Supplementary Data

2.0 Supplementary Data: Method

2.4 Data Preprocessing

The raw data from AirQo and TAHMO for $PM_{2.5}$ and meteorological variables were processed and prepared for subsequent analysis. The data manipulation and pre-processing were performed using Python version 3.10.11(73) and several libraries, including Pandas version 2.0.2, NumPy version 1.23.5, and Matplotlib version 3.7.0.

2.4.1 $PM_{2.5}$ Data Preprocessing

The $PM_{2.5}$ data obtained from AirQo for Jinja and Kampala were analysed using separate Jupyter notebooks. Each dataset contained hourly timestamp information, $PM_{2.5}$ raw values, $PM_{2.5}$ calibrated values, and site names. Jinja had seven sites where the AirQo sensors were deployed, while Kampala had 55 unique sites. The detailed sites where the AirQo sensors were deployed in the two cities are presented in Table 1.

Only data from January 2020 to December 2022 were considered to ensure consistency and reliability. Data from 2019 were excluded due to limited availability and potential data inconsistencies. The next step involved averaging the hourly $PM_{2.5}$ data across all sites within each city, resulting in city-level $PM_{2.5}$ concentrations. This aggregation was performed because the analysis focused on the overall $PM_{2.5}$ levels in each city.

After this, the percentage of missing values was calculated for both Jinja and Kampala datasets. The missing data statistics are recorded in Table 2. The Multiple Imputation by Chained Equation (MICE) algorithm was utilized to impute the missing values. MICE is a multivariate imputation method that takes into account the relationships among variables in the dataset. It generates multiple plausible imputations for missing values by fitting regression models for the incomplete variable using the remaining variables as predictors(74). The imputation process is performed iteratively, resulting in a completed dataset with imputed values. MICE is particularly useful in handling missing data and can handle missingness up to 30%(75, 76).

2.4.2 Meteorological Data Preprocessing

The raw meteorological data obtained from TAHMO for Jinja and Kampala were analysed using separate Jupyter notebooks. Each dataset contained hourly meteorological measurements of atmospheric pressure (kPA), precipitation (mm), temperature average (°C), wind gusts max (m/s), wind speed(m/s), and relative humidity. To ensure consistency with the PM2.5 data, only data from January 2020 to December 2022 were considered, excluding 2019. Some meteorological parameters, including radiation average (W/m), temperature min (°C), temperature max (degrees Celsius), and wind direction (°), were excluded from the datasets as they were deemed unnecessary for the analysis. The percentage of missing values was calculated to handle missing data in the meteorological datasets, similar to the $PM_{2.5}$ data. The missing data statistics are presented in Table 2. The MICE algorithm was then employed to impute the missing values, ensuring a comprehensive dataset for subsequent analysis.

2.4.3 Integration of $PM_{2.5}$ and Meteorological Data

After completing the individual pre-processing steps for both cities' $PM_{2.5}$ and meteorological data, the datasets were joined on the timestamps. The integrated dataset now consisted of $PM_{2.5}$ calibrated values and the various meteorological variables aligned in time, providing a comprehensive dataset for further analysis. The final dataset contained a matrix of 26,304 rows x 8 columns.

Supplementary File SF 1: Sites of AirQo devices in Kampala and Jinja**.**

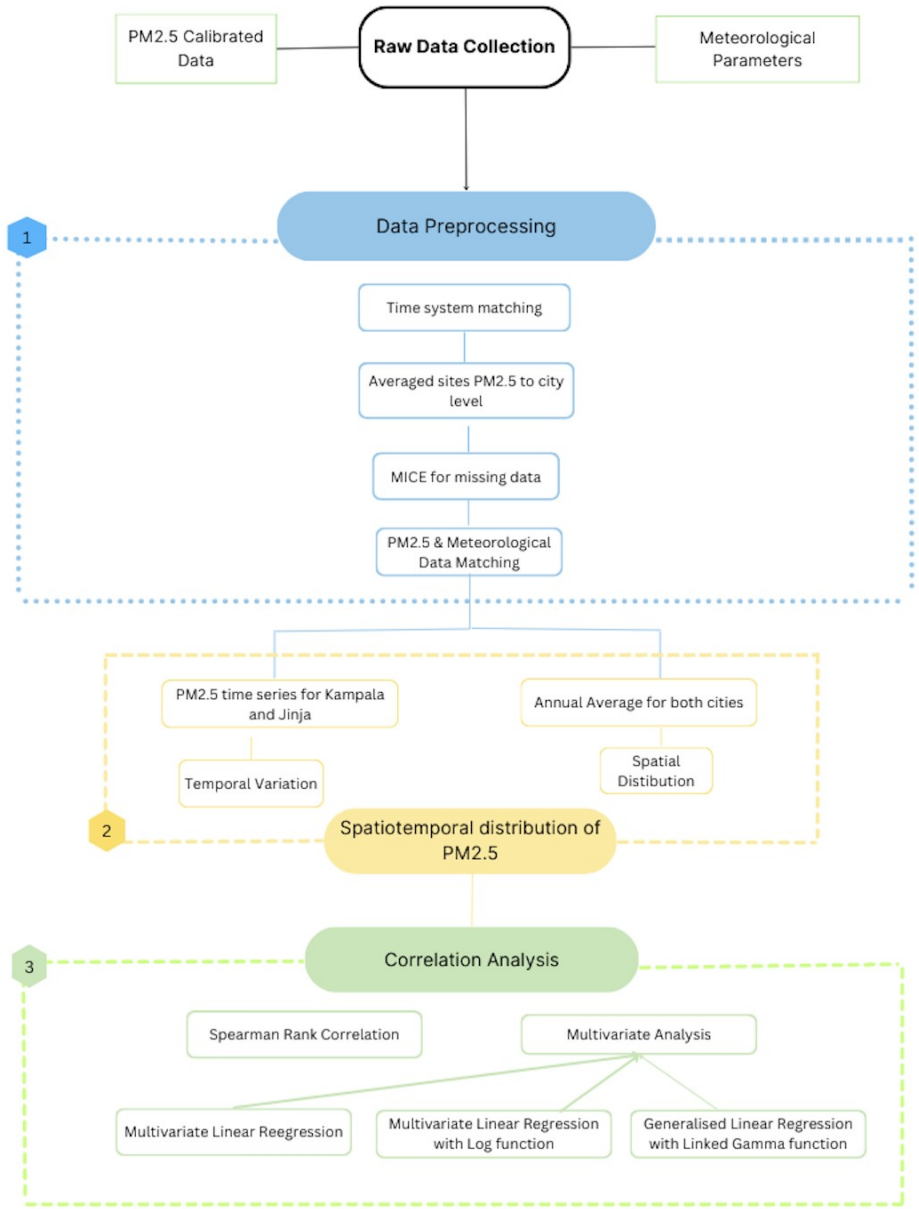
City	Sites
Jinja	Jinja Main Street, Kimaka, City Council, Walukuba East, Mpumudde, Rubaga Ward, YMCA
Kampala	Civic Centre Kampala Central', Bugolobi Nakawa, Rubaga Kampala, Kasubi Rubaga, Bukasa Kira Municipality, Bukoto I Kawempe, Nsambya/US Embassy Makindye, Kalerwe, Kawempe, Butabika Nakawa, Kiwatule Nakawa, Makerere University_01 Kawempe, Nakasero II Kampala, Mbuya II Nakawa, Banda Kampala, Luwafu Makindye, Ntinda Nakawa, Makerere University Weather Station, Bahai Kawempe, Kawempe Industrial Kawempe, Mpererwe

	Kawempe, Kyanja Nakawa, Luzira Nakawa, Rubaga, Kampala Central, Makindye, Kyebando Kawempe, Kawempe, Nakawa, Kampala Central, Makerere, Kawempe, Lower Nsooba, Kawempe, Nakawa, Nakawa, Makindye Lukuli Makindye, Lukuli Makindye, Makindye Nsambya/US Embassy, Busega Rubaga, Makindye KCCA division offices, Kibuli Makindye, Kizungu, Makindye, Ggaba Makindye, St.Ponsiano primary school Makindye, wabigalo, Kansanga Seed Secondary School, Kampala Uganda, Muyenga tank hill ,makindye, Kabalagala town Makindye, Katwe 1, Makindye, Kayan Test Site , Naguru II Nakawa, Kisugu Makindye, Kyebando Kampala, Rubaga Div Rubaga, Katwe 1 Makindye, Muyenga tank hill, Makindye, City parents primary school

Supplementary File SF 2: Missing Data Statistics for $PM_{2.5}$ and Meteorological Parameters in Kampala and Jinja

	Percentage of Missing Data for each Parameter (in %)						
Study city	$PM_{2.5}$	Atmospheric Pressure	Precipitation	Relative Humidity	Average Temperature	Wind speed	Wind gusts
Kampala	16.54	15.37	0.00	15.43	15.37	16.29	16.29
Jinja	24.86	12.59	23.15	12.61	12.57	19.31	19.31

Since the imputation is based on neighbouring values and a significant number of data points were missing in that period, the imputation resorted to using the mean value, resulting in a flat trend. To reduce the noise in the dataset and provide a clearer depiction of the trend, a 30-day moving average was calculated, as shown in Figure 5.



Supplementary File SF 3: Flowchart of the study methodology

Supplementary data: Complete Model Results

MULTIVARIATE LINEAR REGRESSION

Kampala

OLS Regression Results

Dep. Variable:	pm2_5_calibrated_value	R-squared:	0.169
Model:	OLS	Adj. R-squared:	0.169
Method:	Least Squares	F-statistic:	1339.
Date:	Sat, 15 Jul 2023	Prob (F-statistic):	0.00
Time:	00:52:06	Log-Likelihood:	-1.1217e+05
No. Observations:	26304	AIC:	2.244e+05
Df Residuals:	26299	BIC:	2.244e+05
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	208.6426	3.015	69.206	0.000	202.733	214.552
precipitation	-0.5126	0.091	-5.655	0.000	-0.690	-0.335
humidity	-86.8701	1.754	-49.527	0.000	-90.308	-83.432
temp_mean	-3.9473	0.080	-49.170	0.000	-4.105	-3.790
wind_speed	-16.3943	0.413	-39.711	0.000	-17.203	-15.585

Omnibus:	6112.635	Durbin-Watson:	0.233
Prob(Omnibus):	0.000	Jarque-Bera (JB):	16854.350
Skew:	1.236	Prob(JB):	0.00
Kurtosis:	6.044	Cond. No.	724.

Jinja

OLS Regression Results

Dep. Variable:	pm2_5_calibrated_value	R-squared:	0.064
Model:	OLS	Adj. R-squared:	0.064
Method:	Least Squares	F-statistic:	450.2
Date:	Sat, 15 Jul 2023	Prob (F-statistic):	0.00
Time:	00:48:46	Log-Likelihood:	-1.0556e+05
No. Observations:	26304	AIC:	2.111e+05
Df Residuals:	26299	BIC:	2.112e+05
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	102.9843	2.187	47.081	0.000	98.697	107.272
precipitation	-0.7912	0.079	-10.056	0.000	-0.945	-0.637
humidity	-40.1345	1.376	-29.171	0.000	-42.831	-37.438
temp_mean	-2.1865	0.055	-40.064	0.000	-2.293	-2.080
wind_speed	0.7491	0.074	10.128	0.000	0.604	0.894

Omnibus:	16282.613	Durbin-Watson:	0.535
Prob(Omnibus):	0.000	Jarque-Bera (JB):	294530.187
Skew:	2.668	Prob(JB):	0.00
Kurtosis:	18.501	Cond. No.	710.

MULTIVARIATE LINEAR REGRESSION WITH LOG OF TARGET VARIABLE

Kampala

OLS Regression Results

Dep. Variable:	pm2_5_calibrated_value	R-squared:	0.177
Model:	OLS	Adj. R-squared:	0.177
Method:	Least Squares	F-statistic:	1418.
Date:	Sat, 15 Jul 2023	Prob (F-statistic):	0.00
Time:	00:54:12	Log-Likelihood:	-13267.
No. Observations:	26304	AIC:	2.654e+04
Df Residuals:	26299	BIC:	2.658e+04
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	7.4381	0.070	105.978	0.000	7.301	7.576
precipitation	-0.0135	0.002	-6.412	0.000	-0.018	-0.009
humidity	-1.9676	0.041	-48.185	0.000	-2.048	-1.888
temp_mean	-0.0891	0.002	-47.680	0.000	-0.093	-0.085
wind_speed	-0.4199	0.010	-43.691	0.000	-0.439	-0.401

Omnibus:	19.786	Durbin-Watson:	0.230
Prob(Omnibus):	0.000	Jarque-Bera (JB):	22.098
Skew:	-0.019	Prob(JB):	1.59e-05
Kurtosis:	3.137	Cond. No.	724.

Jinja

OLS Regression Results

Dep. Variable:	pm2_5_calibrated_value	R-squared:	0.072
Model:	OLS	Adj. R-squared:	0.072
Method:	Least Squares	F-statistic:	511.8
Date:	Sat, 15 Jul 2023	Prob (F-statistic):	0.00
Time:	00:55:54	Log-Likelihood:	-16648.
No. Observations:	26304	AIC:	3.331e+04
Df Residuals:	26299	BIC:	3.335e+04
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	5.8014	0.074	77.896	0.000	5.655	5.947
precipitation	-0.0307	0.003	-11.461	0.000	-0.036	-0.025
humidity	-1.4058	0.047	-30.011	0.000	-1.498	-1.314
temp_mean	-0.0771	0.002	-41.477	0.000	-0.081	-0.073
wind_speed	0.0421	0.003	16.722	0.000	0.037	0.047

Omnibus:	520.733	Durbin-Watson:	0.390
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1129.053
Skew:	-0.014	Prob(JB):	6.75e-246
Kurtosis:	4.015	Cond. No.	710.

GENERALISED LINEAR REGRESSION MODEL WITH LINKED GAMMA FUNCTION

Kampala

Generalized Linear Model Regression Results

Dep. Variable:	pm2_5_calibrated_value	No. Observations:	26304
Model:	GLM	Df Residuals:	26299
Model Family:	Gamma	Df Model:	4
Link Function:	log	Scale:	0.17188
Method:	IRLS	Log-Likelihood:	-1.0879e+05
Date:	Sat, 15 Jul 2023	Deviance:	4209.6
Time:	01:01:45	Pearson chi2:	4.52e+03
No. Iterations:	100	Pseudo R-squ. (CS):	0.1837
Covariance Type:	nonrobust		

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-5.194e+10	6.09e+10	-0.853	0.394	-1.71e+11	6.74e+10
precipitation	-0.0147	0.002	-6.726	0.000	-0.019	-0.010
humidity	-1.8237	0.042	-43.167	0.000	-1.907	-1.741
temp_mean	-0.0915	0.002	-47.305	0.000	-0.095	-0.088
wind_speed	-0.3901	0.010	-39.235	0.000	-0.410	-0.371
intercept	5.194e+10	6.09e+10	0.853	0.394	-6.74e+10	1.71e+11

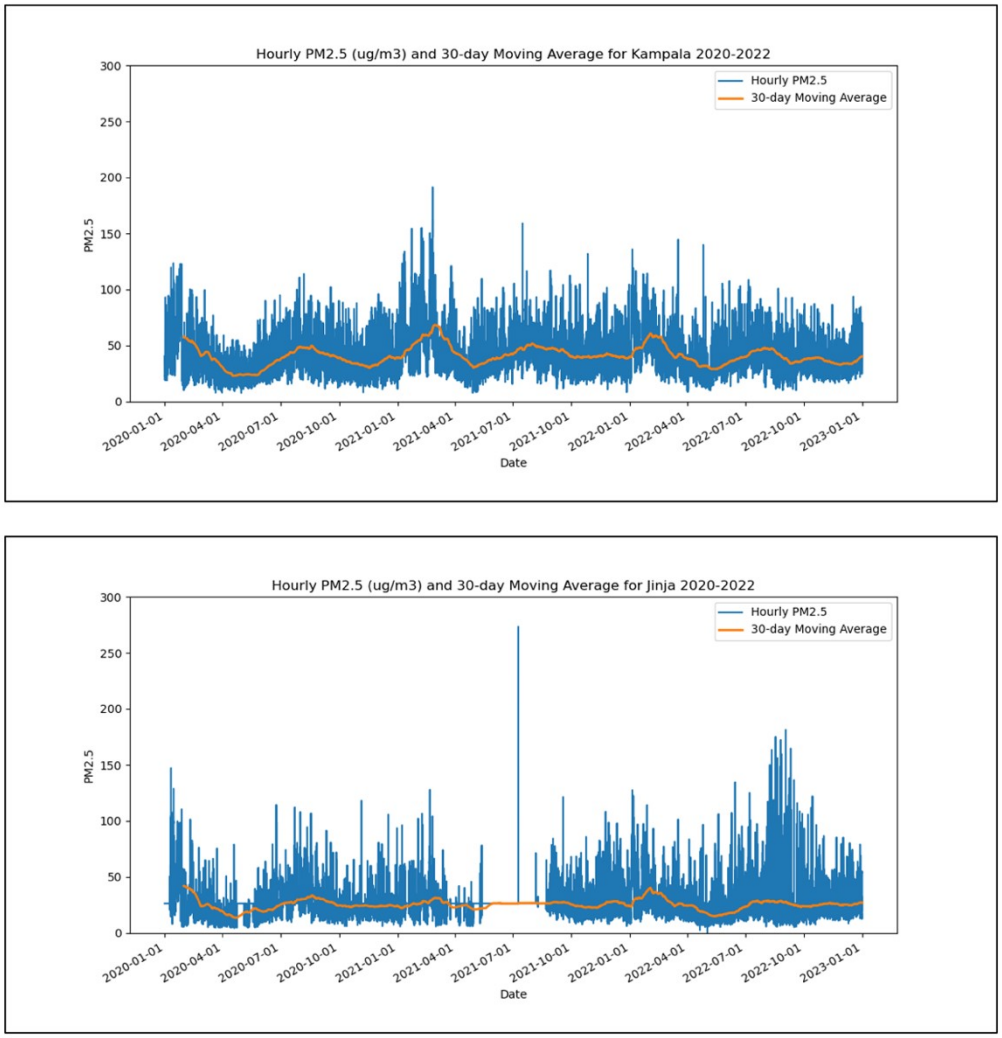
Jinja

Generalized Linear Model Regression Results

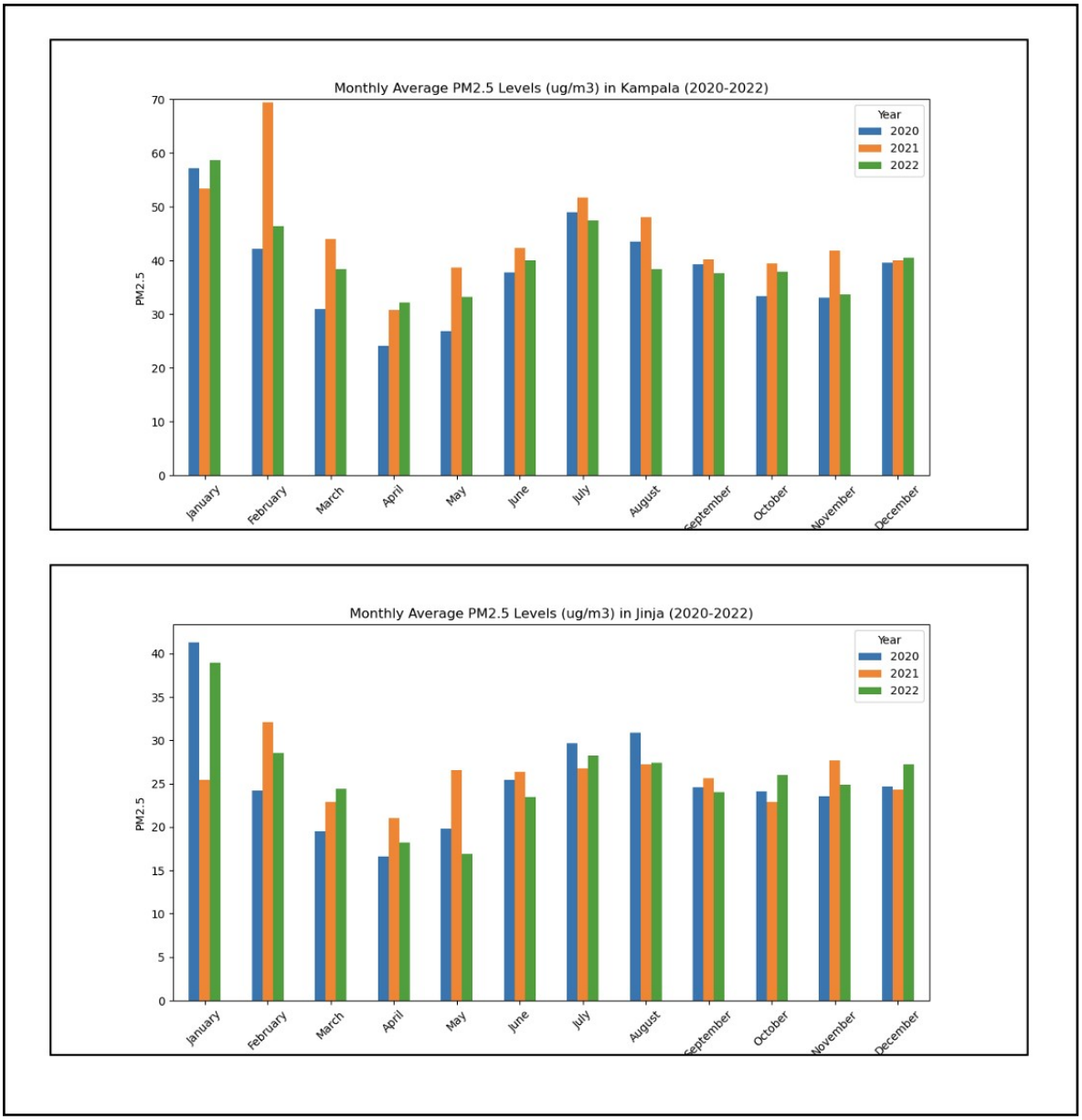
Dep. Variable:	pm2_5_calibrated_value	No. Observations:	26304
Model:	GLM	Df Residuals:	26299
Model Family:	Gamma	Df Model:	4
Link Function:	log	Scale:	0.25885
Method:	IRLS	Log-Likelihood:	-99924.
Date:	Sat, 15 Jul 2023	Deviance:	5537.1
Time:	01:01:02	Pearson chi2:	6.80e+03
No. Iterations:	100	Pseudo R-squ. (CS):	0.06918
Covariance Type:	nonrobust		

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-1.282e+10	7.47e+10	-0.172	0.864	-1.59e+11	1.34e+11
precipitation	-0.0316	0.003	-10.570	0.000	-0.037	-0.026
humidity	-1.3988	0.052	-26.744	0.000	-1.501	-1.296
temp_mean	-0.0821	0.002	-39.550	0.000	-0.086	-0.078
wind_speed	0.0346	0.003	12.314	0.000	0.029	0.040
intercept	1.282e+10	7.47e+10	0.172	0.864	-1.34e+11	1.59e+11

Results



SF 4: Hourly $PM_{2.5}$ (μg/m³) and 30-day Moving Average for Kampala and Jinja from 2020-2022



SF 5:Temporal Variation of $PM_{2.5}$ (µg/m³) in Kampala and Jinja from 2020 to 2022

SF6: Correlation Matrix for Meteorological Parameters in Kampala and Jinja