlands, constituting parts of the Hindu Kush, the Karakoram Range, and the Himalayas; the Indus River basin plain in the center and east (65% of the total area i.e. 796,096 km$^2$); and the Balochistan Plateau in the south and west (Government of Pakistan 2019). Administratively, Pakistan has six units: Punjab, Sindh, Khyber Pakhtunkhwa, Balochistan, Azad Kashmir, and Gilgit Baltistan. Punjab is the most populous (112.38 million; 53%) administrative unit of Pakistan, followed by Sindh (49.05 million; 23%), Khyber Pakhtunkhwa (36.5 million; 17%), and Balochistan (12.7 million; 6%). Balochistan has the largest area (43.6 %) followed by Punjab (25.8%), Sindh (17.7 %), and Khyber Pakhtunkhwa (12.78%). Sindh is the most urbanized and industrialized administrative unit of Pakistan with 52% urban population. Islamabad (2.1 million; 1%) Capital Territory (ICT), a rather small unit in terms of area (0.1 %), is, in fact, the second most urbanized (50.58%) region of Pakistan, and has an annual urbanization rate of 4.91 %. Currently, 10 cities in Pakistan have a population of over one million, and 7 have higher per-capita incomes than the national average (UNDP 2019). The Pakistan economic survey 2018-19 reports a total cropped area of 22.6 million hectares, and agricultural contributions of 18.5 % to the GDP, compared with 20.3% from the industrial sector (Government of Pakistan 2019).
This study covers almost all prominent cities in Pakistan including all administrative units and their capital cities, and the Capital of the country (Figure 1). In summary, the study area analyzes 23 cities from the most populated administrative unit, Punjab; Khyber Pakhtunkhwa is also well-represented by 19 urban centers; Balochistan is the least populated but the largest administrative unit, and is represented by 19 cities; 14 other cities exemplify the diversity of Sindh in the South-East, and 5 cities represent the attractive hilly land of Azad Kashmir.
Figure 1: Geographical and administrative map of Pakistan including a list of cities used in the present study. Cities are characterized using (a) yearly mean CAMS (Copernicus Atmosphere Monitoring Service) reanalysis PM$_{2.5}$ concentrations ($\mu$g/m$^3$) for the years 2003 and 2020, and (b) yearly mean AquaTerra MODIS DTB AOD retrievals at 550 nm from 2003 to 2017. Extremely polluted cities (red color) are defined for PM$_{2.5} > 92.84$ (AOD > 0.6) (3$^{rd}$ quartile), highly polluted cities (brown color) for $45.69 \leq$ PM$_{2.5} \leq 92.84$ (0.3 < AOD < 0.6) (between 3$^{rd}$ and 1$^{st}$ quartiles), and polluted cities (purple color) for PM$_{2.5} < 45.69$ (AOD < 0.3) (1$^{st}$ quartile) using descriptive statistics (Table S1). Cities are not defined as low polluted or clean cities as annual mean PM$_{2.5}$ concentrations for all cities exceed Pakistan’s National Environmental Quality Standards (Pakistan NEQS) for ambient air (<15 $\mu$g/m$^3$ annual mean).

3. Dataset

3.1 AERONET Data

The AERONET (AErosol RObotic NETwork) (Holben et al. 1998; Holben et al. 2001) is a global network of calibrated Sunphotometers coordinated by NASA (National Aeronautics and Space Administration) which provides regular measurements of spectral AOD at 340 nm, 380 nm, 440 nm, 500 nm, 675 nm, 870 nm, 1020 nm, and 1640 nm, and AE at 340–440 nm, 380–500 nm, 440–675 nm, and 500–870 nm at three levels, i.e., Level 1.0 (unscreened), Level 1.5 (cloud-screened), and Level 2.0 (cloud-screened and quality-assured), under cloud-free skies (Smirnov et al. 2000) for every 15 minutes with an uncertainty of 0.01–0.02 (Holben et al. 2001). The present study used Version 3 Level 2.0 AOD at 500 nm (AOD$_{500}$) and AE at 440–675 nm (AE$_{440-675}$) (Giles et al. 2019) obtained from the AERONET website (https://aeronet.gsfc.nasa.gov/) for the Lahore
(31.47987° N, 74.26406° E) and Karachi (24.94574° N, 67.13594° E) sites from 2006 to 2017. The
Lahore and Karachi AERONET sites are located in an urban area, and approximately 20 km away
from the Arabian Sea coast, respectively.

3.2 AquaTerra MODIS Data

In the present study, Aqua and Terra MODIS C6.1 L2 aerosol products at 10 km spatial
resolution are obtained from 2003 to 2017 for Pakistan from the LAADS DAAC
(https://ladsweb.modaps.eosdis.nasa.gov/). The MODIS aerosol product provides DT AOD
retrievals over land and water surfaces (Levy et al. 2013), and DB AOD retrievals only over land
(Hsu et al. 2013). The DT and DB AOD retrievals for different collections are extensively validated
against Sunphotometer (AERONET) measurements at regional (Bilal et al. 2019b; Bilal et al. 2014;
Che et al. 2019; de Leeuw et al. 2018; Fan et al. 2017; Filonchyk et al. 2019; Gupta et al. 2013; He
and Bilal 2016; Shen et al. 2018; Shi et al. 2013; Sogacheva et al. 2018; Wang et al. 2017; Wang
et al. 2019; Xiao et al. 2016; Xie et al. 2011) and global scales (Bilal et al. 2018a; Bilal et al. 2017;
et al. 2014; Sayer et al. 2015; Tong et al. 2020). These studies have reported overestimation and
underestimation in DT and DB AOD retrievals respectively, due to error in the estimated surface
reflectance and aerosol scheme used in the inversion methods, but overall their performance is
satisfactory. Previous studies (Bilal et al. 2018a; Bilal and Nichol 2017; Bilal et al. 2017; Bilal et al.
2018b; Mei et al. 2019; Sayer et al. 2014) have also reported different spatial coverage of DT and
DB AOD retrievals over land due to differences in their approaches, i.e., pixel selection criteria,
estimation of surface reflectance, and the cloud mask. Therefore, a new merged Scientific Data Set (SDS: AOD 550 Dark Target Deep Blue Combined) was introduced which contains only the highest quality DT and DB (DTB) AOD retrievals or their average values (Levy et al. 2013). The purpose of this new dataset is to improve spatial coverage over land (Levy et al., 2013; Sayer et al., 2014), i.e., to retrieve AOD in the same image for those regions where the DT algorithm does not retrieve or the DB algorithm does not retrieve (Bilal et al. 2017; Levy et al. 2013). The merged DTB AOD retrievals have been validated at regional and global scales (Ali and Assiri 2019; Bilal et al. 2018a; Bilal and Nichol 2017; Bilal et al. 2017; Sayer et al. 2014; Sogacheva et al. 2018). However, the new customized method-1 (CM1) (Bilal et al. 2017), which is named Simplified Merge Scheme (SMS) in the later publications (Bilal et al. 2018a; Bilal et al. 2018b), provides equally consistent data quality with the combined DTB AOD retrievals available in C6.1, but with significantly improved spatio-temporal coverage.

3.3 CAMS Data

The Copernicus Atmosphere Monitoring Service (CAMS) reanalysis is an atmospheric composition dataset generated by the European Centre for Medium-Range Weather Forecasts (ECMWF). The global CAMS model combines the satellite-based observations with chemistry-aerosol modeling using the four-dimensional variational (4D-VAR) data assimilation technique to obtain the mass concentration of aerosols and trace gases. CAMS uses the MACCity inventory at 0.5° × 0.5° spatial resolution for anthropogenic emissions which covers the period 1960–2010 (Granier et al. 2011). Detailed information about the model and the emission inventory can be found in (Flemming et al. 2017; Flemming et al. 2015). In this study, the ground-based mass
concentration of particulate matter, including particles with an aerodynamic diameter of less
than 1 µm (PM$_1$), less than 2.5 µm (PM$_{2.5}$), and less than 10 µm (PM$_{10}$) was obtained from the
CAMS reanalysis data for the years 2003 and 2020. PM$_x$ (x = 1, 2.5, & 10) data were used at two
different spatiotemporal resolutions, i.e., (i) CAMS global reanalysis dataset at 0.75° × 0.75°
spatial resolution and 3-hourly temporal resolution from 2003 to 2020, and (ii) CAMS near-real
time dataset at 0.125° × 0.125° spatial resolution and 12-hourly temporal resolution from 2018
to 2020 (Inness et al. 2019). The PM$_x$ data at 0.75° grid size and 3-hourly temporal resolution
were used for long-term climatology and for characterizing extremely polluted cities, whereas,
the CAMS near-real time data at 0.125° grid size and 12-hourly temporal resolution were used
for validation against ground-based PM$_{2.5}$ concentrations obtained from air quality monitoring
stations.

3.4 MERRA-2 Reanalysis Data

The MERRA-2 (Modern-Era Retrospective analysis for Research and Applications, Version 2)
atmospheric reanalysis is the latest data released by the NASA GMAO (Global Modeling and
Assimilation Office) in 2017 (Buchard et al. 2017; Randles et al. 2017). The MERRA-2 aerosol
gridded data, i.e., dust, sea salt, sulfate, black carbon, and organic carbon, are simulated with 72
vertical layers from the surface to higher than 80 km using the GEOS-5 (GMAO Earth system
model version 5) model radiatively coupled to the GOCART (Goddard Chemistry Aerosol
Radiation and Transport) model (Chin et al. 2002; Colarco et al. 2010). For anthropogenic
emissions, MERRA-2 uses EDGAR-4.2 emission inventory at 0.1° × 0.1° spatial resolution and
covers the period 1970–2008 (Janssens-Maenhout et al. 2013). In this study, the MERRA-2
aerosol gridded data (dust, sea salt, sulfate, black carbon, and organic carbon) at 0.5° × 0.625° spatial resolution from 2018 to 2020 were used. More details about MERRA-2 reanalysis data can be found in Randles et al. (2017) and Buchard et al. (2017).

3.5 Ground-based PM$_{2.5}$ Measurements

Ground-based PM$_{2.5}$ measurements were obtained from two different air quality monitoring networks. Firstly, PM$_{2.5}$ data were obtained from 4 air quality stations operated by the US Consulates in Islamabad, Karachi, Lahore, and Peshawar, and secondly, 54 air quality monitoring stations operated by PAQI in Lahore (24 stations), Karachi (15), Islamabad (5) Sialkot (3), Peshawar (2), Rawalpindi (2), Faisalabad (1), Gujranwala (1), and Muridke (1). Due to the lack of a well-developed and standard air quality network of ground-based PM$_{2.5}$ measurements, this study is limited to only these cities for the validation of CAMS and MERRA-2 reanalysis PM$_{2.5}$ gridded data. PM$_{2.5}$ concentrations from the US Consulates are measured by beta gauge attenuation monitors (BAM-1020; Met One Instruments), hereafter referred to as BAM PM$_{2.5}$ concentrations. To increase social awareness in Pakistan, PAQI provides PM$_{2.5}$ data using a nationwide network of low-cost air quality monitors (IQAir AirVisual Pro), hereafter referred to as LCM PM$_{2.5}$ concentrations. In this study, LCM and BAM PM$_{2.5}$ measurements were used for January 2018–December 2019 and January 2019–February 2021, respectively. More details about PAQI (LCM) and US Consulates (BAM) PM$_{2.5}$ data can be found in Shi et al. (2020) and Mhawish et al. (2020), respectively.
3.6 OMI Data

The Ozone Monitoring Instrument (OMI) onboard the Aura satellite was launched in July 2004 as a part of the A-Train satellite constellation. OMI is a hyperspectral sensor that measures the radiation reflected from the earth-atmosphere system, in the wavelength range 250–500 nm and provides daily global coverage at a spatial resolution of 13 × 24 km² at nadir. The OMI OMAERUV algorithm utilizes the sensitivity of near-UV spectral regions to aerosol absorption, and it retrieves absorbing aerosol optical depth (AAOD) at 388 nm (Torres et al. 2013; Torres et al. 2007). Along with the AAOD, the OMAERUV algorithm also provides an ultraviolet Aerosol Index (UVAl), AOD, and Single Scattering Albedo (SSA). OMI also retrieves the atmospheric trace gases O₃, NO₂ and SO₂ (Carn et al. 2017; Krotkov et al. 2017; Krotkov et al. 2016; Li et al. 2017; Li et al. 2013; Veefkind et al. 2006). In this study, OMAERUV version 3 Level 3 daily cloud-screened (cloud fraction < 30 %) NO₂ tropospheric vertical column density (TVCD) (OMNO2e), and SO₂ VCD in the planetary boundary layer (PBL) (OMSO2e) gridded at 0.25° × 0.25° spatial resolution from 2004 to 2019 were used.

3.7 Other Supporting Datasets

Other supporting datasets include (i) annual mean VIIRS nighttime lights data (https://eogdata.mines.edu/products/vnl/) from 2012 to 2019 derived from monthly mean data (Elvidge et al. 2021), (ii) MODIS Collection 6 global monthly Fire Location product (MCD14ML) from 2003 to 2020 (https://firms.modaps.eosdis.nasa.gov/download/), (iv) MODIS Collection 6 Level 3 land cover type product (MCD12Q1) for 2019.
(https://ladsweb.modaps.eosdis.nasa.gov/), and (v) the LandScan population density (https://landscan.ornl.gov/) for 2019 (Rose et al. 2020).

### 4. Research Methodology

To investigate the air pollution scenario over Pakistan and characterize the extremely polluted cities, in this study the following methodology was adopted:

1. MODIS AOD retrievals were obtained from the Scientific Data Set (SDS) “Optical Depth Land and Ocean” and “Deep Blue Aerosol Optical Depth 550 Land Best Estimate”. Only the highest quality-assured DT (QA = 3) and DB (QA ≥ 2) retrievals were used, as recommended by previous studies (Bilal et al. 2013; Levy et al. 2013; Mhawish et al. 2019; Sayer et al. 2013). Pakistan has a variety of land cover types, e.g., snow and mountainous land surface in Northern Pakistan, plain and agricultural land surfaces in Central Pakistan, and arid and desert land surfaces in southern Pakistan, where the DT and DB algorithms overestimate and underestimate, respectively. However, the DT algorithm is unable to provide retrievals over the arid and desert land surfaces of Balochistan. Similar results were observed and reported in our previous study over Pakistan (Bilal et al. 2016).

Therefore, in the present study, we preferred to generate the combined (merged) DTB AOD$_{550}$ retrievals for both Aqua and Terra MODIS data from 2003 to 2017 using the customized method-1 (CM1) (Bilal et al. 2017), which in later publications is named Simplified Merge Scheme (SMS) (Bilal et al. 2018a; Bilal et al. 2018b), i.e., an average of the DT and DB AOD retrievals or the available one with the highest quality flag (Equation 1), to enhance spatio-temporal coverage.
2. Aqua and Terra MODIS may not provide complete spatial coverage due to cloud cover. On days when Aqua provides AOD retrievals, Terra may not, and vice-versa. Therefore, for more complete spatial coverage between Aqua and Terra as well as to represent an average air pollution scenario between morning and afternoon times with a single dataset, the combined AquaTerra DTB AOD retrievals were generated from the Aqua DTB and Terra DTB AOD retrievals using SMS/CM1, i.e., an average of the Aqua and Terra DTB AOD retrievals or the available one (Equation 2).

\[
\text{AquaTerra AOD} = \begin{cases} 
\text{if only Aqua AOD exists} & \rightarrow \text{Aqua} \\
\text{if only Terra AOD exists} & \rightarrow \text{Terra} \\
\text{if both Aqua and Terra AOD exist} & \rightarrow \frac{(\text{Aqua} + \text{Terra})}{2}
\end{cases}
\] (2)

3. The AquaTerra DTB AOD retrievals are validated against Sunphotometer AOD measurements obtained for Lahore (31.480° N and 74.264° E) and Karachi (24.946° N and 67.136° E) AERONET sites. The AERONET Sunphotometer does not provide AOD at 550 nm (AOD\text{550}), AOD\text{550} is interpolated using AOD at 500 nm (AOD\text{500}) and Ångström Exponent at 440-675 nm (AE\text{440-675}) based on the Ångström Exponent empirical formula (Equation 3) (Eck et al. 1999). Collocated AquaTerra and AERONET AOD retrievals were defined as the average of at least two pixels of DTB within a spatial region of 3 × 3 pixels.
(at least 2 out of 9 pixels) centered on the AERONET site and the average of at least two AERONET AOD measurements between 10:00 and 14:30 local solar time.

\[ AOD_{550} = AOD_{500} \left( \frac{550}{500} \right)^{-AE_{440-667}} \]  

4. Accuracy and errors are reported using the Pearson correlation coefficient \( (r) \), the expected error (EE, Equation 4), and relative mean bias (RMB, Equation 5). The slope \( (\beta, \text{Equation 6}) \) and intercept \( (\alpha, \text{Equation 7}) \) between collocated AquaTerra DTB and AERONET AOD data are calculated using the reduced major axis (RMA) regression which incorporates errors in both independent (AERONET) and dependent (MODIS) variables (Bilal et al. 2019a; Harper 2016). The performance of the Terra and Aqua DT, DB, and DTB AOD retrievals is evaluated based on (i) highest correlation coefficient \( (r) \), (ii) highest number of collocated retrievals \( (N) \), (iii) the highest percentage of retrievals within the EE, and (iv) lowest RMB. To evaluate the performance of the collocated retrievals, the following criteria are utilized (Bilal et al. 2017): the DT, DB, and DTB retrievals are considered to be of equal quality if the relative difference is within (1) 5% for the correlation coefficient \( (r) \), (2) 10% for the collocated retrievals, (3) 10% for the percentage of retrievals is within the EE, and (4) RMB < 25%.

\[ EE = \pm (0.05 + 0.20 \times AERONET_{AOD}) \]  

The upper and lower EE envelopes are calculated using Equations 4a and 4b.

\[ Upper \ EE \ envelope = AERONET_{AOD} + |EE| \]  

\[ Lower \ EE \ envelope = AERONET_{AOD} - |EE| \]
Lower EE envelope \(=\) AERONET\(_{AOD}\) \(-\) |EE| \(\) (4b)\)

The percentage of best retrieved MODIS AOD retrievals within the EE is reported using Equation 4c.

\[
%EE = AERONET_{AOD} - |EE| \leq MODIS_{AOD} \leq AERONET_{AOD} + |EE| \quad (4c)
\]

Where |EE| is the absolute value of EE.

\[
RMB = \frac{(MODIS_{AOD} - AERONET_{AOD})}{AERONET_{AOD}} \times 100 \quad (5)
\]

Where, \(MODIS_{AOD}\) and \(AERONET_{AOD}\) are the mean of MODIS and AERONET AOD retrievals, respectively. RMB > 0 represents overestimation in MODIS AOD compared to AERONET AOD, RMB < 0 represents underestimation, and RMB = 0 represents no over- and under-estimations.

\[
\beta = \frac{\sigma_{MODIS_{AOD}}}{\sigma_{AERONET_{AOD}}} \quad (6)
\]

\[
\alpha = MODIS_{AOD} - \left(\frac{\sigma_{MODIS_{AOD}}}{\sigma_{AERONET_{AOD}}}\right) \times AERONET_{AOD} \quad (7)
\]

Where, \(\beta, \alpha, \sigma_{MODIS_{AOD}}, \) and \(\sigma_{AERONET_{AOD}}\) are the slope, intercept, the standard deviation of MODIS AOD, and standard deviation of AERONET AOD, respectively.

5. To show the long-term variation of the mean spatial distributions of AquaTerra AOD over Pakistan, the AOD retrievals from 2003 to 2017 are used to generate monthly mean
spatial AOD maps, and their corresponding pixel counts are calculated for reporting the
retrieval performance of both the DT and DB algorithms.

6. To assure the quality of the PM$_{2.5}$ data, validation of daily average CAMS and MERRA-2
PM$_{2.5}$ data was conducted against in-situ PM$_{2.5}$ measurements obtained from the air
quality monitoring stations. The performance was evaluated based on the correlation
coefficient (r), RMB (Eq. 5), and slope (Eq. 6). MERRA-2 PM$_{2.5}$ concentrations were
calculated based on five aerosol pollutants using Equation 8 (Song et al. 2018), and CAMS
PM$_{2.5}$ and PM$_{10}$ concentrations were calculated using Equations 9 and 10 (Rémy et al.
2019).

\[
PM_{2.5} = [Dust_{2.5}] + [SS_{2.5}] + 1.375 \times [SO_4] + [BC] + 1.6 \times [OC]
\] (8)

Where, Dust$_{2.5}$, SS$_{2.5}$, BC, OC, and SO$_4$ are the GOCART concentrations of dust sea salt,
sulfate, black carbon, and organic carbon particulate matter with a diameter less than 2.5,
respectively.

\[
PM_{2.5} = \rho([SS_1]/4.3 + [SS_2]/4.3 + [DD_1] + [DD_2] + 0.7[OM] + [BC] + 0.7[SU] \\
+ 0.7[NI_1] + 0.25[NI_2] + 0.7[AM])
\] (9)

\[
PM_{10} = \rho([SS_1]/4.3 + [SS_2]/4.3 + [DD_1] + [DD_2] + 0.4[DD_3] + [OM] + [BC] \\
+ [SU] + [NI_1] + [NI_2] + [AM])
\] (10)

Where [SS$_{1,2}$] = sea salt aerosol, [DD$_{1,2,3}$] = desert dust, [NI$_{1,2}$] = nitrate, [OM] = organic
7. To characterize extremely polluted cities in Pakistan, the DTB AOD retrievals from AquaTerra, the PM$_{1}$, PM$_{2.5}$, and PM$_{10}$ from CAMS data, and the SO$_{2}$ and NO$_{2}$ from OMI are used. Polluted months as well as years, for the corresponding polluted cities, are also characterized based on each pollutant.

8. To assess recent changes in air pollutants, the non-parametric Mann Kendall test (Kendall and Gibbons 1990; Mann 1945) associated with Theil-Sen’s slope (Sen 1968; Theil 1992) was used to estimate and detect trends over the main cities of Pakistan from 2003 to 2020. The non-parametric Mann Kendall test is often used to detect monotonic trends in a time series, and is also found to be more suitable for non-normally distributed data, or if the data have some missing observations such as environmental data. Further, the bootstrapping technique was used to eliminate the serial autocorrelation in the monthly mean aggregated time series data and increase the robustness of the test (Hamed and Ramachandra Rao 1998; Salmi et al. 2002). The significance of the calculated trend was assessed using the two-tailed test method at a 95% confidence interval.

9. The NOAA (National Oceanic and Atmospheric Administration) HYSPLIT (Hybrid Single-Particle Lagrangian Integrated Trajectory Model) (Stein et al. 2015), a complete transport, dispersion, and chemical transformation model, is used for back trajectory analysis to determine the origin of air masses (Fleming et al. 2012) and highlight the possible sources of aerosol pollutants affecting the air quality of Pakistan using the PSCF (Potential Source Contribution Function) analysis. In this study, 72 hours HYSPLIT backward trajectories at a height of 500 m above the ground level (AGL) were computed for every 6 hours at
seasonal scales from March 2020 to February 2021 using the GDAS (Global Data Assimilation System) meteorological data at 1° × 1° spatial resolution (available at ftp://arlftp.arlhq.noaa.gov/pub/archives/gdas1) at seasonal scales from March 2020 to February 2021. The PSCF analysis was performed for 4 cities selected based on the availability of ground-based PM$_{2.5}$ measurements from the air quality stations operated by the US Consulates, namely, Peshawar, Islamabad, Lahore, and Karachi. The height of 500 m AGL has been reported very useful as it is the approximate height of the mixing layer (Begum et al. 2005). The backward trajectory clustering and investigation of the origins of the particulate matter at the receptor locations were studied using MeteoInfo TrajStat software (Version 2.0, available at http://meteothink.org/products/trajstat.html) (Wang et al. 2009) in conjunction with HYSPLIT and Geographic Information System (GIS).

The PSCF analysis was performed at 0.5° grid size using 24-hour average ground-based PM$_{2.5}$ concentrations for the days that exceeded the Pak-NEQS 24-hour air quality standards (35 µg/m$^3$). The PSCF value for a specific grid cell was calculated on the assumption that the trajectory endpoint is located within a cell (i, j) and the trajectory is assumed to collect pollutants emitted from different pocket emission sources within that cell (i, j). The PSCF value can be interpreted as a conditional probability describing the potential contributions of a grid cell to the high PM$_{2.5}$ loadings at the receptor site. The error associated with the trajectory is proportional to the distance from the receptor location (Begum et al. 2005). The PSCF value for the $ij$th grid cell can be computed using Equation 11:

$$PSCF(i,j) = \frac{m_{ij}}{n_{ij}}$$  (11)
Where \( n_{ij} \) represents the number of endpoints that fall or pass through the \( ij^{th} \) cell and \( m_{ij} \) denotes for the number of endpoints in the \( ij^{th} \) cell having a higher pollutant concentration than 24-hour Pak-NEQS. The uncertainty arising due to small \( n_{ij} \) is reduced by multiplying an arbitrary weight function \( W_{i,j} \), which is multiplied into the PSCF. In this case, the weight function is given in Equation (12):

\[
W_{i,j} = \begin{cases} 
1.00 & \text{if } n_{ij} > 3\bar{n} \\
0.70 & \text{if } 1.5\bar{n} < n_{ij} \leq 3\bar{n} \\
0.42 & \text{if } \bar{n} < n_{ij} \leq 1.5\bar{n} \\
0.15 & \text{if } n_{ij} \leq \bar{n}
\end{cases}
\] (12)

Where \( \bar{n} \) denotes the average number of endpoints per cell, which is calculated for each cell that has at least one endpoint. Therefore, the Weighted PSCF is expressed as Equation (13):

\[
WPSCF = W_{i,j} \times PSCF (i,j)
\] (13)

5. Results and Discussion

5.1 Aqua and Terra MODIS AOD retrievals

5.1.1 Validation of AOD retrievals against AERONET

Results show that Terra DT (Figure 2a), DB (Figure 2b), and DTB (Figure 2c) AOD retrievals are equally correlated \((r = 0.83)\) with AERONET AOD measurements, and have the same percentage of retrievals within the EE. However, the number of collocated observations for DTB (\( N = 2796 \)) is significantly higher than for DT (\( N = 1437 \)) and DB (\( N = 2486 \)) i.e., 94.6% and 12.5% higher than for DT and DB, individually. The retrievals from DT and DB showed significant overestimation (\( \text{RMB} = 17.17\% \)) and underestimation (\( \text{RMB} = -9.87\% \)) as 34.38% and 28.24% of the retrievals are
above (+EE) and below (-EE) the EE, respectively. These uncertainties appear to be averaged out in the DTB AOD retrievals, as the overestimations and underestimations are fewer than for DT and DB, individually. Furthermore, the RMB (-0.03%) is significantly improved, being 99.9% and 99.8% lower than for DT and DB, respectively. These results indicate the better performance of the Terra DTB AOD retrievals compared to DT and DB over Pakistan. Similar to Terra, the performance of the Aqua DTB AOD retrievals (Figure 2f) is much better than for DT (Figure 2d) and DB (Figure 2e) retrievals, with a significantly higher number of collocated retrievals and lower RMB. However, Aqua performs equally as Terra in terms of correlation and the percentage of retrievals within the EE. It is important to mention that a large number of both DT and DB AOD retrievals were available for Lahore than Karachi and also that DB provides a greater number of AOD retrievals than DT over Pakistan. Based on the superior performance of the Aqua and Terra DTB AOD retrievals, the merged AquaTerra DTB AOD product was generated for further analysis (see Figure S1 in the supplementary data for the validation of AquaTerra DTB AOD retrievals).
Figure 2: Validation of Terra and Aqua DT, DB, and DTB AOD retrievals against AERONET Version 3 Level 2.0 AOD measurements obtained for Lahore (for location, see no. 1 in Fig. 1a) and Karachi (for location, see no. 56 in Fig. 1a) AERONET sites from 2006 to 2017. Where, the red line represents the regression line, the solid black line represents the $y = x$ line, and the dashed black lines represent the upper and lower EE envelopes. The orange points represent AOD pairs at Karachi, the blue dots at Lahore.

5.1.2 Spatial distribution of AOD retrievals

Figure 3 shows the spatial distributions of the monthly mean AquaTerra retrieved DTB AOD as well as corresponding pixel counts (PC) from 2003 to 2017 over Pakistan. Significant monthly variations in both AOD and PC are observed. AOD retrievals are missing over the Gilgit-Baltistan and Jammu & Kashmir (disputed territory) throughout the year, except for January, as the DT and DB algorithms do not provide AOD retrievals over high mountain regions and snow surfaces. The presence of AOD retrievals during January is because the DB algorithm does not use the MODIS snow mask product directly, and the internal snow/cloud mask does not work well over these regions. Surprisingly, high AOD values > 1.0 are observed during June and July over the Northwestern region of Khyber Pakhtunkhwa, which is a high mountainous region with permanent snow cover. These high AOD values over snow-covered regions could be due to an error in the internal snow/cloud mask of the DB algorithm which has missed these pixels during preprocessing, although DT does discard bright pixels during preprocessing. AOD >1.0 is observed in July followed by June and August over Punjab and Sindh mainly attributed to hygroscopic growth of the aerosol particles during summer due to high relative humidity, similar to other
reports using MODIS and MISR aerosol products (Mehta et al. 2016; Mhawish et al. 2021). Most of the major cities of Punjab and Sindh are surrounded by cropland, and the results show that high AOD over Pakistan follows the same spatial pattern as that of the cropland. Significantly higher AOD is observed over cropland than over non-agricultural (i.e., mainly desert) regions throughout the year, even, and especially, during late spring and summer when dust storms are considered a major source of aerosols over Punjab and Sindh. Local anthropogenic aerosol pollutants from urban and industrial emission and agricultural pre- and post-harvest burning may be responsible for the high pollution levels over the region. Over Balochistan, especially over the desert areas, the AOD is low as compared to Punjab and Sindh, but still higher than over other administrative units. Over Punjab the highest AOD values are observed during the post-harvest seasons, i.e., throughout September to November, peaking in November, probably due to biomass (crop residue) burning activities (Jethva et al. 2019; Mhawish et al. 2021). However, if the high AOD levels are only due to local aerosols, then the spatial pattern during each month should be almost the same, but it is not. Therefore, the transport of transboundary aerosols may contribute to Pakistan’s deteriorating air quality. This is confirmed by the well-known smog episodes, occurring every year over Punjab, due to local as well as cross-border crop residue burning activities, which reduce atmospheric visibility to a few meters in urban as well as in rural areas. Overall, much higher AOD levels were observed in Pakistan during June, July, and August (summer) followed by September, October, and November (autumn), March, April, and May (spring), and December, January, and February (winter). The higher AOD retrieval in summer is attributed to several reasons, including (i) hygroscopic growth of aerosol particles due to high relative humidity that increases the extinction efficiency of the atmospheric aerosols (Dickerson
et al. 1997; Li and Wang 2014), (ii) the enhancement of secondary aerosol formation rate due to the enhancement of photochemical reaction under higher temperature (Jacob and Winner 2009; Kulmala et al. 2020), and (iii) the higher contribution of natural aerosols mainly dust during the summer monsoon (Mhawish et al. 2021).

Figure 3 shows a distinct pattern of PC which suggests that the DT and DB algorithms do not perform equally temporally or spatially. For example, between 2003 to 2017, from late spring to early autumn, a large number of AOD retrievals (> 400) per pixel are available over Balochistan and some parts of Punjab, and from late autumn to early spring, a large number of AOD retrievals (> 400) per pixel are available over Sindh and some parts of Punjab. This could be due to the initial preprocessing of the inversion methods or the presence of cloud cover. Only October provides favorable conditions to both the DT and DB algorithms, when more than 400 AOD retrievals are available over Pakistan, except for Gilgit-Baltistan and disputed areas, due to algorithm limitations for snow/ice surfaces.
Figure 3: Monthly mean spatial distributions of AquaTerra DTB AOD$_{550}$ and the total number of corresponding Pixel Counts (PC) over Pakistan from 2003 to 2017. Where, GB = Gilgit-Baltistan, AK = Azad Kashmir, KP = Khyber Pakhtunkhwa, PJ = Punjab, BL = Balochistan, and SN = Sindh.

5.1.3 Characterization of extremely polluted cities using MODIS data

Figure 4a shows the mean AOD$_{550}$ retrievals for 80 cities (Figure 1) obtained from the annual mean AquaTerra DTB AOD$_{550}$ images and categorizes the extremely polluted to polluted cities. The thresholds for polluted and extremely polluted cities are defined based on the values of first (Q1) and third (Q3) quartiles respectively, and these quartiles are calculated by analyzing descriptive statistics (Table S1) for the AOD values extracted for 80 cities. Highly polluted cities are defined based on the AOD range between the first and third quartiles. For example, AOD < 0.3 (1$^{\text{st}}$ quartile) represents polluted cities, AOD > 0.6 represents extremely polluted cities (3$^{\text{rd}}$ quartile), and 0.3 ≤ AOD ≤ 0.6 (between 1$^{\text{st}}$ and 3$^{\text{rd}}$ quartiles) represents highly polluted cities. A total of 21 cities fall within the category of extremely polluted cities (Punjab: 12, Sindh: 7, and Balochistan: 2), 35 cities in the category of moderately polluted cities (Punjab: 11, Sindh 7, Balochistan: 7, Khyber Pakhtunkhwa: 8, Azad Kashmir: 2), and 24 cities in the category of low polluted cities (Punjab: 0, Sindh 0, Balochistan: 10, Khyber Pakhtunkhwa: 11, Azad Kashmir: 3).

The top 3 polluted cities are Jhang, Multan, and Vehari in Punjab, as Punjab is the most urbanized and populated administrative unit (Figures 1b and 4a), with more vehicles and industries, and also faces severe smog episodes and dust storms, resulting in extremely high AOD levels over the region. Along with local anthropogenic aerosols from cropland, urban and industrial emissions, regional transported pollutants may be responsible for Punjab’s severe air pollution problems,
and this can be investigated using the PSCF analysis based on the HYSPLIT air parcel back trajectory analysis and BAM PM$_{2.5}$ concentrations (see section 5.7).

Figure 4b shows the pixel counts (PC) of the daily AOD retrievals for each city from 2003 to 2017. Results show a large number of PC for most cities, indicating that the characterization of extremely polluted to polluted cities is based on a large number of PC, which supports the results (Figure 4a), and demonstrates confidence in the merged AquaTerra DTB AOD retrievals for quantitative research applications over Pakistan. However, the lowest number of PC is observed for the coastal (Ormara and Gwadar) and mountainous (Dir) cities, where the inversion scheme of both the DT and DB algorithms needs to be improved.

The monthly mean AOD retrievals are plotted to identify the high and low polluted months in Pakistan (Figure 4c). The months of June, July, and August are by far the most polluted, with AOD > 1.20 for extremely polluted cities. A similar pattern of monthly variation in AOD is observed for all other cities, though at lower pollution levels. As mentioned in section 5.1.2, these months may be affected by aerosol pollutants from local sources such as agricultural land, urban and industrial regions, and deserts. Figure 4d, showing inter-annual variations, indicates very high AOD levels for extremely polluted cities throughout the last two decades, with annual mean AOD > 0.60, and with the most polluted years being 2004, 2006, 2008, 2016, and 2017.
Figure 4: Characterization of extremely polluted to polluted cities in Pakistan using AquaTerra DTB AOD$_{550}$ retrievals from 2003 to 2017. (a) polluted cities based on mean AOD, (b) pixel counts, (c) polluted months based on mean AOD, and (d) polluted years based on mean AOD.

5.2 CAMS and MERRA-2 reanalysis data

5.2.1 Validation of PM$_{2.5}$ reanalysis data

Previous studies have evaluated the uncertainties in both CAMS and MERRA-2 PM$_{2.5}$ reanalysis data compared to ground-based PM$_{2.5}$ measurements (Cuevas et al. 2015; He et al. 2019; Song et al. 2018; Ukhov et al. 2020). Recently, Ukhov et al. (2020) reported overestimation in CAMS PM$_{2.5}$ over the middle east and west Asia which have been attributed to the deficient size distribution of the emitted dust. Additionally, significant underestimation in MERRA-2 PM$_{2.5}$ was reported over China and India (He et al. 2019; Navinya et al. 2020; Song et al. 2018) which could be due to the lack of nitrate concentrations in the reanalysis data and underestimation of OC emission for urban/suburban areas (Buchard et al. 2016; Provencal et al. 2017).

The results show significant underestimation in both daily (Figures 5a and 5b) and monthly (Figures 5e and 5f) MERRA-2 PM$_{2.5}$ concentrations compared to both BAM and LCM PM$_{2.5}$ measurements with a slope from 0.45 to 0.54 (daily) and 0.30 to 0.52 (monthly), and RMB from -34.2% to -26.8% (daily) and -35.9% to 25.3%. The results also show a weak correlation of MERRA-2 PM$_{2.5}$ data with both BAM and LCM daily ($r = 0.22$–$0.25$) and monthly ($r = 0.10$–$0.27$) PM$_{2.5}$ data. The weak correlation suggests that MERRA-2 PM$_{2.5}$ data based on the GOCART aerosol module is unable to accurately reproduce the temporal variations in PM$_{2.5}$. The significant underestimation
in MERRA-2 PM$_{2.5}$ data over Pakistan is similar to that reports over China (He et al. 2019; Song et al. 2018) and India (Navinya et al. 2020), however, a much weaker correlation was observed than in these regions.

In comparison to the MERRA-2 data, CAMS daily (Figures 5c and 5d) and monthly (Figures 5g and 5h) PM$_{2.5}$ data show a good correlation with both BAM and LCM PM$_{2.5}$ measurements. However, significant over- and under-estimations are observed compared to the ground-based PM$_{2.5}$ measurements. For example, CAMS PM$_{2.5}$ data at 0.75° grid size show a significant overestimation of 30.4% to 55.4% for daily and 30.4% to 57.4% for monthly data. On the other hand, CAMS data at 0.125° grid size show underestimation and overestimation for BAM and LCM measurements, respectively. These results suggest that grid size and ground-based PM$_{2.5}$ measurement methods (e.g., BAM and LCM) play an important role in the overestimation/underestimation of CAMS PM$_{2.5}$ data. For example, against the BAM PM$_{2.5}$ measurements, CAMS data at different grid size show overestimation (0.75° grid) and underestimation (0.125° grid), and CAMS data at the same grid size (0.125°) show underestimation and overestimation against two different ground-based PM$_{2.5}$ measurement methods i.e., BAM and LCM, respectively. It is worth mentioning that both MERRA-2 and CAMS simulate 5 types of fine particulate matter (dust, sea salt, sulfate, organic carbon, and black carbon), without simulating nitrate concentrations. If the lack of nitrate concentrations is the main reason for underestimation in MERRA PM$_{2.5}$ data, as reported by previous studies (Buchard et al. 2016; He et al. 2019; Provencal et al. 2017; Song et al. 2018), then underestimation should be observed in CAMS PM$_{2.5}$ data at 0.75° grid size, but this is not the case. Therefore, the exact reasons for underestimation in both MERRA-2 and CAMS as well as overestimation in CAMS data...
should be thoroughly investigated in future studies. The results show a higher correlation for CAMS monthly data (Figures 5g and 5h) compared to the daily data (Figures 5c and 5d). Although CAMS monthly data at 0.75° grid size show overestimation, they have a good correlation coefficient ($r = 0.72–0.76$) with ground-based PM$_{2.5}$ measurements and could be useful for characterizing pollution levels in the cities of Pakistan compared to the MERRA-2.

Figure 5: Validation of MERRA-2 and CAMS PM$_{2.5}$ reanalysis data against BAM (beta gauge attenuation monitor) PM$_{2.5}$ concentrations for 2019-2020 provided by the US Consulates and LCM (low-cost monitor) PM$_{2.5}$ concentrations for 2018-2019 provided by PAQi. Where, (a) MERRA-2 daily PM$_{2.5}$ vs. BAM daily PM$_{2.5}$, (b) MERRA-2 daily PM$_{2.5}$ vs. LCM daily PM$_{2.5}$, (c) CAMS daily PM$_{2.5}$ vs. BAM daily PM$_{2.5}$, (d) CAMS daily PM$_{2.5}$ vs. LCM daily PM$_{2.5}$, (e) MERRA-2 monthly PM$_{2.5}$ vs. BAM monthly PM$_{2.5}$, (f) MERRA-2 monthly PM$_{2.5}$ vs. LCM monthly PM$_{2.5}$, (g) CMAS monthly PM$_{2.5}$ vs. BAM monthly PM$_{2.5}$, and (h) CAMS monthly PM$_{2.5}$ vs. LCM monthly PM$_{2.5}$.
5.2.2 Characterization of extremely polluted cities using PM$_1$ and PM$_{2.5}$ concentrations

PM$_1$ and PM$_{2.5}$ are fine particulate matter associated with human health issues. PM$_1$ is more harmful than PM$_{2.5}$ as it can reach deeper into the lungs and affect the respiratory system (Liu et al. 2013; Meng et al. 2013). A previous study over China reported that most health issues associated with PM$_{2.5}$ were mainly due to greater contributions of PM$_1$ in PM$_{2.5}$ (Chen et al. 2017).

Figure 6a indicates that the top 10 extremely polluted cities ranked according to PM$_1$ concentrations are Lahore (135.44 µg/m$^3$), Gujranwala (131.99 µg/m$^3$), Okara (107.72 µg/m$^3$), Faisalabad (98.96 µg/m$^3$), Pakpattan (94.06 µg/m$^3$), Jhelum (85.51 µg/m$^3$), Sargodha (84.30 µg/m$^3$), Bhimber (83.99 µg/m$^3$), Gujrat (83.99 µg/m$^3$), and Sialkot (83.99 µg/m$^3$). Similarly, the top 10 extremely polluted cities (Figure 7a) ranked according to PM$_{2.5}$ concentrations are Lahore (170.53 µg/m$^3$), Gujranwala (163.63 µg/m$^3$), Okara (139.43 µg/m$^3$), Faisalabad (129.85 µg/m$^3$), Pakpattan (126.97 µg/m$^3$), Multan (113.09 µg/m$^3$), Bahawalnagar (110.81 µg/m$^3$), Vehari (110.81 µg/m$^3$), Sargodha (109.81 µg/m$^3$), and Jhelum (107.68 µg/m$^3$). The WHO air quality guidelines (AQG) are not yet defined for PM$_1$ as PM$_1$ is not widely monitored as PM$_{2.5}$, therefore the WHO recommended AQG for PM$_{2.5}$ (<10 µg/m$^3$ annual mean) and Pak-NEQS for PM$_{2.5}$ (<15 µg/m$^3$ annual mean) are used for comparison purposes. Not a single city in Pakistan falls within the PM$_{2.5}$ standards defined by Pak-NEQS and WHO, and the values of PM$_1$ and PM$_{2.5}$ respectively for the top 10 cities are 5.6 (8.4) to 9.0 (13.5) times and 7.2 (10.8) to 11.4 (17.1) times greater than the Pak-NEQS (WHO AQG). For PM$_1$ and PM$_{2.5}$, 9 out of 10, and 10 out of 10 cities respectively, are in Punjab. The extremely high pollution level may be due to local anthropogenic activities, confirming the results of a previous modeling study which suggested local anthropogenic activities as the major cause of high particulate concentrations in Pakistan (Shi et al. 2020).
entire country of Pakistan including all cities is exposed to long-term PM$_{2.5}$ concentrations (Figures 1a and 7a), which exceed the Pak-NEQS ($<15 \mu g/m^3$) and 68, 73, and 80, out of 80 cities exceeded the WHO Interim Target-1 ($<35 \mu g/m^3$), Target-2 ($<25 \mu g/m^3$), and Target-3 ($<15 \mu g/m^3$), respectively. These exceedances are set in stark perspective against the much lower recommended WHO AQG for PM$_{2.5}$ of 10 µg/m$^3$. These results suggest that the top polluted cities are extremely hazardous for human health, as an increase of PM$_{2.5}$ by 10 µg/m$^3$ can increase mortality, lung cancer, and cardiopulmonary diseases by 8%, 6%, and 4%, respectively, due to long-term exposure to fine particulates (Pope et al. 2002).
Figure 6: Ranking of extremely polluted to polluted cities in Pakistan using annual mean CAMS PM$_1$ concentrations from 2003 to 2020. Where (a) polluted cities based on yearly mean PM$_1$, (b) polluted months based on PM$_1$, and (c) polluted years based on yearly mean PM$_1$.

Figures 6b and 7b show months with the highest levels of PM$_1$ and PM$_{2.5}$ for the extremely polluted cities. The higher PM$_1$ and PM$_{2.5}$ concentrations were observed in cold months (October to February) with the maximum concentrations in December and January, while warmer months
(March to September) showed lower PM$_x$ concentrations. The high levels of fine particulates in October and November may be attributed to both cross-border biomass burning activities (from India) as well as local anthropogenic activities. As the highest values of fine particulates were observed in December and January which are not the main months of biomass burning activities, these are not likely to be the main source of the high levels of fine particulates pervasive across these high polluted cities. At this time of year, less surface heating and less turbulence due to lower intensity of solar irradiation led to stable and shallow BLH (boundary layer height). Furthermore, with higher concentrations of light-absorbing aerosols, mainly BC, the atmospheric stability increases due to the heating of the upper boundary layer, induced by BC that further lowers the BLH (Ding et al. 2016). Stable atmospheric conditions that imply low BLH and low wind speed are conducive to higher aerosol concentrations and accumulation of particulates near the surface, thus anthropogenic emissions such as fossil fuel combustion, and urban and industrial emissions may linger for long periods (Mhawish et al. 2020). In October and November, internal as well as external (cross-border) biomass (crop residue) burning activities coupled with stable atmospheric conditions have been recognized to cause severe haze and smog episodes, especially over Punjab (Mhawish et al. 2020; Tariq et al. 2015; Tariq et al. 2016). The formation of secondary inorganic aerosol during the haze episodes is also responsible for higher PM$_{2.5}$ concentrations as reported by recent studies over China (Nichol et al. 2020; Zhang et al. 2018).

An increase in PM$_{2.5}$ concentrations was observed in June and July, and PM$_1$ concentrations slightly increased in July. This means that PM$_{2.5}$ exhibited two peaks: the first in winter and the second in summer, whereas a single peak in winter was observed for PM$_1$. The second PM$_{2.5}$ peak in summer may be attributed to the fine particulates from dust, as dust storm activities are very
common in Pakistan during summer, as well as local anthropogenic activities. The lower peak of
PM$_{2.5}$ in the summer, compared to winter, may be due to the unstable atmospheric conditions
due to the higher surface heating by solar irradiation, leading to thermal radiation and strong
mixing conditions which promotes the vertical dispersion of pollutants.

The annual mean concentrations of PM$_1$ (Figure 6c) and PM$_{2.5}$ (Figure 7c) show strong inter-
annual variations with distinct PM$_x$ levels and very poor air quality conditions throughout the last
two decades. The annual mean of PM1 in extremely polluted cities ranges from 63 µg/m$^3$ to
150.19 µg/m$^3$ and from 85 µg/m$^3$ to 187.35 µg/m$^3$ which found 4.2 (6.3)–10 (15) and 5.7 (8.5)–
12.5 (18.7) times greater than the Pak-NEQS (WHO AQG), respectively.
Figure 7: Ranking of extremely polluted to polluted cities in Pakistan using annual mean CAMS PM$_{2.5}$ concentrations from 2003 to 2020. (a) polluted cities based on yearly mean PM$_{2.5}$, (b) polluted months based on PM$_{2.5}$, and (c) polluted years based yearly mean PM$_{2.5}$. Extremely polluted cities are defined for PM$_{2.5}$ > 92.84 (3$^{rd}$ quartile), highly polluted cities for 45.69 ≤ PM$_{2.5}$ ≤ 92.84 (between 3$^{rd}$ and 1$^{st}$ quartiles), and polluted cities for PM$_{2.5}$ < 45.69 (1$^{st}$ quartile) using descriptive statistics (Table S1). Cities are not defined as low polluted or clean cities as annual
mean PM$_{2.5}$ concentrations for all the cities exceed Pakistan’s National Environmental Quality Standards (Pak-NEQS) for ambient air (<15 µg/m$^3$ annual mean).

5.2.3 Characterization of extremely polluted cities using PM$_{10}$ concentrations

Figure 8a shows the ranking of polluted cities based on PM$_{10}$ concentrations. The PM$_{10}$ fraction with an aerodynamic diameter larger than PM$_{2.5}$ (PM$_{10}$-PM$_{2.5}$) mainly originates from natural sources such as desert dust and resuspended soil particles. The top 10 most polluted cities according to the concentrations of coarse particulates (PM$_{10}$) are Lahore (238.9 µg/m$^3$), Gujranwala (229.1 µg/m$^3$), Okara (194.5 µg/m$^3$), Faisalabad (180.6 µg/m$^3$), Pakpattan (177.9 µg/m$^3$), Bahawalnagar (160.6 µg/m$^3$), Vehari (160.6 µg/m$^3$), Multan (157.5 µg/m$^3$), Sargodha (152.3 µg/m$^3$), and Jhelum (149.7 µg/m$^3$). PM$_{10}$ concentrations are 1.2 to 11.9 times higher than WHO AQG for PM$_{10}$ (20 µg/m$^3$ annual mean) for all the cities shown in Figure 8a, suggesting that very poor air quality conditions hazardous for human life prevail in all Pakistani cities. Overall, the PM$_{10}$ temporal trend pattern is very similar to that for PM$_{2.5}$, i.e., December is the month under the greater influence of coarse particles (PM$_{10}$), followed by January, and in summer, July is the most polluted month followed by June (Figure 8b). Similar to the PM$_{2.5}$ variations, PM$_{10}$ also exhibited peaks in both winter and summer. The higher concentrations during December and January may be due to increased anthropogenic emission activities along with stable atmospheric conditions in winter. The higher concentrations during July and June may be due to the frequently occurring dust/sand storms over Pakistan. The pre-harvest, harvesting, and post-harvest burning activities may contribute to higher PM$_{10}$ levels especially during October and
November as these activities produce both fine (PM$_1$ and PM$_{2.5}$) and coarse (PM$_{10}$) particles as reported by Le Blond et al. (2017) over South American countries.

Similar to the annual mean PM$_{2.5}$ variations (Figure 7c), the annual mean PM$_{10}$ concentrations also show distinct interannual variations for all cities (Figure 8c), and severe air pollution levels were observed throughout the last two decades. According to these findings, Pakistani people are not only exposed to long-term PM$_{2.5}$ but also to PM$_{10}$ concentrations exceeding the WHO recommended AQG for PM$_{10}$ (<10 µg/m$^3$). Overall, these results suggested that Pakistani cities are a severe threat to human life due to extremely poor air quality conditions.
Figure 8: Ranking of extremely polluted to polluted cities in Pakistan using annual mean CAMS PM$_{10}$ concentrations for the years 2018 and 2019. (a) polluted cities based on yearly mean PM$_{10}$, (b) polluted months based on PM$_{10}$, and polluted years based annual mean PM$_{10}$. 
5.2.4 PM$_1$/PM$_{2.5}$ and PM$_{2.5}$/PM$_{10}$ ratios

The PM$_x$ ratios are very useful for understanding the contributions among particulate types, as revealed by a study in China where PM$_1$ contributed nearly 80% of PM$_{2.5}$ (Wang et al. 2015), which would have consequences for human health. Over Pakistan, the PM$_1$/PM$_{2.5}$ (Figure 9a) and PM$_{2.5}$/PM$_{10}$ (Figure 9b) ratios are lower than those observed over China (Wang et al. 2015), indicating lower contributions of PM$_1$ to PM$_{2.5}$ and PM$_{2.5}$ to PM$_{10}$. However, the pattern of ratios is similar to that observed for China, i.e., the PM$_1$/PM$_{2.5}$ ratios are higher than PM$_{2.5}$/PM$_{10}$ ratios.

Relatively higher PM$_1$/PM$_{2.5}$ ratios are observed from October to March (Figure 9a), indicating a larger fraction of PM$_1$ in PM$_{2.5}$ due to more anthropogenic activities. The directly emitted PM$_1$ from the automobile and combustion of fossil fuel, and indirectly by formation from precursor gases, are most likely higher from October to March, leading to the enhanced PM$_1$/PM$_{2.5}$ ratio. However, low PM$_1$/PM$_{2.5}$ ratios are observed from April to September in most of the cities, and low ratios during all month are observed in the cities located in Balochistan, indicating a lower contribution of PM$_1$ to PM$_{2.5}$, which is mainly dominated by coarse particles especially during summer (June, July, and August).

Figure 9b shows large contributions of PM$_{2.5}$ to PM$_{10}$ throughout the year with maximum contributions during summer as indicated by the large PM$_{2.5}$/PM$_{10}$ ratios. This suggests that the air quality of these cities is mainly and significantly influenced by fine particulates, largely from anthropogenic sources. The higher PM$_{2.5}$/PM$_{10}$ ratios (Figure 9b) throughout the year also suggested that Gwadar and Ormara, coastal cities in Balochistan, are dominated by PM$_{2.5}$ particles, which indicates that their air quality is influenced by both anthropogenic and natural
sources (dust/sand particles and sea spray aerosols). Gwadar has the deepest seaport in the world and the ship-based emissions may be one of the sources of fine anthropogenic particles throughout the year. However, lower PM$_{2.5}$/PM$_{10}$ ratios are observed for other cities located in Balochistan, indicating the greater influence of coarse particulates (mainly desert dust).

![Figure 9: (a) Monthly mean ratios of PM$_1$/PM$_{2.5}$ and (b) PM$_{2.5}$/PM$_{10}$.

5.2.5 PM$_1$ vs. PM$_{2.5}$ and PM$_{2.5}$ vs. PM$_{10}$

Scatter plots of PM$_1$ vs. PM$_{2.5}$ (Figure 10a) and PM$_{2.5}$ vs. PM$_{10}$ (Figure 10b) show that the PM$_x$ fractions over Pakistan are well-correlated, with Pearson’s correlation coefficients (r) of 0.95 and 0.99, and slopes of 0.90 and 0.70, respectively. The high values of r suggest common sources of
fine and coarse particulates, but slope values suggest larger contributions of PM$_1$ to PM$_{2.5}$ than PM$_{2.5}$ to PM$_{10}$. Overall, both PM$_1$ and PM$_{2.5}$ contribute less to PM$_{2.5}$ and PM$_{10}$, respectively, over Pakistan than over (Wang et al. 2015) as indicated by the PM$_x$ ratios (Figure 9) and slope values (Figure 10), which suggests higher anthropogenic activities in China than in Pakistan. Figures 10a and 10b show some scattered points, within a red circle or ellipse, which represent the data from May to September and these scattered points suggest lower contributions of PM$_1$ in PM$_{2.5}$ and PM$_{2.5}$ in PM$_{10}$, as also indicated by low PM$_x$ ratios (Figure 9).

Figure 10: Scatter plots between (a) PM$_1$ vs. PM$_{2.5}$ and (b) PM$_{2.5}$ vs. PM$_{10}$. Where the red solid line represents the regression line and the black dashed line represents the y = x line. The data points in the red circle and ellipse are explained in the text.

5.2.4 Monthly mean temporal trend of PM$_1$, PM$_{2.5}$, and PM$_{10}$

To investigate the proportions of local and regional contributions to PM$_x$ concentrations, the month-to-month variations of the monthly mean PM$_1$, PM$_{2.5}$, and PM$_{10}$ concentrations for the
The top 10 polluted cities are shown in Figure 11. These cities vary according to population growth, the number of automobiles, urbanization, industrialization, city size, land cover types, and climatic conditions, and PM concentrations are expected to behave differently due to these factors. This study follows the hypothesis of our previous study conducted over Hong Kong (Bilal et al. 2019c) i.e., if the PM concentrations have different magnitudes but follow the same temporal pattern at different locations, they are influenced by local as well as regional contributions. Thus for PM$_1$ concentrations, Figure 11a shows the same pattern for all the different cities, suggesting that both local and regional sources contribute to PM$_1$ concentrations. For both PM$_{2.5}$ (Figure 11b) and PM$_{10}$ (Figure 11c), similar patterns are only evident from September to April, and dissimilar patterns due to variation in magnitudes are evident from May to August, suggesting more local contributions for the summer months of May to August. This local contribution during summer may be attributed to the frequent dust/sand storms. Similarly, PM$_1$, PM$_{2.5}$, and PM$_{10}$ concentrations for Lahore and Gujranwala show the same pattern but with higher magnitudes than other cities from October to January, probably due to Lahore and Gujranwala being the largest cities, with consequently more transport, fossil fuel, and industrial emissions, and some local and cross-border biomass burning activities in autumn (Ali et al. 2013; Tariq et al. 2015; Tariq et al. 2016) along with meteorological impact including shallower boundary layer height and lower wind speed that enhance the accumulation of particulate matter near the surface (Miao et al. 2019; Miao and Liu 2019; Miao et al. 2018; Qu et al. 2017; Sun et al. 2019; Wang et al. 2018).
Figure 11: Monthly variations of PM$_1$, PM$_{2.5}$, and PM$_{10}$ concentrations in the corresponding top 10 polluted cities (see legend). Cities are plotted with the rank of high to low polluted.
5.3 OMI atmospheric trace gases

5.3.1 Characterization of extremely polluted cities using NO₂ data

Tropospheric vertical column density (TVCD) of NO₂ is mainly derived from fossil fuel combustion, industrial emission, automobile emission, biomass burning, natural lightning, and soil microbe emissions (Cheng et al. 2012; Lee et al. 1997; Olivier et al. 1998; Richter and Burrows 2002). It contributes to adverse health, low atmospheric visibility, and poor air quality conditions (Khokhar et al. 2015; ul–Haq et al. 2014). Pakistan’s top ten polluted cities due to NO₂ are those with the highest levels of urbanization, vehicle emissions, and industrialization, suggesting anthropogenic activities to be the major cause. They are Lahore (5.69×10¹⁵ molecules/cm²), Rawalpindi (3.65×10¹⁵ molecules/cm²), Islamabad (3.65×10¹⁵ molecules/cm²), Karachi (3.60×10¹⁵ molecules/cm²), Gujranwala (3.32×10¹⁵ molecules/cm²), Sialkot (2.81×10¹⁵ molecules/cm²), Haripur (2.73×10¹⁵ molecules/cm²), Okara (2.72×10¹⁵ molecules/cm²), Faisalabad (2.72×10¹⁵ molecules/cm²), and Gujrat (2.47×10¹⁵ molecules/cm²) (Figure 12a). Similar results are reported by Tabinda et al. (2019), Ashraf et al. (2013), and Khanum et al. (2017). In terms of data availability from OMI, Figure 12b indicates the largest number of PC available for Lasbela (4168), Awaran (4154), and Panjgur (4140), all located in Balochistan. On a monthly mean basis, NO₂ (Figure 12c) follows the same patterns as observed for PM₁ and PM₂.₅ concentrations; i.e., higher values in winter especially for the extremely polluted cities (Lahore, Rawalpindi, Islamabad, and Karachi), where the higher values are attributed to emissions of automobiles, industries, and fossil fuel combustion, under stable atmospheric conditions. A different trend observed over Balochistan, with higher NO₂ in summer, could be due to natural lightning as reported by Khokhar et al. (2015).
Figure 12d shows that Lahore, Rawalpindi, Islamabad, and Karachi are polluted in all years from 2004 to 2019, subjecting citizens to long-term exposure associated with respiratory diseases, otitis media, and mortality (Latza et al. 2009).
Figure 12: Ranking of extremely polluted to polluted cities in Pakistan using OMI NO₂ TVCDs (molecules/cm²) from 2004 to 2019. (a) polluted cities based on mean NO₂, (b) pixel counts, (c) polluted months based on mean NO₂, and (d) polluted years based on mean NO₂.

5.3.2 Characterization of extremely polluted cities using SO₂ data

Thermal power plants, oil and gas refineries, and metal smelters are the major sources of anthropogenic SO₂ (Dahiya and Myllyvirta 2019). In Figure 13a, extremely polluted to polluted cities are ranked based on OMI-derived SO₂ vertical column density and the top 10 polluted cities are Lahore (10.6×10¹⁵ molecules/cm²), Mirpur (10.5×10¹⁵ molecules/cm²), Gujranwala (10.3×10¹⁵ molecules/cm²), Rawalpindi (10.3×10¹⁵ molecules/cm²), Islamabad (10.3×10¹⁵ molecules/cm²), Sialkot (10.3×10¹⁵ molecules/cm²), Gujrat (10.3×10¹⁵ molecules/cm²), Faisalabad (10.3×10¹⁵ molecules/cm²), Bhimber (10.2×10¹⁵ molecules/cm²), and Jhelum (10.2×10¹⁵ molecules/cm²).

According to the global SO₂ emission hotspot database (Dahiya and Myllyvirta 2019), five oil power plants near Lahore are the main sources of high SO₂ emissions over Lahore. The lower numbers (1080–2520) of successful SO₂ retrievals (Figure 13b) are attributed to the high noise level in retrieving PBL SO₂ that separating the weak SO₂ signal from the noise in radiances is challenging. Therefore, only the stronger SO₂ signal over point sources (e.g., power plants, metal smelters) can be detected, which leads to fewer successful SO₂ retrievals (Fioletov et al. 2011; Li et al. 2017; Li et al. 2020). Temporally, the monthly mean SO₂ (Figure 13C) shows the same pattern as PM₂.₅ and NO₂, being higher in winter. For the top polluted cities, the high SO₂ observed during November, December, and January may be attributed to the power plants and brick kilns (Dahiya and Myllyvirta 2019; Rahman et al. 2000). Like other anthropogenic sources
of SO$_2$ emissions, brick kilns are also considered as the main sources of SO$_2$ and severely bad air quality especially over Punjab (Adrees et al. 2016; Colbeck et al. 2010; Pervaiz et al. 2021; Ur Rehman et al. 2019). Therefore, every year during late autumn and winter, the Government of Pakistan bans these kilns to control pollution levels. The higher SO$_2$ may also be attributed to the stable atmospheric conditions and shallow BLH as this time of year, as less surface heating and less turbulence result from the lower solar radiation. Unlike NO$_2$, SO$_2$ pollutants do not contribute significantly every year to poor air quality in Pakistani cities, as Figure 13d shows particularly high SO$_2$ for the years 2004, 2008, and 2011, and investigation of this would require a separate study.
Figure 13: Ranking of high to low polluted cities in Pakistan in terms of OMI SO\textsubscript{2} VCDs (molecules/cm\textsuperscript{2}) from 2004 to 2019. (a) polluted cities based on mean SO\textsubscript{2}, (b) pixel counts, (c) polluted months based on mean SO\textsubscript{2}, and (d) polluted years based on mean SO\textsubscript{2}.

5.4 Spatial distributions of atmospheric pollutants and trace gases

The purpose of this section is to link the spatial distributions of long-term atmospheric pollutants and trace gases with each other as well as with population density, nighttime lights, land cover types (cropland and urban areas), and presumed vegetation fire activities. Here, the PM\textsubscript{x} data are interpolated using cubic convolution (Keys 1981) from 0.75° grid size to 0.125° grid size to better show smooth spatial distributions over different administrative units. The spatial distributions (Figure 14) showed that Punjab is the most severely affected and most polluted region of Pakistan, followed by Sindh. It is significant that other environmental data including population density (Figure 14g), VIIRS nighttime lights (Figure 14h), cropland (Figure 14i), and vegetation fires (Figure 14j) show similar spatial patterns. It is obvious that vegetation fires would have the same spatial pattern as cropland, but not obvious that population density and nighttime lights would have the same pattern. As nighttime lights and vegetation fires represent human activities, having the same spatial patterns suggests that the majority of human settlements including urban, suburban and, industrial regions, are inter-mixed with cropland. Interestingly, these coincident spatial distributions (population, nighttime lights, land cover, and fires) correspond to the higher ranges of pollutants i.e., AOD > 0.4, PM\textsubscript{1} > 20 µg/m\textsuperscript{3}, PM\textsubscript{2.5} > 40 µg/m\textsuperscript{3}, PM\textsubscript{10} > 60 µg/m\textsuperscript{3}, NO\textsubscript{2} > 1.0×10\textsuperscript{15} molecules/cm\textsuperscript{2}, and SO\textsubscript{2} > 6.5×10\textsuperscript{15} molecules/cm\textsuperscript{2}. These results suggested that atmospheric pollutants and trace gases are mainly emitted from local
anthropogenic sources such as power plants, oil and gas refineries, vehicular emissions, crop residue burning, and industrial activities including construction, manufacturing of cement, ceramic, and bricks, and metals smelting. These anthropogenic sources are mainly responsible for NO$_2$, SO$_2$, and PM$_x$ pollutants (Adrees et al. 2016; Shah et al. 2012; Ur Rehman et al. 2019). Among these anthropogenic sources, brick kilns industries are considered a major source. Small-scale traditional brick kilns, located in rural and suburban areas, produce large amounts of gaseous pollutants (NO$_2$, SO$_2$, O$_3$, and CO) and PM$_x$ due to the usage of low-quality fuels including coal, oil, wood, rice straw, rice husk, rubber tires, bagasse, and corncobs (Adrees et al. 2016; Ishaq et al. 2010). Besides this, the combustion of agricultural biomass and crop residue burning are also contributing to deteriorating rural and urban air quality (Irfan et al. 2015; Irfan et al. 2014). Irfan et al. (2015) reported that Punjab produced more aerosol pollutants than Sindh from crop residue burning and among the crop residues, wheat straw is the main contributor of NO$_x$, SO$_2$, CO$_2$, and CO. Pakistan’s 23.6 million vehicles emitted 58% of the country’s total NO$_2$ emission and 34% is emitted by power plants and industries (Amnesty International 2019; Government of Pakistan 2019; UNDP 2019). Another important source of aerosol pollutants, missed by previous studies, is the burning of solid waste and street garbage which is a common practice in Pakistan, even in major urban cities such as Islamabad, Lahore, Rawalpindi, Faisalabad, Gujranwala, Okara, etc. To support this statement, some illustrations with references are provided in the supplementary data (Figure S2). Figures 14a to 14d show that deserts (see Figure 1 for locations) are also another source of increasing AOD and PM$_x$ levels in Pakistan. Although local anthropogenic activities are the main source of aerosol pollutants and severe air quality problems in Pakistan, transboundary transport of aerosols may also influence Pakistan’s air
quality. To investigate contributions of transboundary transport results from PSCF analyses, integrated with HYSPLIT backward trajectory analysis and ground-based PM$_{2.5}$ measurements, are presented in section 5.7.
Figure 14: Spatial distributions of yearly mean (a) AOD [2003–2017] (b) PM$_1$ [2003–2020] (c) PM$_{2.5}$ [2003–2020], (d) PM$_{10}$ [2003–2020], (e) NO$_2$ [2005–2019], (f) SO$_2$ [2005–2019], (g) Population density [2019], (h) VIIRS Nighttime Lights [2012–2019], (i) Land cover types [2019], and (j) Presumed vegetation fire data [2003–2020].

5.5 Relationship of PM$_x$ with AOD, NO$_2$, and SO$_2$

AOD provides valuable information about the aerosol loading in the atmospheric column, while the PM$_x$ represents the aerosol concentrations near the ground. This section assesses how well do satellite-based AOD describes PM$_1$, PM$_{2.5}$, and PM$_{10}$ by examining the monthly correlation between AOD vs. PM$_x$. We have also examined the monthly correlation between PM$_x$ and the main two trace gases, SO$_2$ and NO$_2$ to understand the common sources and the monthly contribution of secondary aerosols to the PM$_x$ concentration. The relationship between AOD vs. PM$_x$ varies spatially and temporally and influenced by several factors mainly meteorological variables and boundary layer height (Li et al. 2016; Mhawish et al. 2021). The relationship between AOD vs. PM$_x$ shows a higher correlation coefficient from October to January under stable atmospheric conditions and shallower boundary layer height (Figure 15a). This indicates that the majorities of aerosol are within the boundary layer and generated from the local sources. On the other hand, under unstable atmospheric conditions and a deeper boundary layer during April and May, a lower correlation coefficient was found between AOD vs. PM$_x$ ($r < 0.4$). In the rainy season (July to August), the correlation coefficient between AOD and PM$_{10}$ was found higher than PM$_{2.5}$ and PM$_1$, indicating that the AOD better represent the coarser particles than finer particles. The high relative humidity in rainy seasons enhanced the AOD retrieval due to
the hygroscopic growth of aerosol particles, which reflected in a better AOD correlation with PM\(_{10}\) than PM\(_{2.5}\) and PM\(_{1}\). The wash-out of PM\(_x\) due to precipitation, deeper boundary layer, and strong convection during rainy months reduces the PM\(_x\) concentrations, while the AOD retrieval remains high under cloud-free conditions during the inactive rain phase hygroscopic growth of the particles (Mhawish et al. 2021). The results suggested that the AOD can be used as a proxy for air quality under stable atmospheric conditions and dry seasons. On the other hand, the AOD less represents the ground-level PM\(_x\) concentrations during unstable conditions in spring and summer and more influenced by meteorological variables and atmospheric mixing height.

Tropospheric NO\(_2\) and SO\(_2\) are the main precursors for secondary aerosols and consider as co-pollutant of fine particulate matter, which are mainly associated with local anthropogenic sources such as industrial activities, biomass burning, vehicular emission, and thermal power plants. The strong correlation coefficient between PM\(_x\) vs. SO\(_2\) and NO\(_2\) in summer months suggests that the photochemical reactions can contribute to the formation of PM\(_x\). The strong correlation in winter suggests that both trace gases NO\(_2\) and SO\(_2\) originated from the same emission sources of PM\(_x\), mainly domestic heating, industrial activities, and vehicular emissions. While the lower correlation in the summer monsoon suggests the higher contribution of natural sources of PM\(_x\) and the influence of atmospheric stability in the dispersion of air pollutants.
This section presents the annual trends in the six parameters used to assess the air quality in each city of Pakistan. The annual trends were calculated after removing the seasonality from the monthly mean time series data which also accounted for temporal autocorrelation. Figure 16
shows the magnitude of the trends as Sen’s slope over the individual city. A significant positive
trend in PMx was found over most Pakistan cities, particularly in Punjab, Khyber Pakhtunkhwa,
and the Islamabad Capital Territory. The PMx trends found over cities in Punjab range from +0.35
to +1.10 µg/m³ yr⁻¹, +0.42 to +1.52 µg/m³ yr⁻¹, +0.57 to +2.20 µg/m³ yr⁻¹ for PM₁, PM₂.₅ and PM₁₀,
respectively. Correspondingly, Punjab cities show a positive trend in AOD with the strongest
increase over Lahore, reaching 0.008 yr⁻¹. Other cities in Khyber Pakhtunkhwa and Azad Kashmir
also show a positive trend in AOD but with lower magnitudes. The consistent positive trend in
PMx and AOD, particularly over cities in Punjab, suggest increasing aerosol emissions, mainly
from anthropogenic sources and biomass burning which are considered major sources of
ultrafine and fine particles (PM₁ and PM₂.₅) over the region (Alam et al. 2015; Stone et al. 2010).
The positive trends in PMx across major Pakistani cities suggest that local anthropogenic activities
along with biomass burning (particularly in Punjab’s cities) are the main reasons for the increasing
particulate pollution level in Pakistan cities. On the other hand, ~91%, and ~88% of the cities
show a positive trend in trace gases NO₂ and SO₂, respectively. This increase in trace gases would
be a further source of increased particulate pollution, as trace gases facilitate secondary aerosol
formation via gas-to-particle conversion reactions (Seinfeld and Pandis 1998).

In terms of monthly trends, the common feature is that the highest statistically significant
positive trends were found during the cold months (November to February), particularly over
major Punjab cities (Lahore, Faisalabad, and Gujranwala) and Islamabad (Figure S3). These results
suggest that the increasing trend in PMₓ is associated with increasing emission levels during the
cold months, mainly from anthropogenic sources and biomass burning of crop residues. During
the cold months, the air pollutants are trapped near the surface in a shallow boundary layer, and
stable atmosphere conditions (thermal inversion and low wind speed), conducive to the formation of haze. On the other hand, unstable atmospheric conditions with strong mixing in summer enhance pollutant dispersion, and accordingly, pollutant trends in summer varied from statistically non-significant, to low magnitude positive and negative.
Figure 16: Annual mean trend in atmospheric pollutants and trace gases including (a) AOD, (b) PM$_1$, (c) PM$_{2.5}$, (d) PM$_{10}$, (e) NO$_2$, and (f) SO$_2$. The trends were calculated over different periods...
of time, which are indicated on top of each figure at the right-hand side, together with the type of species.

5.7 Potential Source Contribution Function (PSCF) Analysis

This section presents the PSCF analysis which was used to identify the potential source areas for regional as well as long-range transported aerosol pollutants contributing to Pakistan’s air pollution (Begum et al. 2005). PSCF has been widely used and recommended to identify the potential source areas around the receptor locations (Hopke et al. 1995; Lucey et al. 2001; Wang et al. 2012; Wong et al. 2013; Zeng and Hopke 1989). Figure 17 shows that in spring, the potential source areas are mainly located around the cities as well as in other local regions. However, the values of PSCF also indicate the regional transport of aerosol pollutants from Afghanistan and India (see Figure 1 for locations). In summer, pollution in Lahore and Peshawar is mainly from local anthropogenic sources, suggested by high values of PSCF > 0.8. In autumn and winter, Peshawar, Islamabad, and Lahore are mainly influenced by local anthropogenic sources, whereas, Karachi is influenced by both local anthropogenic sources as well as fine dust particles from the Cholistan and Thar deserts (see Figure 1 for locations). Figure 17 shows that Lahore, the top polluted city of Pakistan, is mainly influenced by local potential source areas for aerosol pollutants, during all seasons. This suggests that increases in local anthropogenic activities play an important role in the worsening of Lahore’s air quality. Overall, the higher values of PSCF > 0.6 identify potential source areas which are located both inside and outside of Pakistan, which indicates that the air quality of Pakistan is not only influenced by local sources but also influenced by transport from regional anthropogenic sources areas. However, the main source areas are local.
Figure 17: Potential source contribution function plots for PM$_{2.5}$ at seasonal scales from March 2020 to February 2021 for four receptor cities namely, Peshawar, Islamabad, Lahore, and Karachi (see legend for identification).
6. Conclusions

This study has combined long-term (2003–2020) remote sensing, ground-based, and model simulation datasets to provide the most comprehensive and extensive evaluation ever, of air quality conditions over Pakistan. The analysis includes long-term spatio-temporal distributions of atmospheric pollutants and trace gases, recent long-term trends at the city level, ranking of cities in terms of contributions of air pollution levels into three categories (extremely polluted, highly polluted, polluted cities), and the potential sources of air pollution across Pakistan.

The highest AOD retrievals were observed in the summer months (June to August), mainly attributed to the hygroscopic growth of the aerosol particles during the humid summer season. High AOD levels were also observed during cold months (October to January), mainly over biomass burning affected regions such as Punjab. For PMx and trace gases, highest values were observed during cold months from October to February under stable atmospheric conditions and shallower Boundary layer height, due to higher emission from anthropogenic activities and biomass burning. The higher correlation between columnar aerosol loading (AOD) and the ground-level PM$_{2.5}$ concentrations under stable and dry atmospheric conditions in cold months suggests that AOD can be used as a proxy for air quality. On the other hand, under humid unstable conditions, AOD does not represent the ground-level PM$_{2.5}$ which is lower during the summer months.

The CAMS PM$_{2.5}$ data showed better agreement with ground-based PM$_{2.5}$ concentrations compared to MERRA-2 reanalysis PM$_{2.5}$ data and was used to rank the cities in terms of fine particulate levels. The 18-year average of PM$_{2.5}$ concentrations for the 80 cities of Pakistan
revealed that a total of 21 cities fall within the category of extremely polluted cities ($PM_{2.5} > 92.84$) (namely Punjab: 17, Khyber Pakhtunkhwa: 3, Azad Kashmir: 1), 40 cities fall within the category of highly polluted cities ($45.69 < PM_{2.5} < 92.84$) (namely Punjab: 6, Sindh 14, Balochistan: 3, Khyber Pakhtunkhwa: 13, Azad Kashmir: 4), and 19 cities fall within the category of polluted cities ($PM_{2.5} < 45.69$) (namely Balochistan: 16, and Khyber Pakhtunkhwa: 3). No single city in Pakistan falls within the $PM_{2.5}$ standards defined by Pak-NEQS and WHO, and the values of $PM_1$ and $PM_{2.5}$ respectively for the top 10 cities are 5.6 (8.4) to 9.0 (13.5) times and 7.2 (10.8) to 11.4 (17.1) times greater than the Pak-NEQS (WHO AQG). The annual exposure map showed that the whole country and people of Pakistan are exposed to high long-term $PM_{2.5}$ concentrations as the annual mean concentrations for all cities exceeded the Pak-NEQS ($<15 \mu g/m^3$), and 68, 73, and 80 cities exceeded the WHO Interim Target-1 ($<35 \mu g/m^3$), Target-2 ($<25 \mu g/m^3$), and Target-3 ($<15 \mu g/m^3$), respectively. In terms of pollution sources, the study concludes that biomass (crop residue) burning activities are not likely to be the main source of severe air quality conditions in Pakistan, as the highest values of $PM_{10}$ were observed in December and January and similar results were also observed for $NO_2$ and $SO_2$ which are known to be associated with anthropogenic activities including transport and industrial emissions. Interestingly, higher levels of AOD, $PM_1$, $PM_{2.5}$, $PM_{10}$, $NO_2$, $SO_2$, population density, nighttime lights, and vegetation fire activities showed the same spatial pattern as cropland, and this is because most of the major cities as well as rural areas of Pakistan are surrounded by cropland. This strongly suggests that Pakistan’s extreme air pollution problems are mainly derived from local anthropogenic activities. This is also confirmed by the PSCF ($> 0.6$) analysis based on HYSPLIT air parcel back trajectories and ground-based $PM_{2.5}$ concentrations.
Significant positive trends in the concentrations of AOD, \( \text{PM}_{1} \), \( \text{PM}_{2.5} \), \( \text{PM}_{10} \), \( \text{NO}_2 \), and \( \text{SO}_2 \) were observed from November to February, particularly over Lahore, Islamabad, Gujranwala, and Faisalabad, and these results suggest increasing emission levels, mainly from anthropogenic sources including crop residue burning.

The final remark of this study is that all cities in Pakistan have been exposed to long-term \( \text{PM}_{x} \), \( \text{NO}_2 \), and \( \text{SO}_2 \) concentrations throughout the last two decades. The pollution levels in these cities imply extremely poor air quality conditions, mainly due to local anthropogenic activities, which severely threaten to human life. This comprehensive study, based on long-term multi-source aerosol information, may be considered a baseline study by the Ministry of Climate Change, Pakistan, and other policymakers, to mitigate air pollution problems in Pakistan.

**CRediT authorship contribution statement**

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The authors would like to acknowledge NASA's Level-1 and Atmosphere Archive and Distribution System (LAADS) Distributed Active Archive Center (DAAC) (https://ladsweb.modaps.eosdis.nasa.gov/) for MODIS data, the Copernicus Atmosphere Monitoring Service (CAMS) for air quality data (PM$_{1}$, PM$_{2.5}$, and PM$_{10}$), the Goddard Earth Science DISC (https://daac.gsfc.nasa.gov/) for OMI data, Principal Investigators of Lahore AERONET site, and the NOAA Air Resources Laboratory (ARL) for the provision of the HYSPLIT air parcel back trajectories (https://www.ready.noaa.gov) used in this publication. The authors are grateful to Mr. Abid Omar, founder of Pakistan Air Quality Initiative (PAQI), and gratefully acknowledge the U.S. Department of State for providing the open access ground-based PM$_{2.5}$ data. The authors also acknowledge the use of data from NASA's Fire Information for Resource Management System (FIRMS) (https://earthdata.nasa.gov/firms), part of NASA's Earth Observing System Data and Information System (EOSDIS). The authors are thankful to Mr. Pravash Tiwari for helping in PSCF analysis and Dr. Devin White (Oak Ridge National Laboratory) for MODIS Conversion Tool Kit (MCTK).
Funding

This work was supported by the National Key Research and Development Program of China (2016YFC1400901), the Special Project of Jiangsu Distinguished Professor (R2018T22), the National Natural Science Foundation of China (41976165), Jiangsu Technology Project of Nature Resources (KJXM2019042), and the Startup Foundation for Introduction Talent of NUIST (2017r107). Additional support came from the New Mexico State University College of Agriculture Consumer and Environmental Sciences’ Agricultural Experiment Station.

Reference


algorithm with improved cloud screening for Sun photometer aerosol optical depth (AOD) measurements. *Atmospheric Measurement Techniques, 12*, 169-209


Khanum, F., Chaudhry, M.N., & Kumar, P. (2017). Characterization of five-year observation data of fine particulate matter in the metropolitan area of Lahore. *Air Qual Atmos Health, 10*, 725-736


Li, L., & Wang, Y. (2014). What drives the aerosol distribution in Guangdong--the most developed province in Southern China? *Sci Rep*, 4, 5972


aerosol optical depth retrievals over land with airborne sunphotometer measurements during ARCTAS summer 2008. *Atmospheric Chemistry and Physics, 14*, 2015-2038


WHO (2018a). 9 out of 10 people worldwide breathe polluted air, but more countries are taking action. In. Geneva


85