Scientific Research Publishing

Journal of Applied Mathematics and Physics, 2024, 12, 256-267
https://www.scirp.org/journal/jamp
ISSN Online: 2327-4379
ISSN Print: 2327-4352

Modeling and Characterization of Fine Particulate Matter Dynamics in Bujumbura Using Low-Cost Sensors

Egide Ndamuzi[1,2], Rachel Akimana[2], Paterne Gahungu[3], Elie Bimenyimana[4]

[1]Doctorate School of the University of Burundi, Bujumbura, Burundi
[2]Department of Physics, University of Burundi, Bujumbura, Burundi
[3]Department of Mathematics, Imperial College London, London, UK
[4]Climate and Atmosphere Research Centre (CARE-C), The Cyprus Institute, Nicosia, Cyprus
Email: paterne.gahungu@ub.edu.bi

How to cite this paper: Ndamuzi, E., Akimana, R., Gahungu, P. and Bimenyimana, E. (2024) Modeling and Characterization of Fine Particulate Matter Dynamics in Bujumbura Using Low-Cost Sensors. *Journal of Applied Mathematics and Physics*, **12**, 256-267.
https://doi.org/10.4236/jamp.2024.121020

Received: December 19, 2023
Accepted: January 28, 2024
Published: January 31, 2024

Abstract

Air pollution is a result of multiple sources including both natural and anthropogenic activities. The rapid urbanization of the cities such as Bujumbura, economic capital of Burundi, is one of these factors. The very first characterization of the spatio-temporal variability of $PM_{2.5}$ in Bujumbura and the forecasting of $PM_{2.5}$ concentration have been conducted in this paper using data collected during a year, from August 2022 to August 2023, by low-cost sensors installed in Bujumbura city. For each commune, an hourly, daily and seasonal analysis was carried out and the results showed that the mass concentrations of $PM_{2.5}$ in the three municipalities differ from one commune to another. The average hourly and annual $PM_{2.5}$ concentrations exceed the World Health Organization standards. The range is between 28.3 and 35.0 μg/m^3. In order to make a prediction of $PM_{2.5}$ concentration, an investigation of Recurrent Neural Networks with Long Short-Term Memory has been undertaken.

Keywords

Particulate Matter, Recurrent Neural Networks, Calibration

1. Introduction

Air pollution remains one of the high threats to global health and environment. Household combustion, motor vehicles, industrial facilities, waste burning and forest fires are common sources of air pollution. $PM_{2.5}$ can enter the bloodstream, primarily resulting in cardiovascular and respiratory diseases [1]-[7]. According to the WHO, there are 4.2 million premature deaths every year as a result of exposure to ambient (outdoor) air pollution and 3.8 million for household (indoor) exposure, primarily related to smoke from cooking and heating. Approximately 90% of these deaths are estimated to occur in low- and middle-income countries, particularly in Sub-Saharan Africa. The estimates from WHO show that 80% of people living in urban areas are exposed to air pollution, threatening lives, productivity and economies. Efforts to reduce air pollution across the East African region have been undermined by the lack of air quality oversight in key areas. A few reference monitors are deployed in some parts of Sub-Saharan Africa. Air quality monitoring is currently guided by the use of low-cost sensors in many regions [3] [8]. Low-cost particulate matter sensors are becoming more widely available and are being increasingly deployed in ambient and indoor environments because of their low cost and ability to provide high spatial and temporal resolution PM information [9].

2. Materials and Methods

2.1. Study Area

This study was carried out in Bujumbura (3˚21'40.536"S, 29˚20'52.4976"E; 774 m asl), the largest and economic capital city of Burundi, located on the coast of Lake Tanganyika (**Figure 1(a)**). The city of Bujumbura is characterized by rather stable temperatures throughout the year ranging from 19˚C to 29˚C and about 835mm of rainfall annually [10] [11]. An annual rainfall peak is found in the two months of April and May, where rainfall often reaches or exceeds 90 mm per month, and a minimum from June to August, where rainfall is rare and sporadic [11]. Four seasons characterize Bujumbura city: the long dry season (June-August), Short wet season (September-December), Short dry season (mid January-mid February), and long wet season (February-May). Slight seasonal variations throughout the year characterize the average hourly wind speed in Bujumbura city: the windiest part of the year (April-October) and the calmest wind period of the year (October-April). As for wind direction, a predominant hourly average varies throughout the year in Bujumbura most often, on one hand from the south (January-February; June-September) and on the other hand most from the east (February-June; September-January) [12].

Bujumbura is one of the most densely populated cities in Burundi with 11,668 inhabitants per km^2. The city's rapid urbanization has shifted disproportionately between the growth of population and urbanized area [13], which is associated with increase in air pollution levels due to anthropogenic activities.

Three low-cost sensors have been installed in Bujumbura at different locations within the city. These three sites were selected to represent the main sources of air pollution in the city of Bujumbura, such as residential areas, industrial zones and business centres. The LCSs were installed in the urban commune of Mukaza (3˚19'12"S, 29˚22'19.2"E), in Ntahangwa (3˚20'57.768"S, 29˚21'0"E) and in Muha (3˚26'24"S, 29˚22'19.2"E) (**Figure 1(b)**).

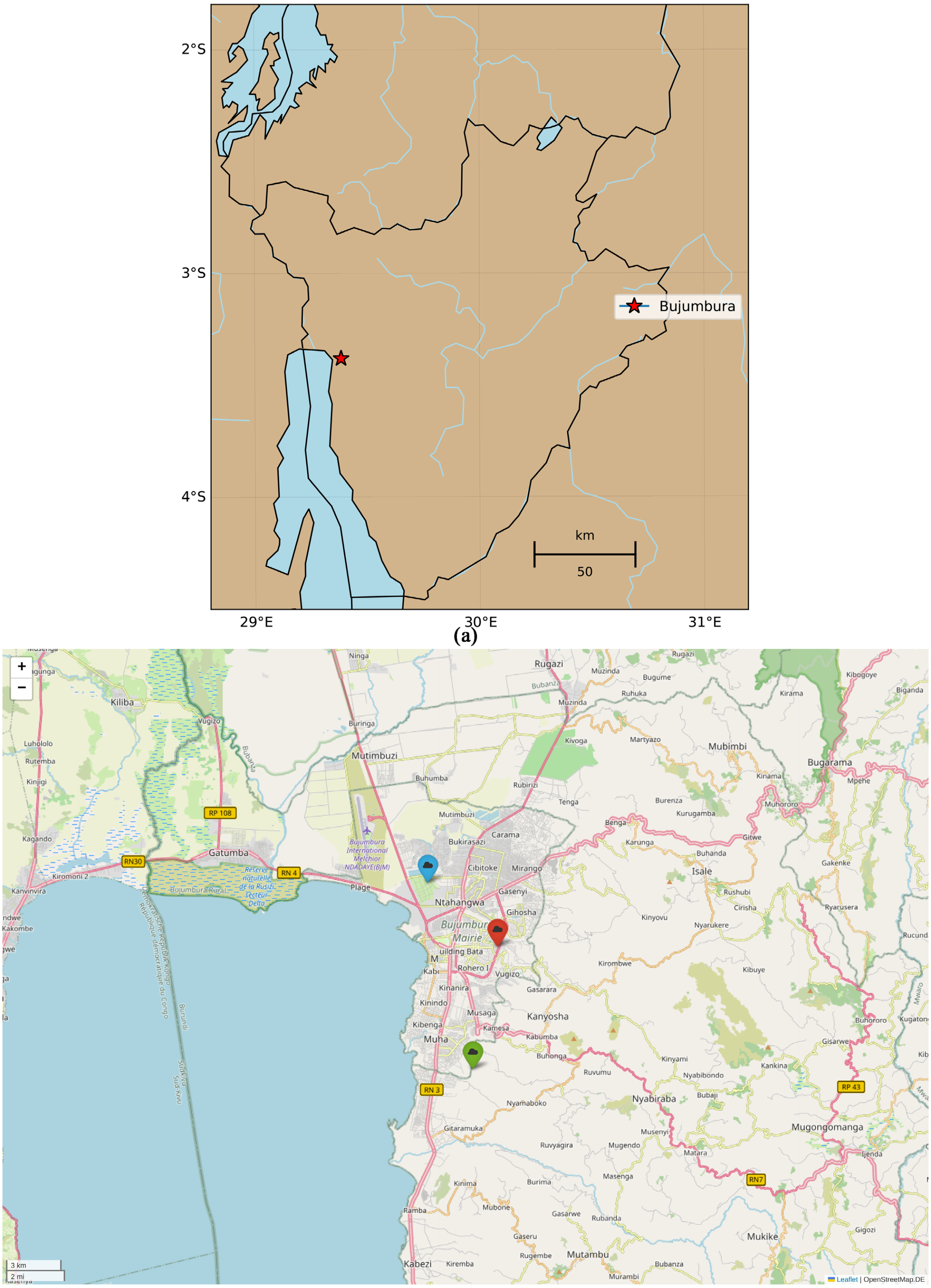

Figure 1. The map of Burundi showing the location of the city of Bujumbura in Burundi (a) and PM measurement sites in Bujumbura (b).

2.2. Instrumentation

Particulate matter measurements were conducted for a complete year (August 2022 to August 2023) in order to capture seasonal variability of PM concentrations in Bujumbura. PM concentrations were monitored by PurpleAir air quality sensors (PA-II-FLEX) [14]. Plantower PMS 5003 optical sensors and a Bosch BME280 sensor are used to estimate $PM_{2.5}$ mass concentrations and temperature, relative humidity, and pressure respectively [2] [15]. The functionality of these devices allows transmission of data via wireless connectivity in real time and recorded to an on-board micro SD card. The devices measure sensor readings in six size bins ranging from 300 nm to 10 μm at approximately a 60-s interval. A proprietary algorithm, sometimes CF = ATM algorithm, converts raw sensor measurements to PM mass using assumptions about particle shape and density [2].

2.3. Data Correction

Despite the availability of low-cost particle sensors, their limited accuracy means that the data can not be used reliably without correction. The common method used for data correction (calibration) is to co-locate the LCSs with reference monitors [16]. Due to absence of reference instruments, we opted to use a transparent and reproducible alternative method ALT CF3 of calculating $PM_{2.5}$ from the number of particles in three size categories. Note that this approach was applied successfully in a couple of studies [17] [18]. Basically, the PM concentrations data collected by PurpleAir low-cost sensors is corrected using the particle concentrations in three size bins as expressed in Equation (1):

$$\mathrm{PM}_{2.5} = 3\left(\alpha X + \beta Y + \gamma Z\right), \tag{1}$$

where the number 3 is the Calibration Factor (CF) [17]. X, Y and Z are particle numbers per deciliter in three size categories, i.e. X = 0.3 μm - 0.5 μm, Y = 0.5 μm - 1.0 μm and Z = 1.0 μm - 2.5 μm, given directly by the PurpleAir LCSs. The estimate mass concentration in each size category, α, β and γ respectively is given as a product of water density (1 g/cm³), the particle volume ($V = \frac{4}{3}\pi r^3$) and the number of particle (N); $r = \frac{d}{2}$ with d the geometric mean of each size boundaries. It can also be approximated by the midpoint. Hence, $\alpha = 0.00030418$, $\beta = 0.0018512$ and $\gamma = 0.02069706$.

2.4. Mathematical Modeling

Linear models, sometimes are not sufficient to capture the real-world phenomena, thus nonlinear models are necessary [19]. The Artificial Neural Networks (ANNs) are among of nonlinear models used for nonlinear regression and classification tasks. Neural networks architectures have been widely used to perform forecasting tasks. In some cases, Neural Networks are viable competitors for various traditional time series models [3] [20] [21]. A recurrent neural network is adapted to work for time series data or data that involves sequences. RNNs have the concept of memory that helps them store the states or information of previous inputs to generate the next output of the sequence, *i.e.* in RNN, the information cycles through the loop in the hidden layer such that the information derived from earlier inputs remains in the network [22]. The input layer at time step *t*, $x_t \in \mathbb{R}$ such that we can extend it to d-dimensional feature vector. Then, x_t processes the initial input and passes it to the middle layer h_t which consists of multiple hidden layers, each with its activation functions, weights, and biases. In simplest terms, Equations (2) and (3) define how an RNN evolves over time. The output y at time t is computed as:

$$y_t = f\left(h_t, w_y\right) = f\left(w_y \cdot h_t + b_y\right) \tag{2}$$

where $\cdot$ is the dot product, $w_y \in \mathbb{R}^m$ are weights associated with hidden to output units with m number of hidden layers, and $b_y \in \mathbb{R}$ is the bias associated with the feedforward layer. The RNN can remember its past by allowing past computations h_{t-1} to influence the present computations h_t.

$$h_t = g\left(x_t, h_{t-1}, w_x, w_h, b_h\right) = g\left(w_x x_t + w_h h_{t-1} + b_h\right), \tag{3}$$

where $w_x \in \mathbb{R}^m$ are weights associated with inputs in recurrent layer and $b_h \in \mathbb{R}^m$ is the bias associated with the recurrent layer. f and g are activation functions. In the recurrent neural network, any activation function we like in t time can be used. Long Short-Term Memory (LSTM) are a special kind of RNN, capable to overcome the vanishing/exploding gradient problem, so RNNs can safely be applied to extremely long sequences. The accuracy of the model is based on the metrics such as Root Mean Square Error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} \tag{4}$$

and Absolute Mean Error (AME):

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \tag{5}$$

where n denotes the total number of observations, y_i denotes the observed values, and $\hat{y}_i$ denotes the predicted values.

In this work, we have used the "mean-squared-error" loss function in tensorflow, "Adam" optimizer and "sigmoid" activation function. We have split the data with 80% for training and 20% for validation.

3. Results and Discussions

The first-ever characterization of the spatio-temporal variability of $PM_{2.5}$ in Bujumbura using low-cost sensors between August 2022 to August 2023 revealed that $PM_{2.5}$ mass concentrations in the municipalities of Bujumbura differ from one commune to another. A very high $PM_{2.5}$ annually mean concentration was observed in Muha (35.0 μg/m^3), and the lowest in Ntahangwa (28.3 μg/m^3). For Mukaza, 32.8 μg/m^3 $PM_{2.5}$ annually mean concentration was observed.

3.1. Comparison between CF1 and ALT CF3 Algorithms

Plantower algorithm (CF1) was compared with the ALT CF3 approach for the three sites (Ntahangwa, Muha and Mukaza) and the corresponding results are shown in **Figure 2**. Overall good agreement between the two algorithms was observed in terms of temporal trends as demonstrated by strong correlations between the two methods. However, higher concentrations were observed for CF1 compared to ALT CF3 (almost two times higher). This is in accordance with what is reported in previous studies where an overestimate of $PM_{2.5}$ concentrations of approximately 40% were found for the PurpleAir LCSs compared to the reference instruments [16] [17] [23]. In fact, estimates provided by Plantower [24], the manufacturer of the sensors used in PurpleAir monitors using the (CF1 or CF-ATM) method was inferior in several respects [4] [9] (lower precision, higher detection limit, less improved size distribution), $PM_{2.5}$ concentrations overestimation compared to this alternative. Hence, the precision of LCSs is still not comparable to reference-grade measurements [25].

3.2. Diurnal Variability of $PM_{2.5}$

Diurnal cycles of $PM_{2.5}$ concentrations across the three sites are shown in **Figure 3** where the error bars represent the standard deviation. Mean hourly $PM_{2.5}$ mass concentrations of 32.9 μg/m³, 24.2 μg/m³ and 31.4 μg/m³ have been observed in Mukaza, Ntahangwa and Muha respectively.

As shown in **Figure 3**, two peaks of $PM_{2.5}$ were observed at the three sites in the morning and evening. In general, the city of Bujumbura has traffic in the form of a transport network oriented towards the city center which is particularly noticeable during rush hours. For the Muha and Ntahangwa, a $PM_{2.5}$ morning peak can generally be due to the road traffic. The population of these two municipalities often goes at time to the city center, *i.e.* towards Mukaza, a center considered as of business, commerce and administration. For the Mukaza, the $PM_{2.5}$ morning peak can also be associated with the road traffic, *i.e.* the arrival of those who come from these two other municipalities. According to the

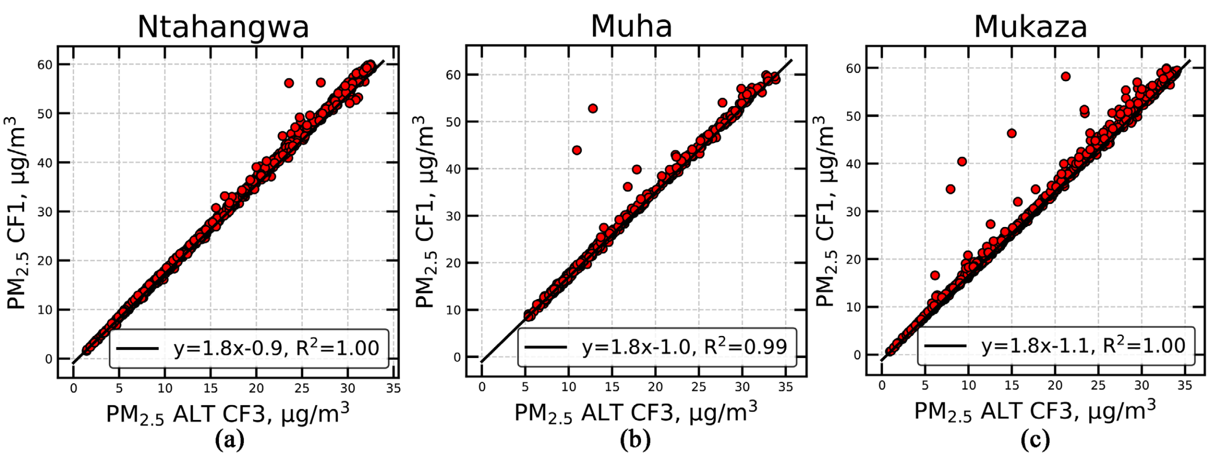

Figure 2. ALT CF3 algorithm vs CF1 Plantower algorithm.

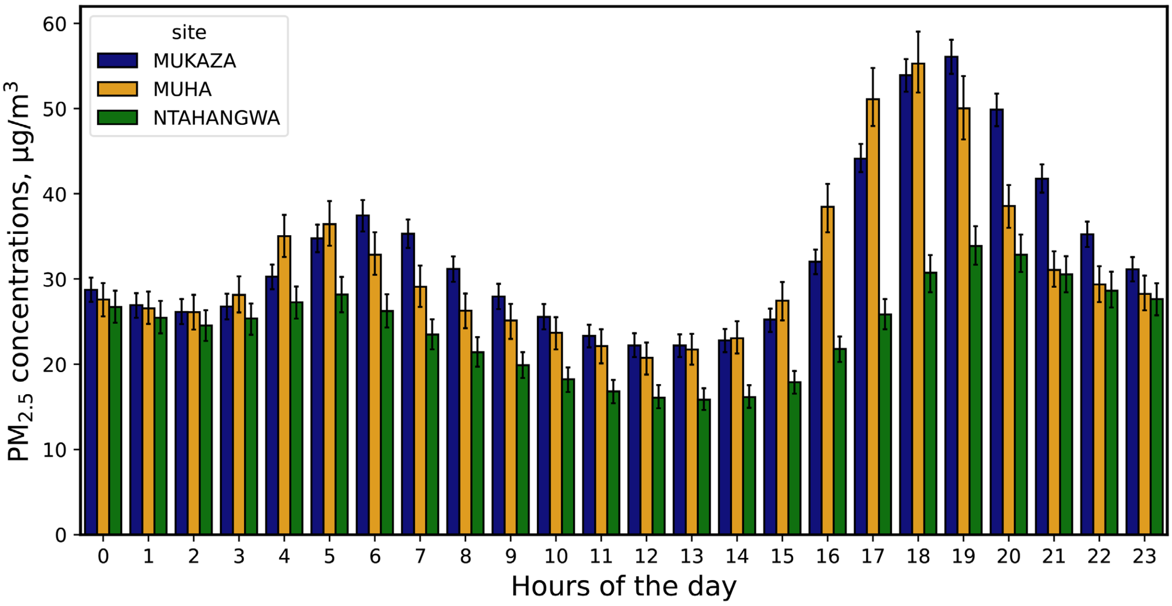

Figure 3. Hourly mean value of $PM_{2.5}$ mass concentrations in Bujumbura.

study carried out by JICA (Japanese International Cooperation Agency) [26], 80% of vehicles leave the outskirts every morning heading towards the Bujumbura city center.

After the morning peaks, the concentration of $PM_{2.5}$ decreases. The other peak is observed towards the evening which can also be associated with traffic, the return of those who have been at work and domestic activities such as cooking. Weather conditions such as wind direction could also explain the observed $PM_{2.5}$ diurnal variation given that Bujumbura is a coastal city and is likely to be affected by sea breeze and land breeze phenomena.

3.3. Seasonal Variability of $PM_{2.5}$

The seasonal variability of $PM_{2.5}$ mass concentrations is shown in **Figure 4**. It is shown that the $PM_{2.5}$ mass concentrations increased significantly during the dry season and decreased during the rainy season. Several factors may explain this phenomenon. During the dry season, there is an increase in re-suspended dust as a result of dry surface and higher wind speed, as well as other human activities such as bush fires that are likely to occur during dry period. The slight increase in $PM_{2.5}$ mass concentrations observed around February-March is probably due to burning of agricultural residues since at that time people are preparing their fields for the main growing season.

3.4. Daily Mean $PM_{2.5}$ Concentrations

As shown in **Figure 5**, the mean daily mass concentrations of $PM_{2.5}$ indicate that WHO guidelines of 25 μg/m^3 as a 24-hour mean was exceeded in all municipalities of Bujumbura during the dry seasonal. Considering the $PM_{2.5}$ annually mean mass concentrations, it is shown also that air PM pollution is higher in Muha

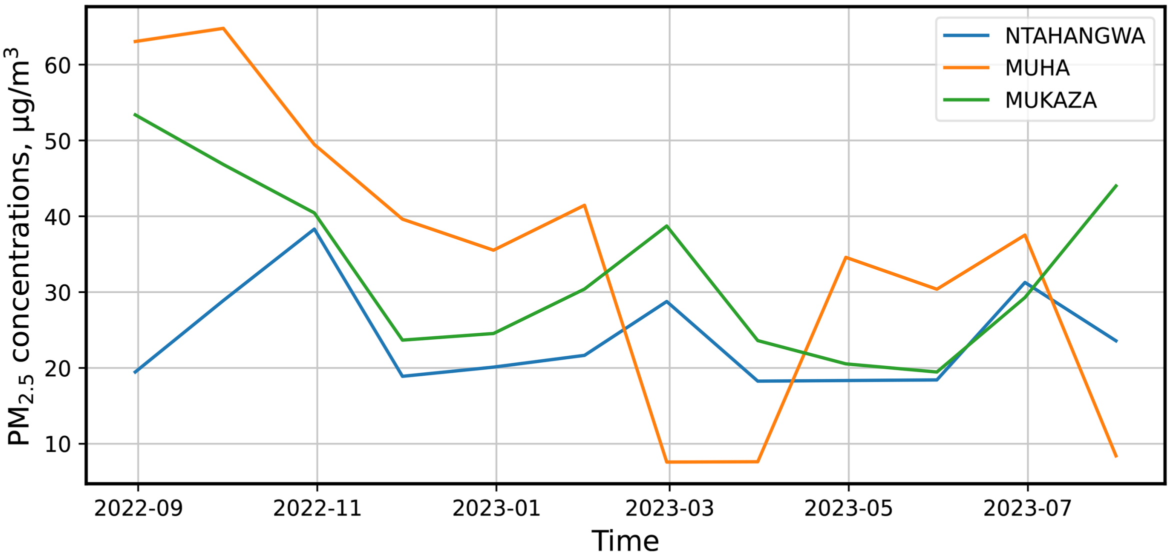

Figure 4. Seasonal variability of $PM_{2.5}$ mass concentrations in Bujumbura.

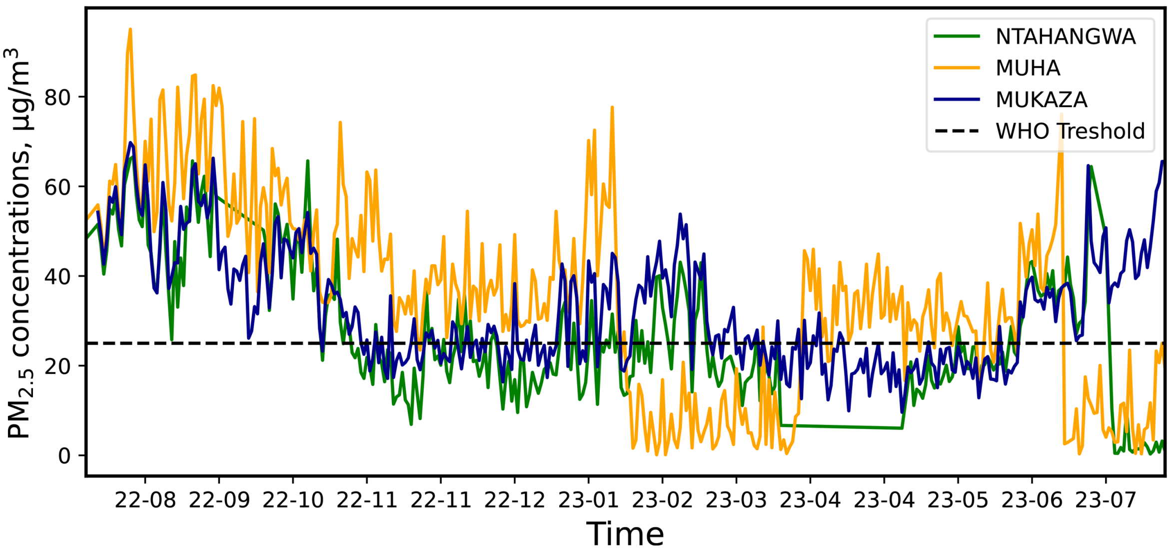

Figure 5. Daily variability of $PM_{2.5}$ mass concentrations in Bujumbura.

than in Mukaza and Ntahangwa municipalities.

The higher $PM_{2.5}$ concentrations in Muha compared to other municipalities can be explained by several factors. In Mukaza and part of Ntahangwa, there are public trash cans in some places for better waste management, which is not the case for Muha and that causes waste burning in this locality. Mukaza is largely commercial, part of Ntahangwa is industrial, while Muha is largely residential where wood or charcoal is largely fuel used for cooking. Furthermore, road dust re-suspension can be very important in Muha compared to the other sites given the lower quality of roads networks in this locality. In addition, the Muha commune is bordered with the crop fields of non-urban communes such as Kanyosha and Kabezi, which contributes to air pollution with field waste fires. The time series of the three sites were also compared between them using Pearson correlation as shown in **Figure 6**. Overall, weak to moderate correlation was observed suggesting that those sites are affected by different sources. Mukaza was the least correlated with other sites indicating different more localized emission sources in that municipality.

3.5. Forecasting $PM_{2.5}$ Concentrations Using Long Short-Term Memory Model

Forecasting air pollution in Bujumbura with Long Short-Term Memory Model has been carried out and results are given in **Figure 7**. $PM_{2.5}$ hourly mean concentrations are used to train the model. Model evaluation is given at each site by root mean square error and absolute mean error. The RMSE is in range of 8.2 μg/m^3 to 12.8 μg/m^3 and the MAE is in range of 5.3 μg/m^3 to 6.8 μg/m^3. This shows that the recurrent neural network method is a potential forecasting model in Bujumbura for $PM_{2.5}$ time series data. A good correlation is obtained with r^2, in the range of 0.7 to 0.8.

4. Conclusion

The very first characterization of spatio-temporal variability of $PM_{2.5}$ in Bujumbura has been explored. Hourly, daily and seasonal analysis was carried out on the three communes of Bujumbura. The analysis has shown that the $PM_{2.5}$ concentration is location-dependent in the study area and time of the day. The study has given qualitative indications about potential main sources in Bujumbura such as the transport and cooking activities. The use of electric vehicles and clean cooking energy could reduce significantly the levels of pollution in Bujumbura. An investigation of Recurrent Neural Networks—Long Short-Term Memory has given good performance as a potential forecasting model in Bujumbura. Future

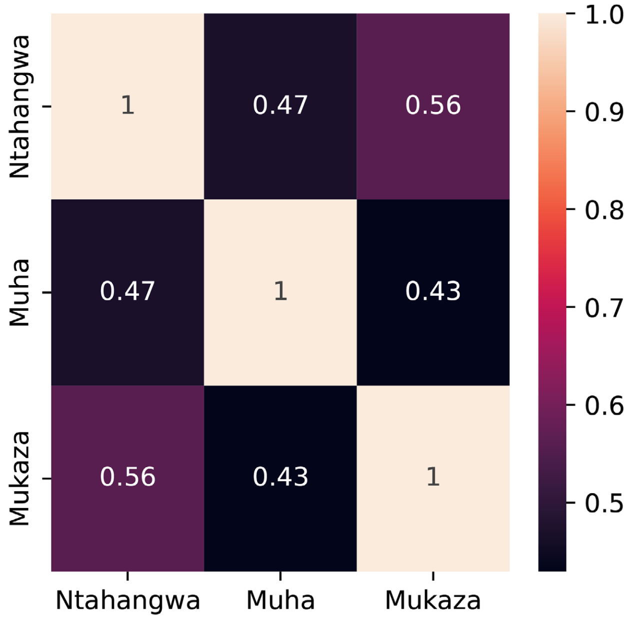

Figure 6. Correlation of $PM_{2.5}$ mass concentrations in Bujumbura municipalities.

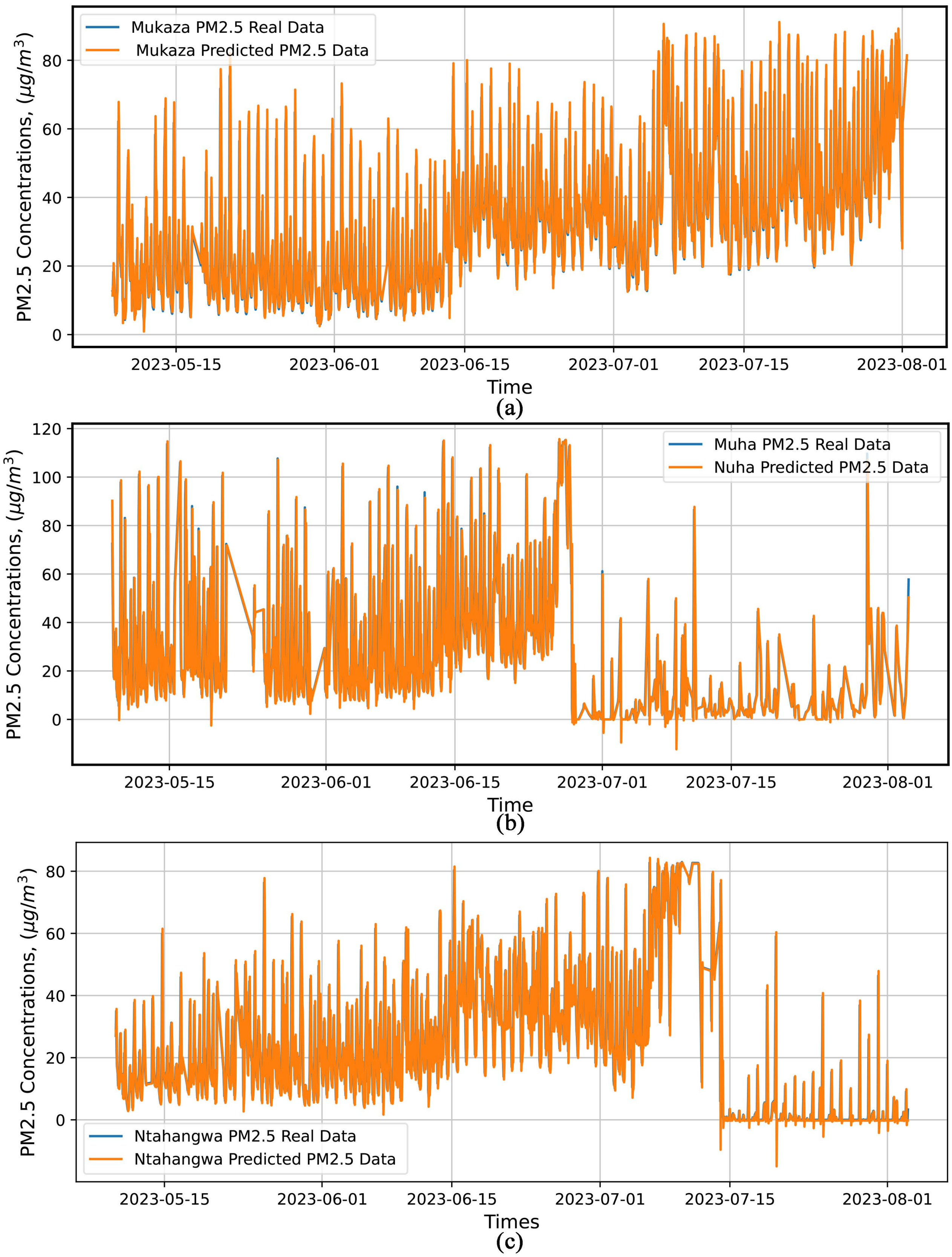

Figure 7. Bujumbura $PM_{2.5}$ forecasting. (a) Mukaza $PM_{2.5}$ forecasting; (b) Muha $PM_{2.5}$ forecasting; (c) Ntahangwa $PM_{2.5}$ forecasting.

work will explore the impact of air pollution exposure on health in Bujumbura.

Acknowledgments

The author, Egide NDAMUZI, would like to thank the African Institute for Mathematical Sciences, https://www.nexteinstein.org/, with financial support from the Government of Canada, provided through Global Affairs Canada, https://www.international.gc.ca/, and the International Development Research Centre, https://idrc-crdi.ca/en who funded this work.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Health Effects Institute (2020) State of Global Air 2019. https://www.stateofglobalair.org/report

[2] McFarlane, C., Isevulambire, P.K., Lumbuenamo, R.S., Ndinga, A.M.E., *et al.* (2021) First Measurements of Ambient $PM_{2.5}$ in Kinshasa, Democratic Republic of Congo and Brazzaville, Republic of Congo Using Field-Calibrated Low-Cost Sensors, *Aerosol and Air Quality Research*, **21**, Article ID: 200619. https://doi.org/10.4209/aaqr.200619

[3] Gahungu, P. and Kubwimana, J.R. (2022) Trend Analysis and Forecasting Air Pollution in Rwanda. ArXiv: 2205.10024.

[4] Gupta, P., Doraiswamy, P., Levy, R., Pikelnaya, O., Maibach, J. and Feenstra, B. (2018) Impact of California Fires on Local and Regional Air Quality: The Role of a Low-Cost Sensor Network and Satellite Observations. *GeoHealth*, **2**, 172-181. https://doi.org/10.1029/2018GH000136

[5] Paterne, G., Jean, R.K., Lionel, B.M. and Egide, N. (2022) Modelling Spatio-Temporal Trends of Air Pollution in Africa. ArXiv: 2208.12719v1.

[6] Cohen, A.J., Brauer, M., Burnett, R., Anderson, H.R., Frostad, J., Estep, K., *et al.* (2017) Estimates and 25-Year Trends of the Global Burden of Disease Attributable to Ambient Air Pollution: An Analysis of Data from the Global Burden of Diseases Study 2015. *The Lancet*, **389**, 1907-1918. https://doi.org/10.1016/S0140-6736(17)30505-6

[7] Ling, S.H. and van Eeden, S.F. (2009) Particulate Matter Air Pollution Exposure: Role in the Development and Exacerbation of Chronic Obstructive Pulmonary Disease. *The International Journal of Chronic Obstructive Pulmonary Disease*, **4**, 233-243. https://doi.org/10.2147/COPD.S5098

[8] Stavroulas, I., Grivas, G., Michalopoulos, P., Liakakou, E., Bougiatioti, A., Kalkavouras, P., *et al.* (2020) Field Evaluation of Low-Cost PM Sensors (Purple Air PA-II) under Variable Urban Air Quality Conditions, in Greece. *Atmosphere*, **11**, Article 926. https://doi.org/10.3390/atmos11090926

[9] Kelly, K.E., Whitaker, J., Petty, A., Widmer, C., Dybwad, A., Sleeth, D., Martin, R. and Butterfield, A. (2017) Ambient and Laboratory Evaluation of a Low-Cost Particulate Matter Sensor. *Environmental Pollution*, **221**, 491-500. https://doi.org/10.1016/j.envpol.2016.12.039

[10] Ministry of Environment, Agriculture and Livestock (Republic of Burundi) (2019)

Third National Communication on Climate Change (TNCCC). https://bi.chm-cbd.net/sites/bi/files/2021-06/tcncc_en.pdf

[11] Rhugwasanye, C., Rao, C.H., Agarwal, S. and Rajya, K. (2022) Rainfall Trend Analysis and IDF Curve Generation for Bujumbura Mairie in East Africa. *International Journal of Special Education*, **37**, 14636-14649.

[12] https://weatherspark.com/y/95879/Average-Weather-in-Bujumbura-Burundi

[13] Kabanyegeye, H., Sikuzani, Y.U., Sambieni, K.R., Masharabu, T., Havyarimana, F. and Bogaert, J. (2020) Trente-trois ans de dynamique spatiale de l'occupation du sol de la ville de Bujumbura, République du Burundi. *Afrique SCIENCE*, **18**, 203-215.

[14] https://www2.purpleair.com/

[15] Ardon-Dryer, K., Dryer, Y., Williams, J.N. and Moghimi, N. (2020) Measurements of $PM_{2.5}$ with PurpleAir under Atmospheric Conditions. *Atmospheric Measurement Techniques*, **13**, 5441-5458. https://doi.org/10.5194/amt-13-5441-2020

[16] Li, J.Y., Mattewal, S.K., Patel, S. and Biswas, P. (2020) Evaluation of Nine Low-Cost-Sensor-Based Particulate Matter Monitors. *Aerosol and Air Quality Research*, **20**, 254-270. https://doi.org/10.4209/aaqr.2018.12.0485

[17] Wallace, L., Bi, J.Z., Ott, W.R., Sarnat, J. and Liu, Y. (2021) Calibration of Low-Cost PurpleAir Outdoor Monitors Using an Improved Method of Calculating $PM_{2.5}$. *Atmospheric Environment*, **256**, Article ID: 118432. https://doi.org/10.1016/j.atmosenv.2021.118432

[18] Wallace, L. (2022) Inter-Comparison of PurpleAir Sensor Performance over Three Years Indoors and Outdoors at a Home: Bias, Precision, and Limit of Detection Using an Improved Algorithm for Calculating $PM_{2.5}$. *Sensors*, **22**, Article No. 2755. https://doi.org/10.20944/preprints202202.0142.v1

[19] Computer Science Department University of Toronto (2012) Machine Learning and Data Mining. https://www.dgp.toronto.edu/~hertzman/411notes.pdf

[20] Khashei, M. and Bijari, M. (2010) An Artificial Neural Network (p, d, q) Model for Time Series Forecasting. *Expert Systems with Applications*, **20**, 479-489. https://doi.org/10.1016/j.eswa.2009.05.044

[21] Chen, Y., Yang, B., Dong, J. and Abraham, A. (2005) Time-Series Forecasting Using Flexible Neural Tree Model. *Information Sciences*, **174**, 219-235. https://doi.org/10.1016/j.ins.2004.10.005

[22] Wu, Z. and Christofides, P.D. (2019) Economic Machine-Learning-Based Predictive Control of Nonlinear Systems. *Mathematics*, **7**, Article 494. https://doi.org/10.3390/math7060494

[23] Barkjohn, K.K., Gantt, B. and Clements, A.L. (2021) Development and Application of a United States-Wide Correction for $PM_{2.5}$ Data Collected with the PurpleAir Sensor. *Atmospheric Measurement Techniques*, **14**, 4617-4637. https://doi.org/10.5194/amt-14-4617-2021

[24] https://www.plantower.com/en/

[25] Bi, J., Wildani, A., Chang, A.H.H. and Liu, Y. (2020) Incorporating Low-Cost Sensor Measurements into High-Resolution $PM_{2.5}$ Modeling at a Large Spatial Scale. *Environment Scientific Technology*, **54**, 2152-2162. https://doi.org/10.1002/essoar.10501744.1

[26] Ministry of Transport, Posts & Telecommuniccation and Ministry of Public Works and Equipment (Republic of Burundu) and Japan International Cooperationagency (2008) The Emergency Study on Urban Transport in Bujumbura. https://openjicareport.jica.go.jp/pdf/11882628.pdf