A neural network-based method for modeling PM 2.5 measurements obtained from the surface particulate matter network

Nnaemeka Onyeuwaoma · Daniel Okoh · Bonaventure Okere

Received: 24 August 2020 / Accepted: 4 April 2021 / Published online: 12 April 2021
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2021

Abstract Air pollution is a global problem; hence, many countries devoted lots of resources towards its study and possible eradication. The major parameter indicator for air quality is the particulate matter (PM). These particles, especially PM$_{2.5}$, are injurious to health either under high concentration levels or after a long-term exposure. PM$_{2.5}$ particles are known to cause lung and respiratory diseases, cardiovascular diseases, and even cancer. In this research, artificial neural networks were used to train PM 2.5 measurements obtained from the Surface Particulate Matter Network (SPARTAN). The training was done using inputs that indicate time series of the measurements and the prevailing atmospheric conditions. The developed models were used to estimate PM 2.5 over a sub-Saharan site in Ilorin. Our study considered meteorological parameters and aerosol optical depth (AOD) as inputs for the neural networks. The targets are PM 2.5 measurements obtained from SPARTAN. Our models showed very high correlation with measured data. Apart from the data generated using model $p$ which has a correlation of 0.0009, the correlation $R^2$ for other models ranges from 0.59 to 0.95 which has a good performance. The model PRB estimated both low and high PM better while others either under or over predict emission scenarios.

Keywords Particulate matter · Artificial neural network · Aerosols · Meteorology · Pressure

Introduction

Air pollution is the contamination of the indoor or outdoor environment by any chemical, physical, or biological agent that modifies the natural characteristics of the atmosphere (WHO, 2017). A critical air pollutant is particulate matter (PM), which is measured in micrograms per cubic meter ($\mu$g/m$^3$). The major components of PM are sulfate, nitrates, ammonia, sodium chloride, black carbon, and mineral dust (Amegah and Mensah, 2017; WHO, 2017), including the novel SARS-CoV-2 (Anchordoqui et al., 2020). Results have shown that SARS-CoV-2 virions have similar characteristic like other air pollutant with a determined settling velocity depending on the particulate size (see Anchordoqui et al., 2020; Anchordoqui & Chudnovsky, 2020). Furthermore, the main sources of PM are household combustion devices, motor vehicles, industrial facilities, poor waste management system, forest fires (Brauer et al., 2012), and human reflexes (sneezing and exhalation) (Anchordoqui et al., 2020). In the study of air quality, atmospheric aerosol, which is a mixture of solid and liquid particulate matter (PM), is also a major pollutant.
Particulate matter, which is used to segregate between aerosols, is majorly discussed in air quality-related researches as either PM$_{2.5}$ (these are particle loading with a diameter of 2.5 µm or less (PM$_{2.5}$), ultrafine particles, and soot) or PM10 (loading of particles with a diameter less than or equal to 10 µm (PM$_{10}$)). These particles, especially PM$_{2.5}$, are injurious to health either under high concentration levels or after a long exposure. PM$_{2.5}$ particles are known to cause lung and respiratory diseases, premature death (Wang & Christopher, 2003), cardiovascular diseases (Weber et al., 2010), and even cancer (Pope et al., 2002).

Air pollution in Africa, as at today, is responsible for a high number of premature deaths (Mbewu & Mbanya, 2006; WHO, 2013) and also hinders economic development with a resultant cost of 2.7% of GDP of the continent. The report by Vidal (2016) (see www.theguardian.com/global-development/2016/oct/20/air-pollution-deadlier-africa-than-dirty-water-or-malnutrition-oecd) highlighted that polluted air kills approximately 712,000 yearly in Africa, compared to 542,000 from unsafe water, 275,000 from malnutrition, and 391,000 from unsafe sanitation. This translates to an increase of 36% in the death rate as a result of outdoor air population from 1990 to 2013 (Brauer et al., 2012). The WHO reported that this mortality is mainly due to exposure to inhalable particulate matter, particles of 10 microns or less in diameter (PM$_{10}$), which cause cardiovascular and respiratory disease and cancers. The economic loss as a result of air pollution in Africa was put at $447bn in 2013 (Vidal, 2016), which amounts to a huge economic loss compared to poverty index and average per capita income in Africa. Hence, air pollution poses more risk in Africa than malnutrition and diseases. Air pollution sources in Sub-Saharan African (SSA) can be natural or anthropogenic. Petkova et al. (2013) and Naidja et al. (2018) reported that biomass burning, poor waste management, and the high number of diesel-powered vehicles are the predominant sources of anthropogenic PM, while Saharan dust and savanna fires are the main natural sources of particulate matter. Within the sub-Saharan region, particles from the Sahara Desert and in particular the Bodélé depression greatly affect the PM all year round (Onyeuwaoma et al., 2015). Therefore, there is the need to understand the role of these aerosol sources on the sub-Saharan public health. Furthermore, air pollution episodes can also be attributable to power generation, increased industrialization, burning of refuse, millions of steel diesel electricity generators, cars which have had the catalytic converters removed, petrochemical plants, the construction sector (Burciaga, 2020), and indoor cooking with inefficient fuel stoves (Mulenga & Siziya, 2019). In spite these aforementioned sources, knowledge of the sources of air pollution, its magnitude, and impact in much of Africa is still sketchy. In order to tackle air quality headlong, the World Health Assembly (WHA), in its Sixty-ninth assembly in 2016, proposed a more vigorous and comprehensive research on air quality to quantify and ascertain its actual health impact. To compound the issue, most Sub-Saharan Africa (SSA) countries have no actual historical long-term air quality monitoring station (Petkova et al., 2013). This underscores the need to use other proxy data to apportion the source and quantity of PM 2.5 emission within this region. On the other hand, Brauer et al. (2012) opined that PM 2.5 concentration in SSA stood at around 100 µg/m$^3$ compared to <20 µg/m$^3$ in the Europe and North America. This implies that in SSA, the air quality scenario is tending towards the Asian experience. Therefore, in order to effectively tackle the problems associated with air quality, there must be a strong political commitment aimed at re-writing the narrative. This lack of will/commitment according to Amegah and Agyei-Mensah (2017) is attributable to the following reasons:

- Lack of reliable data on air pollution levels due to weak and non-existent air quality monitoring network of stations.
- Lack of local evidence on the environmental and human health impact of air pollution, and the magnitude of the associated health risk.

Given this, there is the need for an up-to-date data on the state of the air quality in the region in order to achieve the Sustainable Development Goals (SDG) on air quality, other efforts geared towards the provision of clean air had been the Clean Air Initiative in Sub-Saharan Africa (CAI-SSA) which operated from 1998 to 2002, but this program focused majorly on vehicular pollutant emission reduction instead of air quality monitoring and assessment (see Bultynck and Reliquet (2003)). Consequently, the risks associated with long-term exposure to PM lead the US Environmental Protection Agency (EPA) in 1997 to set up a new National Ambient Air Quality Standards (NAAQS), while China in 1982 set
up the Ambient Air Quality Standards (AAQS) (Zhang et al., 2015) to monitor and control air pollution in China; hence, SSA countries should follow suit. In view of this, the established risky conditions for PM$_{2.5}$ concentration exposure starts from 40.5 µm$^{-3}$ especially for individuals suffering from heart or lung ailments, older adults, and children (Gorai et al., 2018). Unfortunately, most of the studies used in establishing these epidemiological evidences were conducted in Europe and North America where PM data are available, while regions prone to bad air quality like Africa have no PM data (Snider et al., 2015) nor air quality guidelines (Petkova et al., 2013). To buttress this point, World Health Organization reported that air quality monitoring in SSA is only conducted in 9 out of the 27 countries within the region (WHO, 2013). To compound the issue, most available data are given at a price. Outside the SSA, a lot of efforts had been committed towards air quality research; for instance, Sajjadi et al. (2017) studied the concentration of PM 2.5 and PM 10 in Sabzevar, Iran, using measurements from 48 stations. Elsewhere in China, Zhang et al. (2006) and Vedal et al. (2017) showed that there have been intensive studies on how to curb air pollution and its health effects. Zhu (2017) opined that the Chinese scientific communities are already familiar with air pollution and its control since the 1970s till date, and their results are considered during policy formulation. To this effect, Lim et al. (2012) reported that the Global Burden of Disease (GBD) assessment attributed 3.2 million premature deaths per year to ambient PM$_{2.5}$, which made PM$_{2.5}$ exposure a major cause of premature death globally. This implies that, in assessing the GDB, the volume of PM$_{2.5}$ emitted must be well integrated (Lim et al., 2012; WHO, 2014), for environmental performance indicators (EPI, 2014). Therefore, to understand the air quality within any given environment, it needs to be studied in relation with other local prevailing factors. This gave rise to various PM models as shown in see section “PM measurements”.

**PM measurements**

**AOD-PM relationship**

To estimate PM from AOD, there exist explicit formulations for it depending on the data available; van Donkelaar et al. (2015) related AOD with PM$_{2.5}$; thus,

$$M_{2.5,d,\Delta z} = \left[\frac{4}{3} \left(\frac{r_{2.5,d,\text{eff}}}{r_{2.5,\text{eff}}}\right)^{2} \left(\frac{\tau_{2.5,\Delta z}}{\tau}\right)\rho_{2.5,d,\text{eff}} \Delta z \Delta e^{-1}\right] \tau$$

(1)

where $M_{2.5,d,\Delta z}$ is the total fine dry aerosol mass between the surface and altitude $z$, $r_{2.5,d,\text{eff}}$ is the fine dry effective radius, and $\tau_{2.5,\Delta z}$ is the fine AOD between the surface and altitude $z$. All parameters in the bracketed expression except $\tau$ refer to representative values between the ground and height $z$. We refer to $M_{2.5,d,\Delta z}$ as remote-sensed PM$_{2.5}$.

Liu (2015) expressed the relationship thus:

$$C = \frac{4\rho_{e}}{3Q} \times \frac{f_{\text{PBL}}}{H_{\text{PBL}}} \times \text{AOD}$$

(2)

$\rho$=particle density, $Q$=extinction coefficient, $r_{e}$=effective radius, $f_{\text{PBL}}$=% AOD$_{550\text{nm}}$ in PBL, $H_{\text{PBL}}$=mixing height, $C$=PM 2.5 equations 1, 2, and 3 can all be used to arrive at PM estimation depending on the data available.

In this research, multiple regression analysis of the form

$$PM\ 2.5 = \beta_0 + \beta_x \times \tau + \sum_{n=1}^{m} (\beta_n \times M_n) + \delta_i$$

(3)

$\beta_0$, $\beta_x$, and $\beta_n$ are the regression parameters, $\tau$ and $M_n$ are the independent, $n (=1, 2, 3...)$ which are the meteorological parameters of interest) is the index for gage $n$, $\delta_i$ is the model error.

Consequently, the models in Eqs. 1 and 2 were jetisoned in this study because most of the input parameters are not readily available in our study area. Hence, the network of stations available does not capture data on them. Therefore, we relied on the available data to decipher the PM concentration within our study location. To assess the level of PM exposure requires the establishment of a network of stations for continuous measurement and monitoring such as the Surface Particulate mAtter Network (SPARTAN), which was initiated to cover some select locations across the globe (Snider et al., 2015). Though some of these stations are currently not functional especially those located within SSA, outside the SPARTAN but within the sub-Saharan Africa (except Republic of South Africa), it is not on record of any existing continuous PM monitoring station (see Fig. 1). Most of the times, measurements are carried out as campaigns by individual scientists as the need arises due to the cost involved.
Once air quality is quantified, its health implications can then be established for SSA, which will translate to national/regional policy on air quality and national air quality standards.

Due to the lack of in situ air quality monitoring stations, surface PM loading prediction becomes a verifiable option (Mulenga & Siziya, 2019) using retrieval of columnar aerosol content and meteorological data. Wang and Christopher (2003) correlated AOD, which is a measure of the extinction of the solar beam by particulate matter (NOAA, 2018) and surface PM$_{2.5}$ concentrations and obtained a correlation of 0.7. This indicates that aerosol particles are well-mixed within the lower boundary layer. On the other hand, Snider et al. (2015) and van Donkelaar et al. (2015) opined that satellites offer a concrete platform for long-term monitoring of PM$_{2.5}$ concentrations at the surface, while taking dependent factors such as aerosol vertical distribution, humidity, and aerosol composition into consideration as model input parameters. Such models will involve using columnar aerosol content and meteorological data. Hence, archived satellite data that had been in existence for over two decades can be a veritable tool in studying the global PM trend and the extent of air quality degradation worldwide.

Therefore, the aim of this study is to generate models for predicting and quantifying PM 2.5 over a moribund SPARTAN station in SSA using aerosol, meteorological parameters, and archived SPARTAN PM 2.5 data as variables. The best performing model will be used to extrapolate the air quality situation in Nigeria while putting the local meteorology into consideration. These models will be the first comprehensive air quality model for Nigeria, which considered the aerosol loading and different meteorology simultaneously.

The flow of this study is as follows: In section “Introduction”, we discussed the general theoretical background to this work and outlined some of the established PM models. In section “Methodology”, we will discuss the methodology employed in this research, coupled with a brief discussion on the study area and the data used. Finally, results and discussions are presented in “Results and discussions” with the conclusion.

**Methodology**

**Region and site descriptions**

The study area for this research is a site within the Sub-Saharan Africa (SSA). SSA is the geographical area of the continent of Africa that lies south of the Sahara. According to the United Nations, it consists of all African countries that are fully or partially located south of the Sahara. Generally, Africa is the hottest continent on earth; dry lands and deserts comprise 60% of the entire land surface. The Sahara Desert which is the world’s largest desert, with a temperature above 37.78 °C (100 °F), is located within the SSA. Furthermore, SSA is characterized by sparse rainfall. The locations for this study are Ilorin in Nigeria located at latitude 8.484, longitude 4.675. This location had at some point a PM monitoring and meteorological stations operated by Surface PARTICulate mAtter Network (SPARTAN) instrument. The data from this station will be used.
as the base data (Snider et al., 2015) for integration into the models. Studies carried out in Ouagadougou, Burkina Faso (Boman et al., 2009), eastern Botswana (Chimidza & Moloi 2000), and Dar es Salaam (Tanzania) (Eliasson et al., 2009) show that soil dust is the major component of particulate matter in most parts of SSA. Nyanganyuraa et al. (2007) showed that biomass contributes substantial percentage of particulate matter concentration Zimbabwe (Fig. 2).

Data

The data for this work are.

- PM 2.5 was acquired from the archive of the Surface PARTICulate mAtter Network (SPARTAN) for Ilorin.
- Meteorological parameters of relative humidity (R), temperature (T), pressure (P), and fractional relative humidity (F) were used to understudy the role of meteorology in PM abundance; hence, climate has direct effect on air quality (Mulenga & Siziya, 2019). These parameters were substituted interchangeable to create different models. Previous researches observed that at high RH (Wang & Christopher, 2003), the hygroscopic effect sets in thereby changing the size distribution and the scattering efficiencies of particles (Wang & Martin, 2007). Conversely, Gupta and Christopher (2009) reported that high temperatures enhance the photochemical reactions which in turn lead to increased production of PM$_{2.5}$ particles in the atmosphere, which implies that temperature is a very important factor to be considered in PM estimation. Furthermore, pressure was included in the analysis to establish its relationship with particulate matter concentration.
- Total particle light scattering (B) was used in place of aerosol optical depth (AOD) data because

![Fig. 2 Map showing the study area, Ilorin (source Oluyemi et al., 2016)](image_url)
of its availability within the temporal scale. Hence, the available AOD data had a lot of missing days. Furthermore, B gives information on the aerosol loading within the environment (see Eq. 4).

\[ B = 3f(RH)[(NH_2)_2SO_4 + NH_4NO_3] + 4[1.4OC] + 1[Soil] \]

(4)

The dry scattering efficiencies were assumed to be 3 \( m^2/g \) for (AMSUL) and \( NH_4NO_3 \) (AMNIT), 4 \( m^2/g \) for organic mass (OM), and 1 \( m^2/g \) for fine soil. The \( f(RH) \) term accounts for the increase in Bsp caused by hygroscopic growth of AMSUL and AMNIT (RH=relative humidity) (Tang & Munkelwitz, 1994). All these data were acquired alongside the PM data at the SPARTAN station.

**Data integration**

In order to integrate all the data to arrive at the different models, we used artificial neural network (ANN), which had proved to be a veritable statistical instrument in this regard. Neural networks are basically structured in 3 layers: an input layer, a hidden layer, and an output layer (Fig. 3).

The input layers are the meteorological parameters and B, while the target is the measured PM\(_{2.5}\) and the expected output the predicted PM\(_{2.5}\). The networks were trained using the Levenberg-Marquardt (LM) algorithm, which minimizes the error functions that arise during neural network training and saves time (Jang et al., 1997). Furthermore, MATLAB tansig function used to transfer functions between the input layer and the hidden layer and between the hidden layer and the output layer was applied as in Okoh et al. (2019), which underlining equations are given in Eqs. 5 and 6.

\[ H_m = \tanh(I_{wm} \times I_m + B_1) \]  
\[ O_m = \tanh(H_{wm} \times H_m + B_2) \]

(5)

(6)

where Eq. 5 is the equation connecting the input layer matrix \( I_m \) to the hidden layer matrix \( H_m \), and Eq. 6 connects the hidden layer matrix to the output layer matrix \( O_m \). \( I_m \) contains inputs for the neural network, \( H_m \) contains intermediary values computed within the hidden layer, and \( O_m \) contains outputs from the neural network. \( I_{wm} \) and \( H_{wm} \) are, respectively, weight matrices for the input layer and the hidden layer, while \( B1 \)
and B2 are, respectively, bias vectors for the input layer and the hidden layer. The input weight matrices and the bias vectors contain constants for a given trained neural network.

After training, it is the weight matrices and the bias vectors that characterize the trained neural network; they do not change unless the neural network is retrained. Twelve different models used in the predictions as shown in section “Results and discussions” (Figs. 6, 7, 8, and 9) were generated.

The data used for this research were split in time into three sets: 70% was used for training, 15% was used for validation, and another 15% was set aside for testing. During the training, each network had varying number of neutrons in the input layer. For each, 100 trained networks were generated which we used the validation data set to compute the root-mean-square error (RMSE) (see Eq. 7) and the least RMSE was used as the number of hidden layer neurons for generation of the network predictions for the PM measurement (see Fig. 4)

$$\text{RMSE} = \sqrt{(p - m)^2}$$ (7)

where

\[ p = \text{predicted values (result from the neural network)} \]
\[ m = \text{measured values (results from the ground station)} \]

An example of the neutral network architecture is as shown in Fig. 5 which is a 5–10-1.

**Results and discussions**

In this section, we present the discussions on the results obtained from the neural network analysis from the different combinations of meteorological parameters.

Figure 6 shows the plots of modelled (from B, FB and P) and measured data. The B plot shows the modeling of PM from total particle light scattering (B) only. The result from this plot indicates that B can predict PM 2.5 with an $R^2$ of 0.956 (see scatter plot) and a root mean squared error (RMSE) of 2.73 $\mu m^{-3}$. It was further observed that there was a slight over prediction of PM by the model when the PM reading is below 10 $\mu m^{-3}$. In general, B shows a very high predictability of PM. The plot FB shows that at some points, the model over predicts (as indicated by the spikes) PM 2.5 both at high and low
PM concentration, which resulted in a correlation of 0.612 and a RMSE of 8.11 µm$^{-3}$ (see Table 1). Between the measured and modelled data, the plots indicate a moderate predictability of PM from FB. Furthermore, when P alone was used as the predictor variable, the result shows an over estimation of

**Fig. 5** Schematic illustration of the neural network with architecture 5–10-1

**Fig. 6** Model results in comparison with measured data, when B, FB, and P were used in the neural network prediction. The red lines show the modelled data while blue lines show the measured data alongside the scatter plot.
### Table 1  Summary of results from the models

| Parameters | $R$-squared | Adjusted $R$-squared | F-statistic vs. constant model | $p$ value | Root mean squared error |
|------------|-------------|----------------------|--------------------------------|-----------|-------------------------|
| TB         | 0.937       | 0.937                | 8.77e+03                       | 0         | 3.26                    |
| TRB        | 0.845       | 0.845                | 3.2e+03                        | 4.6e-240  | 5.13                    |
| TFB        | 0.71        | 0.71                 | 1.44e+03                       | 3.45e-160 | 7.01                    |
| RB         | 0.844       | 0.844                | 3.17e+03                       | 4.07e-239 | 5.15                    |
| PTRB       | 0.838       | 0.837                | 3.04e+03                       | 2.21e-234 | 5.24                    |
| PTRB       | 0.642       | 0.642                | 1.06e+03                       | 2.25e-133 | 7.78                    |
| PRB        | 0.751       | 0.751                | 1.78e+03                       | 7.84e-180 | 6.49                    |
| PFB        | 0.741       | 0.741                | 1.68e+03                       | 1.41e-174 | 6.62                    |
| PFB        | 0.596       | 0.595                | 866                            | 1.03e-117 | 8.27                    |
| PB         | 0.639       | 0.639                | 1.04e+03                       | 2.4e-132  | 7.81                    |
| P          | 0.000951    | -0.000748            | 0.56                           | 0.455     | 13                      |
| PB         | 0.612       | 0.611                | 927                            | 5.87e-123 | 8.11                    |
| B          | 0.956       | 0.956                | 1.28e+04                       | 0         | 2.73                    |

**Fig. 7** Model results in comparison with ground-based data, when PRB, PB, and PFB were using in the neural network prediction. The red line shows the modelled data while blue line shows the ground-based data alongside the scatter plot.
PM in all concentration scenario. The correlation of P model stood at 0.000951 and RMSE of 13 µm$^{-3}$ which does not in any way replicate the measured data. This indicates that P alone cannot be used to determine air quality situation.

Figure 7 on the other hand shows the estimation of PM from PRB, PB, and PFB, respectively. The predictive model from PRB shows a near perfect fitting of both measured and predicted data with a correlation of 0.741 and RMSE of 6.62 µm$^{-3}$. This implies that a combination of these meteorological parameters can effectively predict PM during low and high episodes. Also, PB on the other hand shows a high PM predictability with an $R^2$ of 0.639 and 7.81 µm$^{-3}$ RMSE. A combination of PB that was observed slightly over predicts PM especially when the value is below 10 µm$^{-3}$. For PFB, the results indicate an under estimation of low PM as seen by the spikes pointing downwards. Subsequently, when the measured values are above 10 µm$^{-3}$, there is a fair representation from this model. The correlation and RMSE for this model stood at 0.596 and 8.27, respectively.

Meanwhile, Fig. 8 shows the results from models generated from combinations of RB, PTRB, and PTFB, respectively. The RB model shows that the modelled PM values fall within the measured PM and perfectly followed the trend of PM emission.

Statistically, the result showed a correlation of 0.844 and 5.15 RMSE, respectively. This result implies that a combination of RB can give the average value of PM in a given time. Hence, this model can be used to show the probability of the PM concentration. With the combination of PTRB, the result

![Fig. 8 Model results in comparison with ground-based data, when RB, PTRB, and PTFB were using in the neural network prediction. The red line shows the modelled data while blue line shows the ground-based data alongside the scatter plot](image-url)
showed a correlation of 0.838 and an RMSE of 5.24. This model over predicts PM when the values are above 10 µm⁻³. Using PTFB as a predictor combination gave an average prediction value at an $R$ square of 0.64 and an RMSE of 7.78 µm⁻³. The curve further showed that all the predicted values fall within the measured range, but slightly over predicts PM when it falls below 10 µm⁻³. In Fig. 9, we showed the results obtained from models TRB, TB, and TFB. The plot TRB showed similar results obtained in Figs. 6 and 8 (see B) but an $R$ squared of 0.84 and RMSE of 5.13 µm⁻³. The difference between models TRB and B is that low PM values cannot be easily obtained from B.

Furthermore, the model TB over predicted low ranged values but does better when the PM values are relatively high with an $R$ square of 0.93 and RMSE of 3.26 µm⁻³ showing similar results like in B. A comparison of B and TB showed that the influence of temperature in the model is limited when other meteorological parameters are removed (see Fig. 6). Meanwhile, TFB showed an incoherent prediction pattern, where the predicted and measured values do not match. In most cases, low measured value corresponds to the peak modelled values though the $R$ squared and RMSE are 0.71 and 7.01, respectively.

On the individual performance of the model, B, FB, RB, and PB show that at high PM concentration, these models perform well, but when the value goes below 10 µm⁻³, the models do not balance well again as it over predicts the PM concentration. P alone on the other hand as a sole predictor does not have a definite pattern. Subsequently, PFB and PRB on the other hand predict both high and low scenarios effectively.

Fig. 9 Model results in comparison with ground-based data, when TRB, TB, and TFB were using in the neural network prediction. The red line shows the modelled data while blue line shows the ground-based data alongside the scatter plot.
The introduction of the parameters R and F in the models determines the direction of the predicted values. When F is introduced, the model will under predict low PM events while R captures low events more accurately. This implies that to capture low emission events, R must be put into consideration in the model.

When the temperature component was considered in the model (PTRB, PTFB, TRB, TB, TFB), it resulted thus; for TB, the measured PM was well predicted when the value is above 10 µm$^{-3}$, while lower values were over predicted. Meanwhile, on addition of R, it was able to stabilize the model by capturing values lower than 10 µm$^{-3}$ effectively. Further to this, the substitution of R with F was able to predict PM greater than 10 µm$^{-3}$, while under predicting lower values. It was also observed that when the T component was added to PRB, it counters the effect of effect of R given rise to the same result as was observed in RB (see Fig. 9). Similarly, T has the same effect when added to PFB; hence, the model over predicts PM values less than 10 µm$^{-3}$. These results further showed that temperature (T) as a meteorology factor in PM prediction counteracts the effect of relative humidity in the models contrary to the submission by Gupta and Christopher (2009).

Finally, the aerosol component (B) in the models that was observed replicates PM when the values are above 10 µm$^{-3}$, while an addition of pressure (P) and relative humidity (R & F) takes care of low emission scenarios. This implies that there is a strong relationship between low PM and parameters P, R, and F. Consequently, in order to predict PM above 10 µm$^{-3}$, B can be used independently, but for a holistic prediction, the combination of PRB is most appropriate.

Figure 10 is the result of covariance analysis of the measured and modelled PM 2.5, which shows that only three models (p, b, and tb) had varied mean from the measured data. This implies that meteorological impacts on PM vary from one parameter to another.

When compared with other models, our models performed relatively better except for p (0.0009) and pfb (0.59). For instance, Wang et al. (2016) using linear regression obtained a correlation ranging between 0.44 and 0.55, while Zheng et al. (2016) arrived at a correlation of 0.77 using linear mixed effects models. Lin et al. (2015), on the other hand, arrived at a correlation of 0.7, 0.77, and 0.83 for three urban locations studied in China. Using Bayesian linear regression, Lv et al. (2016) obtained an $R^2$ of 0.75; also, Chen et al. (2018) using machine learning achieved an $R^2$ ranging between 0.36 and 0.9. Again, using the XGBoost model, Gui et al. (2020) obtained an $R^2$ of 0.79, 0.8, 0.86, 0.81 for daily, monthly, seasonally, and annual measurements, respectively. These comparisons
invariably show that machine learning is well situated to replicate measured PM with minimal deviations. Generally, our models showed very high correlation with measured data which ranges from 0.59 to 0.95 with the exception of P which had a correlation of 0.0009.

Conclusion

A neural network-based method was used to develop different models for predicting PM 2.5 in a sub-Saharan African site using meteorological and aerosol parameters. The results obtained were compared with measured data to ascertain the performance of the individual models against the measured data. The models performed differently, some predicted PM very well when the value is above a certain threshold, while PRB predicated both high and low scenarios effectively. Subsequently, model P showed no resemblance with the measured data at a correlation of 0.0009. Generally, the results from models showed that air quality can be inferred from meteorology in areas where there is dearth of PM data. These models can be used to estimate PM$_{2.5}$ and extrapolate in other environments that lack PM data.

Therefore, in this research, we have succeeded in developing models from meteorological parameters for determining air quality (PM 2.5) in sub-Saharan Africa. The choice of the model(s) by an individual user depends on the data available at the disposal of the user, while taking into consideration the established accuracy of level of the chosen model. Finally, further work is ongoing to determine and spatially characterize the variation of aerosol loading in sub-Saharan Africa from 1998 to 2020.

Acknowledgements We thank the SPARTAN project for its effort in establishing and maintaining Ilorin site. The SPARTAN network was initiated with funding from the Natural Sciences and Engineering Research Council of Canada.

Author contribution ON conceived and designed the project and wrote the first draft of the manuscript. OD wrote some of the codes used in analyzing the data. ON, OD, and OB contributed to the writing of the manuscript and given critical comments on the draft manuscript. All authors agree with manuscript results and conclusions, reviewed and approved the final manuscript.

Data availability The data used for this research will be made available to interested users on request. They are open source data from SPARTAN network of stations.

Code availability The codes will be made available if requested for.

Declarations

Competing interests The authors declare no competing interests.

References

Amegah, A. K., & Agyei-Mensah, S. (2017). Urban air pollution in Sub-Saharan Africa: Time for action. *Environmental Pollution*, 220, 738–743. https://doi.org/10.1016/j.envpol.2016.09.042

Anchordoqui, L. A., & Chudnovsky, E. M. (2020). A physicist view of COVID-19 airborne infection through convective airflow in indoor spaces. *Science Medicine Journal*, 2, Special Issue "COVID-19". https://doi.org/10.28991/SciMedJ-2020-02-SI-5

Anchordoqui, L. A., Dent, J. B., & Weiler, T. J. (2020). A physics modeling study of COVID-19 transport in air. *Science Medicine Journal*, 2, Special Issue "COVID-19". https://doi.org/10.28991/SciMedJ-2020-02-SI-7

Boman, J., Linden, J., Thorsson, S., Holmer, B., & Eliasson, I. (2009). A tentative study of urban and suburban fine particles (PM$_{2.5}$) collected in Ouagadougou, Burkina Faso. *Spectrometry: An International Journal*, 38:354–362. https://doi.org/10.1002/xrs.1173

Brauer, M., Amann, M., Burnett, R. T., Cohen, A., Dentener, F., Ezzati, M., et al. (2012). Exposure assessment for the estimation of the global burden of disease attributable to outdoor air pollution. *Environmental Science and Technology, 2012*, 652–660. https://doi.org/10.1021/es2025752

Bultynck, P., & Reliquet, C. (2003) 1998–2002 PROGRESS REPORT on World Bank Clean Air Initiative In Sub-Saharan African Cities. The World Bank. https://documents.worldbank.org/en/publication/documents-reports/documentdetail/344531468767960431/clean-air-initiative-in-sub-saharan-african-cities-1998–2002-progress-report. Accessed 20 Jun 2020.

Burciaga, U. M. (2020). Sustainability assessment in housing building organizations for the design of strategies against climate change. *HighTech and Innovation Journal*, 1(4), https://doi.org/10.28991/HiJ-2020-01-04-01

Chen, G., Li, S., Knibbs, L. K., Hamm, N. A. S., Cao, W., Li, T., Guo, J., Ren, H., Abramson, M. J., & Guo, Y. (2018). A machine learning method to estimate PM2.5 concentrations across China with remote sensing, meteorological and land use information. *Science of the Total Environment*, 636, 52–60. https://doi.org/10.1016/j.scitotenv.2018.04.251
Chimidza, S., & Moloi, K. (2000). Identification of sources of aerosol particles in three locations in eastern Botswana. *Journal of Geophysical Research: Atmospheres*, *105*(D14):811–818. https://doi.org/10.1029/2000JD900182

Eliasson, I., Jonsson, P., & Holmer, B. (2009). Diurnal and intra-urban particle concentrations in relation to wind-speed and stability during the dry season in three African cities. *Environmental Monitoring and Assessment*, *154*(1–4):309–324. https://doi.org/10.1007/s10661-008-0399-y

Environmental Performance Index (EPI). (2014). *Environmental Performance Index 2014 CRC Analysis and Recommendations*. http://publications.jrc.ec.europa.eu/repository/bitstream/JRC89939/lnha26623en.pdf. Accessed 30 May 2018.

Gorai, A. K., Tchounwou, P. B., Biswal, S. S., & Tuluri, F. (2018). Spatio-temporal variation of particulate matter (PM2.5) concentrations and its health impacts in a mega city. *Delhi in India. Environmental Health Insights*, *12*, 1–9. https://doi.org/10.1177/1178630218792861

Gupta, P., Christopher, S. A., Wang, J., Lee, Y., & Kuma, N. (2006). Satellite remote sensing of particulate matter and air quality assessment over global cities. *Atmospheric Environment*, *40*(30), 5880–5892. https://doi.org/10.1016/j.atmosenv.2006.03.016

Gupta, P., & Christopher, S. A. (2009). Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: multiple regression approach. *Journal of Geophysical Research*, *114*, D14205. https://doi.org/10.1029/2008JD011496

Gü, K., Chea, H., Zengc, Z., Wangc, Y., Zhaid, S., Wangc, Z., Luoc, M., Zhanga, Z., Liaof, T., Zhaoa, H., Lia, L., Zhenga, Y., & Zhanga, X. (2020). Construction of a virtual PM2.5 observation network in China based on high density surface meteorological observations using the Extreme Gradient Boosting model. *Environment International*, *141*, 105801. https://doi.org/10.1016/j.envint.2020.105801

Jang, J. S. R., Sun, C. T., & Mizutani, E. (1997). *Neuro fuzzy and soft computing*. Prentice-Hall.

Lin, C., Li, C., Yuan, Z., Lai, A. K. H., Li, C., & Fung, J. C. H. (2015). Using satellite remote sensing data to estimate the high-resolution distribution of ground-level PM2.5. *Remote Sensing of Environment*, *156*, 117–128. https://doi.org/10.1016/j.rse.2014.09.015

Lim, S. S., Vos, T., Flaxman, A. D., Danaei, G., Shibuya, K., Adair-Rohani, H., et al. (2012). A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*, *380*, 2224–2260. https://doi.org/10.1016/S0140-6736(12)61766-8

Liu, Y. (2015). Particulate Matter Air Quality from Space – Advanced Statistical Modeling https://appliedsciences.nasa.gov/sites/default/files/2020-11/ParticulateMatter_Part3. Accessed 21 January, 2019

Lv, B., Hu, Y., Chang, H. H., Russell, A. G., Cai, J., Xu, B., & Bai, Y. (2016). Daily estimation of ground-level PM2.5 concentrations at 4 km resolution over Beijing-Tianjin-Hebei by fusing MODIS AOD and ground observations, *Science of the Total Environment*. https://doi.org/10.1016/j.scitotenv.2016.12.049

Mbewu, A., Mbanya, J. C. (2006). Cardiovascular disease. In: Jamison, D. T., R. G. Feachem, M. W., Makgoba, E. R., Box, F. K., Baingana, K. J., Hofman, K. O., Rogo, E. (Eds), *Disease and Mortality in Sub-Saharan Africa 2 The International Bank for Reconstruction and Development / The World Bank.*

Mulenga, D., & Siziya, S. (2019). Indoor air pollution related respiratory ill health, a sequel of biomass use. *Science of Medicine Journal*, *1*:1. https://doi.org/10.28991/SciMedJ-2019-0101-5

Naidja, L., Ali-Khodja, H., & Khardi, S. (2018). Sources and levels of particulate matter in North African and Sub-Saharan cities: a literature review. *25*:12303. *Environmental Science and Pollution Research International*, *25*(13):12303–12328. https://doi.org/10.1007/s11356-018-1715-x

NOAA. (2018). https://www.esrl.noaa.gov/gmd/grad/surfrad/aod/. Accessed 26 Jul 2018.

Nyanganyurua, D., Maenhaut, W., Mathuthu, M., Makarau, A., & Meixner F. X. (2007). The chemical composition of tropospheric aerosols and their contributing sources to a continental background site in northern Zimbabwe from 1994 to 2000. *Atmospheric Environment*, *41*(12):2644–2659. https://doi.org/10.1016/j.atmosenv.2006.11.015

Okoh, D., Seemala, G., Rabiu, B., Habarulema, J. B., Jin, S., Shiokawa, K., Otsuka, Y., Aggarwal, M., Uwamahoro, J., Mungufeni, P., Segun, B., Obafaye, A., Ellahony, N., Okonkwo, C., Tshisaphungo, M., & Shetti, D. (2019). A neural network-based ionospheric model over Africa from Constellation Observing System for Meteorology, Ionosphere, and Climate and Ground Global Positioning System observations. *Journal of Geophysical Research: Space Physics*, *124*. https://doi.org/10.1029/2019JA027065

Oluyemi, J., Yinusa, M., Abdullateef, R., & Adejoke, J. (2016). Utilization of Herbal Medicine Among Inhabitants of an Urban Centre in North-Central Nigeria. *Algerian Journal of Natural Products* 4(3):367–378. https://doi.org/10.5281/zenodo.254532

Onyeuwaoma, N. D., Nwofor, O. K., Chineke, T. C., Eguaroje, E. O., & Diike, V. N. (2015). Implications of MODIS impression of aerosol loading over urban and rural settlements in Nigeria: Possible links to energy consumption patterns in the country. *Atmospheric Pollution Research*, *6*, 484–494. https://doi.org/10.5094/APR.2015.054

Petkova, E. P., Jack, D. W., Volavka-Close, N. H., & Kinney, P. L. (2013). Particulate matter pollution in African Cities. *Air Quality Atmosphere and Health*, *6*, 603–614. https://doi.org/10.1007/s11869-013-0199-6

Pope, C. A., Burnett, R. T., Thun, M. J., Calle, E. E., Krewski, D., Ito, K., & Thurston, G. D. (2002). Lung cancer, cardiopulmonary mortality and long-term exposure to fine particulate air pollution. *The Journal of the American Medical Association*, *287*, 1132–1141. https://doi.org/10.1001/jama.287.9.1132

Sajjadi, S. A., Zolfaghari, G., Adabe, H., Allahabadid, A., & Delsouza, M. (2017). Measurement and modeling of particulate matter concentrations: Applying spatial analysis and regression techniques to assess air quality. *Methods*, *X4*, 372–390. https://doi.org/10.1016/j.mex.2017.09.006

Snider, G., Weagle, C. L., Martin, R. V., van Donkelaar, A., Conrad, K., Cunningham, D., et al. (2015). SPARTAN:
a global network to evaluate and enhance satellite-based estimates of ground-level particulate matter for global health applications. *Atmospheric Measurement Techniques, 8*, 505–521. https://doi.org/10.5194/amt-8-505-2015

Tang, I. N., & Munkelwitz, H. R. (1994). Water activities, densities, and refractive indices of aqueous sulfates and sodium nitrate droplets of atmospheric importance. *Journal of Geophysical Research, 99*(18), 801–808. https://doi.org/10.1029/94JD01345

van Donkelaar, A., M. R. V., Brauer, M., & Boys, B. L. (2015). Use of satellite observations for long-term exposure assessment of global concentrations of fine particulate matter. *Environmental Health Perspectives, 123*, 135–143. https://doi.org/10.1289/ehp.1408646

Vidal, J. (2016). Environmental sustainability. https://www.theguardian.com/global-development/2016/oct/20/air-pollution-deadlier-africa-than-dirty-water-or-malnutrition-oe.cd. Accessed 20 Jun 2017.

Vedal, S., Han, B., Xu, J., Szpiro, A., & Bai, Z. (2017). Design of an air pollution monitoring campaign in Beijing for application to cohort health studies. *International Journal of Environmental Research and Public Health, 14*, 1580. https://doi.org/10.3390/ijerph14121580

Wang, J., & Christopher, S. A. (2003). Intercomparison between satellite-derived aerosol optical thickness and PM2.5 mass: implications for air quality studies. *Geophysical Research Letters, 30*(21), 2095. https://doi.org/10.1029/2003GL018174

Wang, J., & Martin S.T (2007). Satellite characterization of urban aerosols: Importance of including hygroscopicity and mixing state in the retrieval algorithms. *Journal of Geophysical Research: Atmospheres, 112*, D17203. https://doi.org/10.1029/2006JD008078

Wang, Y., Jiang, H., Zhang, S., Xu, J., Lu, X., Jin, J., & Wang, C. (2016). Estimating and source analysis of surface PM2.5 concentration in the Beijing-Tianjin-Hebei region based on MODIS data and air trajectories. *International Journal of Remote Sensing, 37*(20), 4799–4817. https://doi.org/10.1080/01431161.2016.1220031

Weber, S. A., Engel-Cox, J. A., Hoff, R. M., Prados, A. I., & Zhang, H. (2010). An improved method for estimating surface fine particle concentrations using seasonally adjusted satellite aerosol optical depth. *Journal of the Air & Waste Management Association, 60*, 574–585. https://doi.org/10.3155/1047-3289.60.5.574

WHO. (2017). Air Pollution. World Health Organization website. https://www.afro.who.int/health-topics/airpollution. Accessed 30 April, 2018

World Health Organization (WHO). (2013). Public Health and Environment (PHE). Database: outdoor air pollution in cities. http://www.who.int/phe/health_topics/outdoorair/databases/en/. Accessed 26 Jul 2018.

World Health Organization (WHO). (2014). Burden of Disease from Ambient Air Pollution for 2012 Summary of Results. http://www.who.int/phe/health_topics/outdoorair/databases/AAP_BoD_results_March2014.pdf. Accessed 30 Apr 2018.

Zhang, W. J., Sun, Y. L., Zhuang, G. S., & Xu, D. Q. (2006). Characteristics and seasonal variations of PM2.5, PM10, and TSP aerosol in Beijing. *Biomedical and Environmental Sciences, 19*(6):461–8. PMID: 17319272.

Zhang, H., Wang, S., Hao, J., Wang, X., Wang, S., Chai, F., & Li, M. (2015). Air pollution and control action in Beijing. *Journal of Cleaner Production*. https://doi.org/10.1016/j.jclepro.2015.04.092

Zheng, Y., Zhang, Q., Liu, Y., Geng, G., & He, K. (2016). Estimating ground-level PM2.5 concentrations over three megalopolises in China using satellite-derived aerosol optical depth measurements. *Atmospheric Environment, 124*, 232–242. https://doi.org/10.1016/j.atmosenv.2015.06.046

Zhu, T. (2017). Air pollution in China: scientific challenges and policy implications. *National Science Review, 4*, 800. https://doi.org/10.1093/nsr/nwx151

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.