Arabian Journal of Chemical and Environmental Research

Vol. 08 Issue 02 (2021) 159–180

Statistical analysis of the relationship between lightning activity and average surface wind speed

M. W. Onah[1]*, J. A. Adéchinan[2], F. K. Guédjé[1], H. Kougbéagbédè[1] and E. B. Houngninou[1]

[1]*Laboratoire de Physique de l'Atmosphère, Département de physique, Faculté des Sciences et Techniques, Université d'Abomey-Calavi, Bénin*

[2]*Département de Physique, Faculté des Sciences et Techniques, Université Nationale des Sciences, Technologies, Ingénierie et Mathématiques, Bénin*

Received 12 Mars 2021, Revised 26 April 2021, Accepted 30 April 2021

Abstract

This paper analyses the yearly variability of data from three synoptic stations. These data are cross-referenced with the lightning data of the area. The resulting linear or polynomial regression models revealed the same description of the relationship between the mean number of lightning flashes and the mean surface wind speed. A correlation of 0.75, 0.89 and 0.90 is significantly established between the data from Kandi, Natitingou and Parakou stations respectively. A coefficient of determination of 0.56; 0.80 and 0.81 is significantly obtained respectively for these stations by linear regression and then 0.56; 0.84 and 0.85 by polynomial regression. The F-test showed that the fits of the two models are equal. However, the coefficient of determination is higher with the polynomial regression. All other things being equal, when the average surface wind speed increases by 1m/s, the average number of lightning bolts increases by 8400 according to Kandi, 12674 according to Natitingou and 8847 according to Parakou. More than 80% of the variability in the average number of lightning flashes is explained by the average surface wind speed.

Keywords: WWLLN *lightning, wind speed, wind rose, modelling, prediction, statistical testing.*

*Corresponding author.
E-mail address: waidionah@yahoo.fr

ISSN : 2458-6544

1. Introduction

Thunder and lightning are complex weather phenomena. For a thunderstorm to originate and develop, three essential factors must be present: humidity, instability and a dynamic trigger. Thunderstorms can occur as a result of daytime heating, the passage of cold, moist air over a warmer surface, or by orographic lift or by a convergence of surface winds. Depending on their origin, they are classified as frontal (or cyclonic), orographic and thermal (or heat-related) thunderstorms. Topography is therefore a very important factor. Thunderstorms can occur if an unstable flow of moist air is lifted by a mountain range. In this case, these thunderstorms line up along the windward side of the mountain range, and last as long as the airflow feeds them.

Being a meteorological phenomenon, it is established that thunderstorms and lightning have intimate relationships with other meteorological phenomena such as climate change. Several studies in various localities and at different scales have analysed the link between thunderstorms or the lightning they produce. For example, lightning and the production of NOx, which is a greenhouse gas, [1–14], lightning and rainfall accompanying thunderstorms [15–22] or deep conversion and lightning activity [23–26], lightning and temperature [21, 27, 28], lightning and atmospheric pressure [29], lightning and relative humidity [30], lightning and insolation [31, 32], lightning and wind speed [33], lightning and climate change [21, 28, 34–51]. All of these models provide climatology and even forecasting. Despite their complexity, the forecasting of thunderstorms is still possible. Indices are defined and calculated and then used to predict the risk of thunderstorms locally. These indices are defined according to the type of climate such as Adedokun 2, TTI Mod Index and Faust for temperate climates or the indices of [52] and [53] which are indices of instability based on linear regression with several predictors [54].

The availability of certain data encourages the linking of data that is not always easily accessible. In this way, it will be possible to know the variability of some variables that have the variability of others. The genesis of a storm cell is a function of three essential factors as mentioned above. These factors are also linked to other parameters which are more easily measurable. Statistical modelling, which is a formalised representation of a phenomenon, is a means used to explain the links between the different variables. The construction of models is therefore a delicate process that requires a certain amount of skill in order to highlight reality. It is therefore a schematic or simplified representation of a complex reality. The choice of model is also decisive in order to reveal this reality. Thus an unwise choice can lead to an erroneous representation of reality. It is therefore sometimes necessary to use several models for the same phenomenon in order to identify the optimal model.

In this paper, the north of Benin is taken as an example to analyse the relationship between the average surface wind speed and the average number of lightning flashes. Since it is very often easier to have data

related to wind than to lightning, the latter are chosen as endogenous variables here. Two types of models are examined to establish the links: linear regression and polynomial regression. Linear regression sometimes reflects a coarse relationship between the variables, which is why it is refined by polynomial regression. A model selection is then made from the statistical tests.

In this paper, the north of Benin is taken as an example to analyse the relationship between the average surface wind speed and the average number of lightning flashes. Since it is very often easier to have data related to wind than to lightning, the latter are chosen as endogenous variables here. Two types of models are examined to establish the links: linear regression and polynomial regression. Linear regression sometimes reflects a coarse relationship between the variables, which is why it is refined by polynomial regression. A model selection is then made from the statistical tests.

This work is organised as follows: the next section presents the study area, the data used and the adapted methodology, then the third section presents the results.

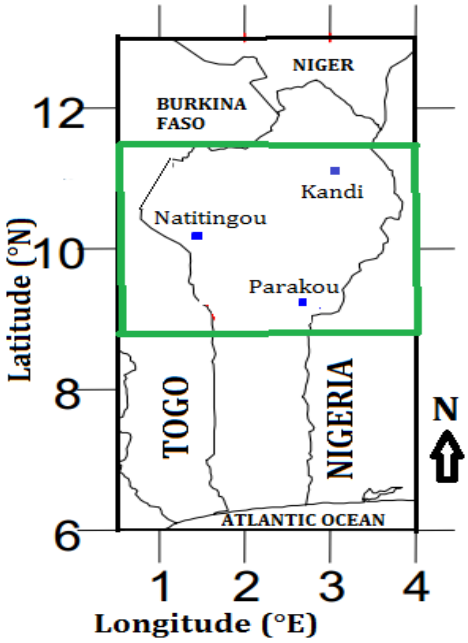

Figure 1. Map showing the study area.

2. Materials and methods

2.1. Presentation of the study area

The study area consists of two topographic units: the crystalline peneplain and the sandstone plateau. The presence of mountain ranges should be noted. Their altitudes in relation to sea level vary between 510 and 700 m. The peneplain is dotted to the south with a multitude of isolated hills. It is connected to the Atacora massif to the west and to the Kandi plateau to the north and north-east. These hills, although not very high, are the major feature of the topography. Despite their modest altitude, they influence the flows. These reliefs increase daytime heating, disturb the currents, aggravate turbulence and favour the

ascent of air masses. Their presence explains the increased importance of thunderstorm events in this region [55, 56].

The region is subject to a Sudanese type climate characterised by a single dry season and a single wet season. The rainy season in this area is from March to October [57]. Thunderstorms mainly occur from late spring to late summer, but they are particularly numerous and violent near the reliefs, even modest ones [56, 58].

Three synoptic stations cover the study area. These are the stations of Kandi, Natitingou and Parakou. These stations belong to the observation network of the National Meteorological Directorate and have produced more than 50 years of climatological data. **Fig1** shows the distribution of the synoptic stations over the study area and **Table 1** specifies their geographical coordinates. There are several reasons for choosing this area. The availability of data and the orography of the area are the most relevant. Indeed, this orography favours the uplift of winds at the surface. The boundaries of the study area are indicated in **Fig1** by the green line.

Table 1. Geographical coordinates of the synoptic stations.

Station	Latitude	Longitude	Altitude
Kandi	11°8N	2°56 E	290m
Natitingou	10°19N	1°23 E	460m
Parakou	9°21N	2°36E	392m

2.2. Data

2.2.1. Flash data

The flash data used in this study comes exclusively from the World Wide Lightning Location Network (WWLLN). This is a real-time global flash detection network with worldwide coverage. WWLLN has more than 70 sensors around the world today [59–61]. Each station in the network consists of a 1.5 m antenna, a GPS (global positioning system) receiver, a receiver for very low frequency electromagnetic radiation (VLF) called lightning sferics and a computer with internet connection. To locate lightning, the technique known as TOGA (Time of Group Arrival) [62–69]. Residual minimization methods are used in the processing to create high quality data for each station. WWLLN data processing ensures that the residual time is less than 30ms and that the data provided by the network corresponds to lightning strikes detected by at least five stations, [64, 68]. The accuracy of lightning location on the network is 5 km, [70, 71]. Thirteen parameters are measured: date, time in UTC (Universal Time Coordinate), latitude and longitude in fractions of a degree, residual error in microseconds (always < 30), the number of

stations involved in locating the lightning strike (always $\geqslant 5$), the energy radiated at very low frequency by the lightning in joules, the uncertainty on the energy radiated in joules, the sub-group of stations located between 1000 and 8000 km from the strike, used to estimate the energy. Each line in the database represents one recorded flash. The data of this study cover the period from January 2005 to December 2017.

2.2.2. Data on climatological variables

Data on climatic variables were collected from the meteorological stations of Kandi, Natitingou and Parakou indicated in **Figure 1**. In these synoptic stations in Benin, devices were installed in a meteorological park for hourly and daily measurements of climatological and meteorological parameters. These devices include the anemometer and weather vane, which are placed about 10metres above the ground to capture wind speed and direction; the weather shelter at two metres to measure data on temperature and humidity of the ambient air; the Campbell heliograph to capture and measure the duration of insolation during the course of a day; and a rain gauge to collect and measure the height of rainfall. The data used covers the period from 1980 to 2019 and includes wind direction (wind), wind speed (speed), air temperature (temp), dew point temperature (dewt), atmospheric pressure (pres) and relative humidity (hr). For this study, only the first two variables are used.

2.3. Methods

2.3.1. Data processing

WWLLN data is available for the entire study area. They are directly used to determine variables at different time scales (hourly, daily, monthly, annual and interannual). Indeed, each return arc in the database is tracked by at least five sensors.

As for the data from the synoptic stations, the percentages of missing values are determined. **Figure 2** shows the percentage of missing values for each series. The data on direction show a high rate of missing data with more than 50% at the Natitingou station. **Figure 3** illustrates the number of gaps shared by several variables. 15 lines of data for Kandi station are missing values, 17 and 14 for Natitingou and Parakou respectively. To fill the gaps, the average of the period of each series was used.

A visualization between the variable to be explained, which is the average number of flashes, and the potential explanatory variables is made. **Figure 4** shows this relationship, specifies the correlation coefficients and indicates the distribution of the different variables. The analysis of this **Figure 4** confirmed the choice of the mean wind speed at the surface as the explanatory variable. In order to analyse wind direction and speed, this second variable is taken into account in the analysis.

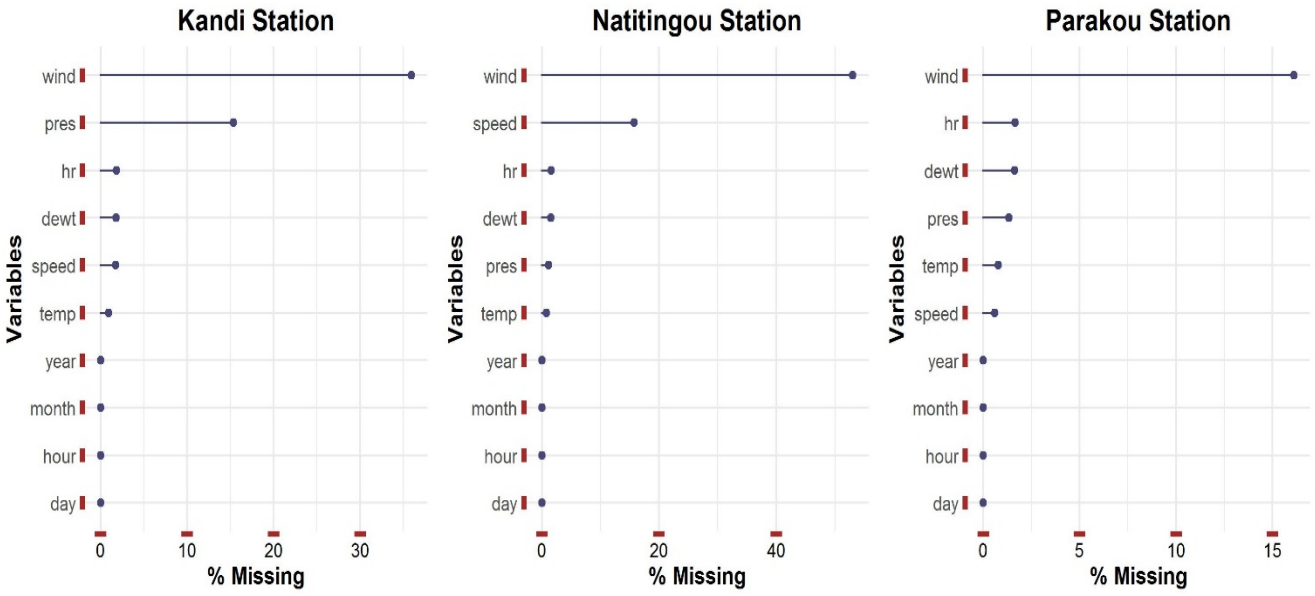

Figure 2. Visualization of the percentage of missing values.

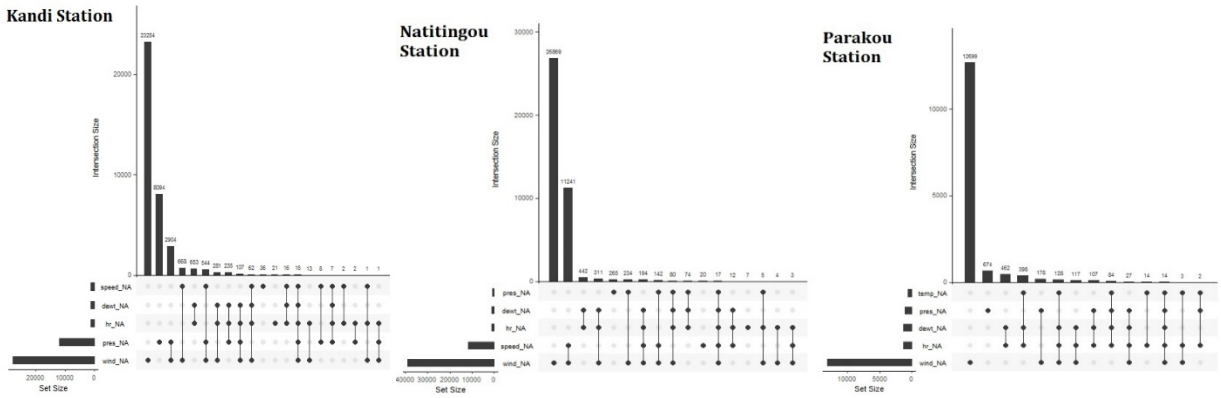

Figure 3. Visualization of the relationships between missing values.

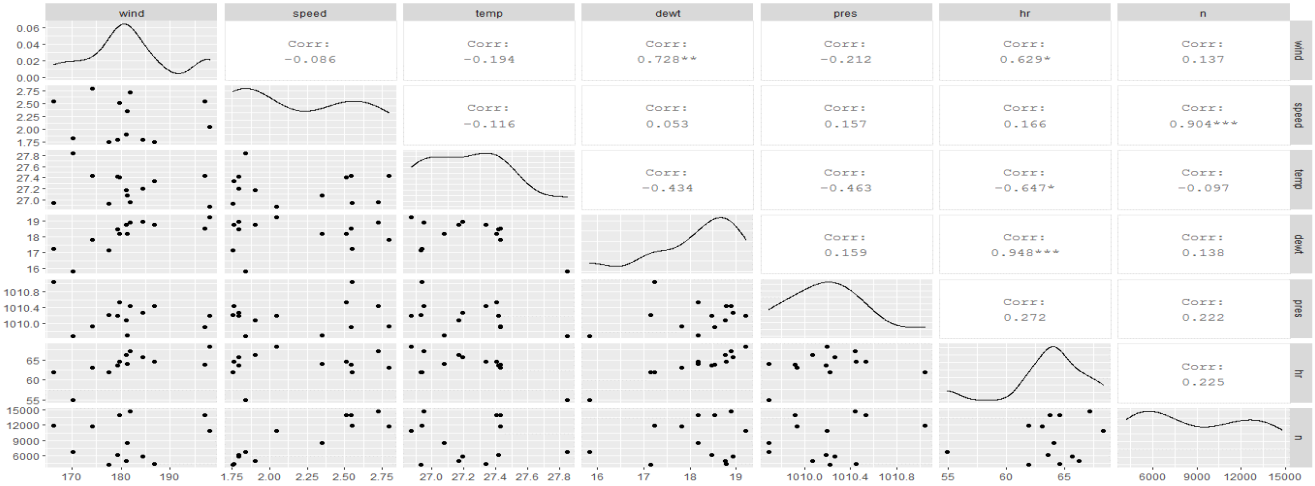

Figure 4. Scatter plot of all pairs of variables.

The Gaussian noise hypothesis makes it possible to obtain the law of estimators and thus to carry out

hypothesis tests on the parameters of the model. This hypothesis is important for this study since the annual averages of the study period of the WWLLN data cover only thirteen years. The unknown model parameters are β_0, β_1 and σ^2. A graphical inspection of the relationship between the average number of flashes and the average wind speed, at different time scales, revealed a noticeable trend with annual averages. Thus, an estimation of the model parameters of the annual averages is made. By choosing y_i as the mean number of flashes and xi as the mean surface wind speed the following model is studied:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i , \qquad i = 1,...n \tag{2}$$

where ε_i are independent random variables of zero expectation and constant variance σ^2, regardless of i. The estimators are obtained by minimising the least squares criterion:

$$S(\alpha_0, \alpha_1) \sum_{i=1}^{n} (y_i - \alpha_0 - \alpha_1 x_i)^2 \tag{3}$$

Several tests are carried out. After an overall evaluation of the model, the interpretation of the coefficients is made. Confidence and prediction intervals for new values are given. If x_0 designates a new observation of the mean surface wind speed, the mean flash value is a realization of the random variable:

$$y_0 = \beta_0 + \beta_1 x_0 + \varepsilon_0 \tag{4}$$

The predictor of y_0 for the new value x_0 is given by:

$$\hat{Y}_0^P = \hat{\beta}_0 + \hat{\beta}_1 \hat{X}_1 \tag{5}$$

The level prediction interval 1-α for y_0 which allows to find two random bounds which will frame the random variable y0 with a probability equal to 1- α :

$$IP_{1-\alpha}(Y_0|x_0) = \left[\hat{Y}_0^P \pm t_{1-\alpha/2}^{(n-2)} \hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}}\right] \tag{6}$$

An unknown fixed value estimator:

$$E(Y_0 | X = x_0) = \beta_0 + \beta_1 x_0 \tag{7}$$

is given by :

$$\hat{E}(Y_0|X = x_0) := \hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 \tag{8}$$

with a confidence interval of level 1-α of E(Y0|X=x0) :

$$IC_{1-\alpha}(\beta_0 + \beta_1 x_0) = \left[\hat{Y}_0^P \pm t_{1-\alpha/2}^{(n-2)} \hat{\sigma} \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}}\right] \tag{9}$$

The analysis of the residuals made it possible to examine the basic assumption of the linear model. The test for assessing the significance of the linear link between the two variables is valid, if the residuals are independent; distributed according to a Normal distribution with a null mean and are homogeneously distributed, i.e. with a constant variance.

Graphic analysis, Jarque and Bera's and Shapiro's non-normality test are used to establish the normality of error terms. The hypothesis of normality of the error terms plays an essential role because it will specify the statistical distribution of the estimators. It is therefore thanks to this hypothesis that statistical inference can be made. The hypothesis of normality can be tested on the variables of the model or on the error terms of the model. The hypothesis test is as follows: Ho: X follows a normal law N(m, σ) against H1: X does not follow a normal law N(m, σ). The Jarque-Bera statistic is defined by:

$$JB = n\left(\frac{S^2}{6} + \frac{(k-3)^2}{24}\right) \quad (10)$$

Where:$S = \frac{\mu_3}{\sigma^3}; = \frac{\mu_4}{\sigma^4}$; $\mu_P = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})P$, respectively, S is the asymmetry coefficient (Skewess) and k the application coefficient (kurtosis). The JB statistic follows a Chi-Two law with two degrees of freedom under the assumption of normality.

The Shapiro-Wilk test is based on the W-statistic. Compared to other tests, it is particularly powerful for small populations (n ⩽ 50). The test statistic is written as follows:

$$W = \frac{\left[\sum_{i=1}^{\frac{n}{2}} a_i \left(x_{(-i+1)} - x_{(i)}\right)\right]^2}{\sum_{i=1}^{\frac{n}{2}}(x_i - \bar{x})^2} \quad (11)$$

Where: x(i) corresponds to the sorted data series ; $\frac{n}{2}$ is the entire part of the $\frac{n}{2}$ report; a_i are constants generated from the mean and variance covariance matrix of the quantiles of a sample of size n according to the normal distribution. The W statistic can therefore be interpreted as the coefficient of determination (the square of the correlation coefficient) between the series of quantiles generated from the normal distribution and the empirical quantiles obtained from the data. The higher the W statistic, the more credible the compatibility with the normal distribution.

The Breusch-Pagan test is used to check the homogeneity of the residues. It tests the hypothesis of homoscedasticity of the error term of a linear regression model. The Breusch-Pagan statistic:

$$BP = nR^2 \quad (12)$$

Which follows χ^2(K-1) with K the number of coefficients to be estimated, n the number of values used and R^2 the coefficient of determination. If the Breusch-Pagan statistic is higher than the one read in the Chi-Deux table for a certain level of risk of error of the first species (5% being the value generally retained), then the null hypothesis of homoscedasticity is rejected.

The Durbin-Watson test is used to analyse the independence of the residues.

$$DW = \frac{\sum_{i=2}^{n}(e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2} \quad (13)$$

with e residue. This statistic, noted d or DW, has a value between 0 and 4. If it is close to zero, the

autocorrelation is positive, values around 2 show an absence of autocorrelation and if it is close to 4, there is negative autocorrelation (values both above and below the trend).

To ensure that the linear regression is not coarse, a polynomial regression was performed. The polynomial model consists of representing the relationship between the explanatory variable Y and an explanatory variable X in a non-linear form of the type:

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \cdots + \beta_P X^P + \varepsilon \quad (14)$$

This model is a multiple regression model with p degrees being the successive powers of the explanatory variable. It should be noted that polynomial regression belongs to the family of linear models because linear refers to the parameters of the model and the fact that their effects are added together. Moreover, linear regression is a polynomial regression of degree 1. For this study p = 2, i.e. a polynomial regression of degree 2.

To compare the fit of the two models, an F-test is used. This test is defined by:

$$F = \frac{\frac{RSS_1 - RSS_2}{nbparam_2 - nbparam_1}}{\frac{RSS_2}{n - nbparam_2}} \quad (15)$$

where: Index 1 refers to the linear model and index 2 to the polynomial model, RSS the sum of the residual squares, nb_{param} number of parameter of the models which is equal to 2 for linear regression (intercept and slope) and 3 for polynomial regression (intercept and both slopes), n the number of data.

3. Results and discussion

3.1. Comparison of station data

The daily, monthly and annual averages of the data for each station show good correlation. **Figure 5** and **6** summarize the observations made by the three synoptic stations for the variables wind direction and mean wind speed. **Figure 5** shows the comparative evolution of the average speed of the three synoptic stations. The inter-annual variability of the three stations is consistent. An almost identical trend can therefore be seen for all three stations. This observation ensures that the correction of the data did not significantly affect them. **Figure 6** shows the comparative variation in average wind direction for the three synoptic stations. Despite the extensive correction of the data, a highly correlated trend in the station data can be noted. It should be noted that the different stations do not have the same rate of deficiency. Despite the remarkable correlation of the data, the analysis took into account the data from each station. The identification of the link between the mean wind speed and the mean number of flashes is done on different time scales (daily, monthly and yearly). Only the annual scale produced convincing results. Therefore, only the results of these data will be presented in the following.

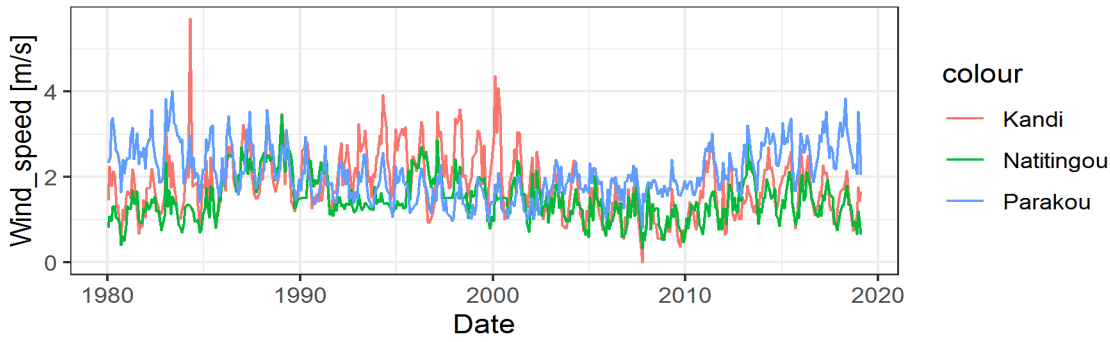

Figure 5. Climatology of the average monthly wind speed of the three stations.

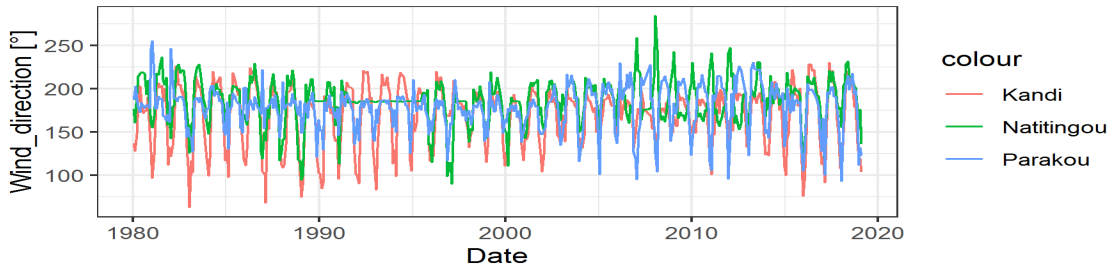

Figure 6. Climatology of the average monthly wind direction of the three stations.

3.2. Analysis of raw data

Figure 7 shows the relationship between the annual average wind speed and the average number of flashes. The same trend is noted for both variables, which bodes well for a possible correlation between these variables. It should be noted that the average speed is higher through the data from the Parakou station than for the other two stations. The inter-annual variation of the two series is the same.

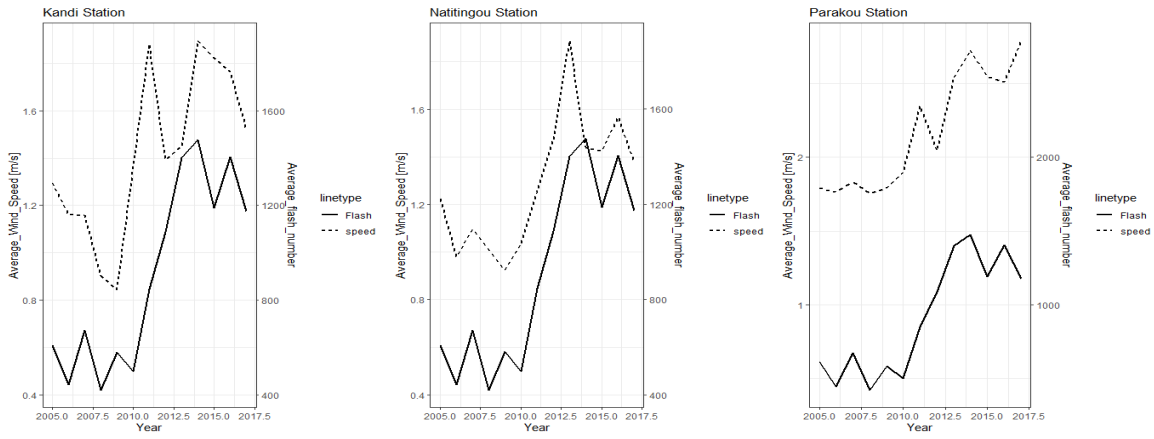

Figure 7. Comparison of average wind speed and average number of flashes at each station.

Figure 8 shows the wind rose obtained from the data from each station. The directions are generally the same but with a difference in the proportions of the speeds. The prevailing winds are SN (South North),

SW-NE (South West - North East) and SE-NW (South East - North West), with a more representative average speed of 2.09m/s at the Parakou station.

Figure 9 details the monthly wind rose for each station in order to contrast it with the average monthly numbers of flashes. **Figure 10** displays the average monthly numbers of flashes for the study period. The evolution of the winds is identical for the three stations. From March to October the dominant direction is SN and contrasts with the monthly average number of flashes. Indeed, the average number of flashes increases from March to October as well. Likewise, the percentage of winds in September is at its maximum as is the average number of flashes.

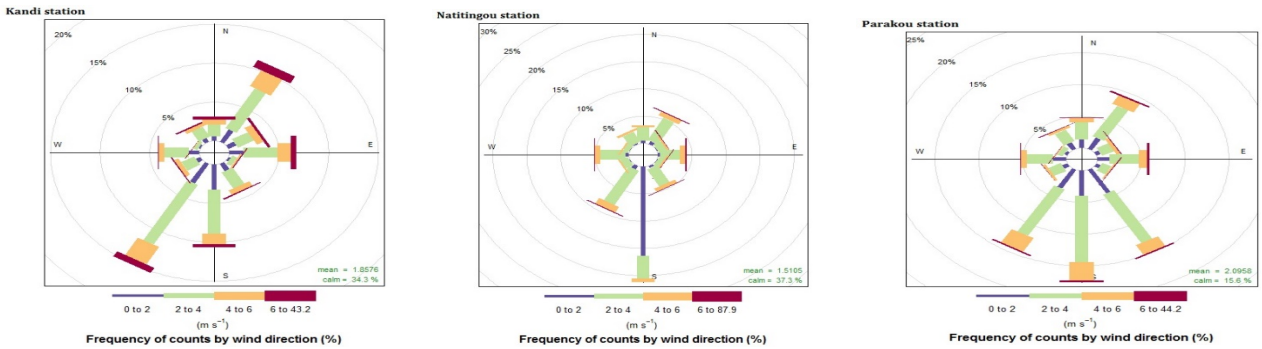

Figure 8. Compass rose obtained from the data of each synoptic station.

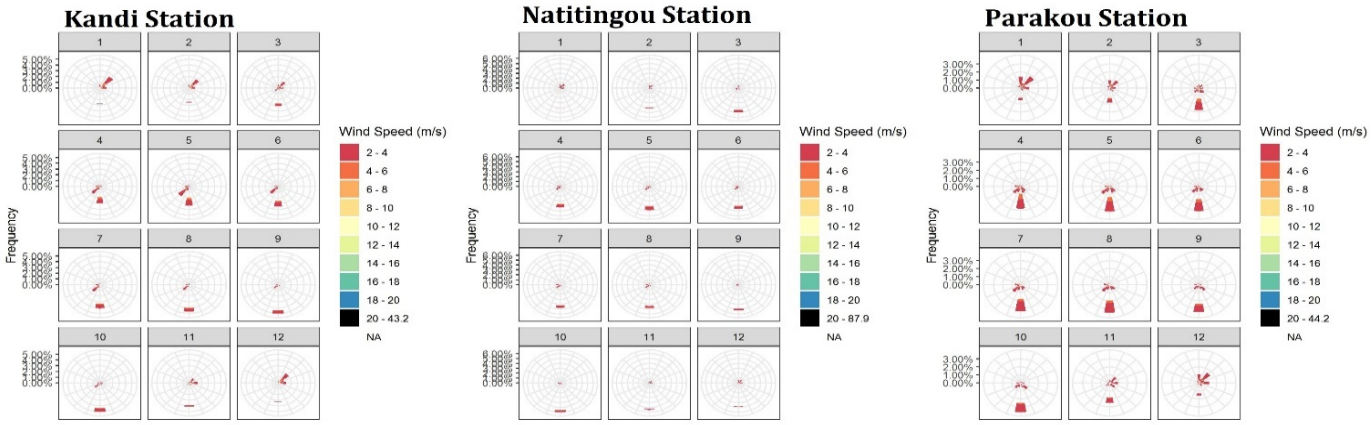

Figure 9. Compass rose of the monthly data for each station.

3.3. Modelling

3.3.1. Linear regression

The correlation coefficient of the two sets of variables is determined per station and the nullity test of the coefficient to show that it is significant as summarized in **Table 2**. This coefficient is approximately 0.9 for the Natitingou and Parakou stations and 0.75 for Kandi.

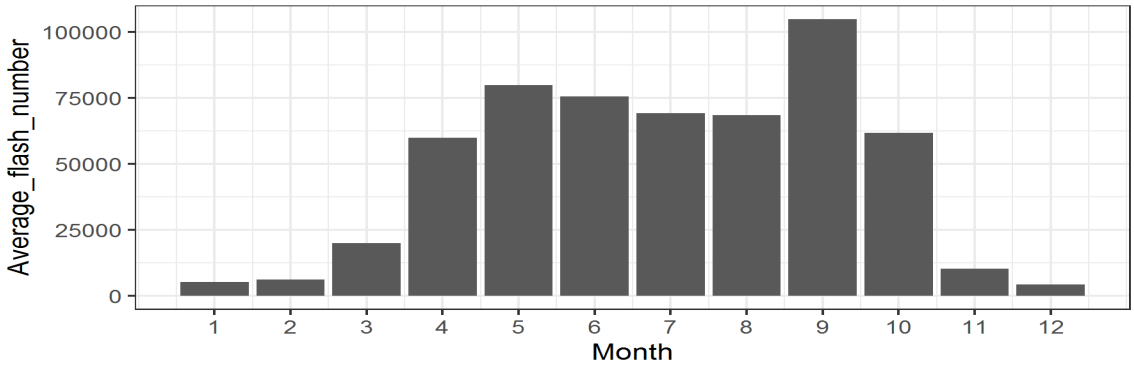

Figure 10. Monthly average of the number of flashes during the study period.

Table 2. Results of the nullity test of the correlation coefficient by station.

Station	Result of the nullity test
Kandi	cor = 0.7491992 p-value = 0.003201 interval: 0.3375564 - 0.9202906
Natitingou	horn = 0.8955086 p-value = 3.498e-05 interval: 0.6800860 - 0.9685832
Parakou	horn = 0.9035666 p-value = 2.286e-05 interval: 0.7021461 - 0.9710919

Figure 11 shows for each station that a significant autocorrelation is present for lag 1, i.e. between the residuals of one line of the data table and those of the next line. The results of the Durbin-Watson test, on the other hand, summarised in **Table 3**, show that there is no significant autocorrelation between the residuals of one row of the data table and those of the next row because the p-values are higher than 0.05. The independence of the residuals can be accepted.

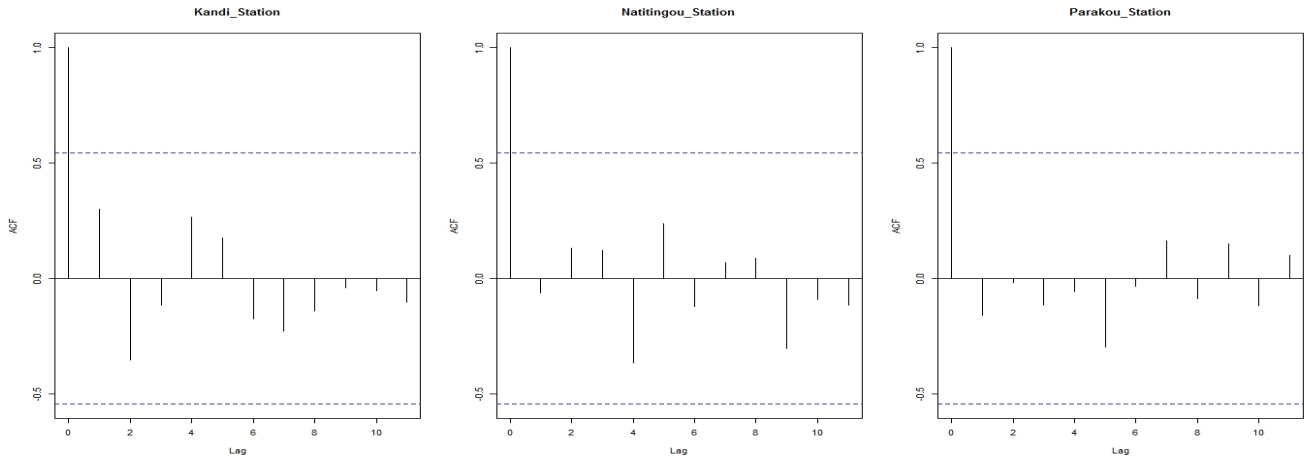

Figure 11. Autocorrelation analysis plot of residuals per station.

Figure 12 gives for each synoptic station, four graphs for the analysis of linearity (Residual vs Fitted); residue normality (Normal Q-Q) and residue homogeneity (Scale-Location). The analysis of these graphs shows that the normality is globally satisfactory in Parakou, unsatisfactory in Natitingou and very little in Kandi. In the same order, the homogeneity of the residues and the linearity of the model should

be noted. These observations are reinforced by statistical tests. The Jarque and Bera test does not allow to conclude to the non-normality of the errors as shown in **Table 4**. Similarly, the Shapiro test does not reject the normality of the residuals for both models, again according to **Table 4**.

Table 3. Results of the Durbin-Watson test by station.

Station	Result of the Durbin-Watson test			
	lag	Autocorrelation	D-W Statistic	p-value
Kandi	1	0.3007709	1.312228	0.11
Natitingou	1	-0.1579085	2.089957	0.88
Parakou	1	-0.06098996	1.927588	0.708

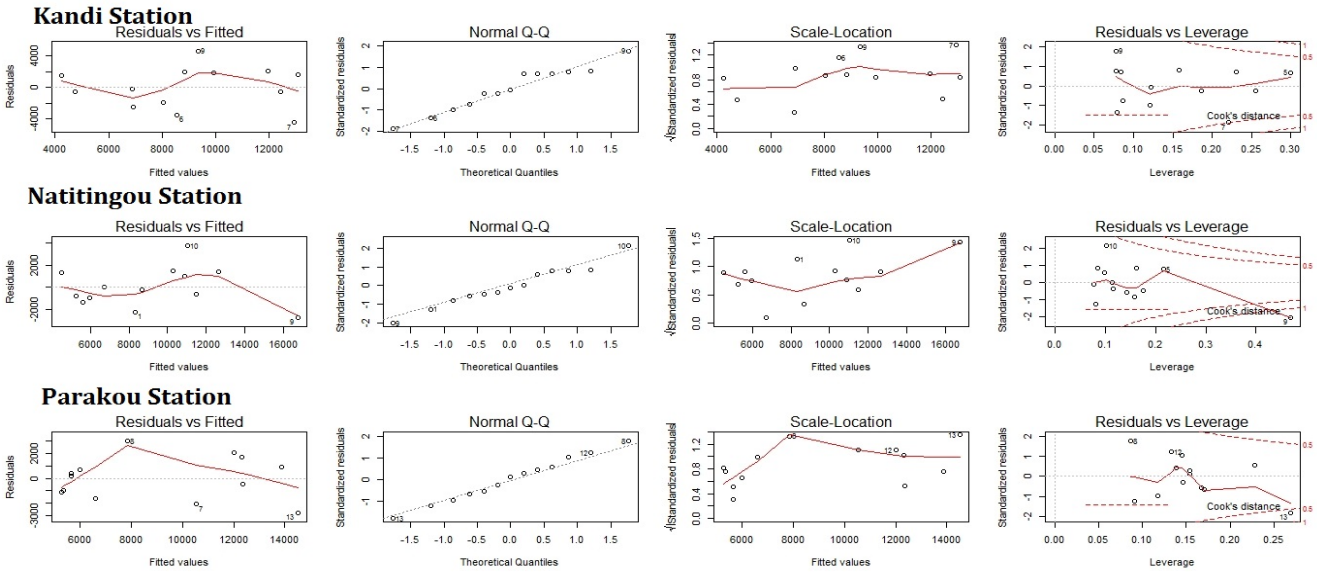

Figure 12. Analytical plot of normality and homogeneity of residues per station.

Table 4. Residue normality test result.

Results of the Jarque and Bera test		
Station	Linear regression	Polynomial regression
Kandi	X-squared = 0.35616, p-value = 0.8369	X-squared = 0.5122,p-value = 0.7741
Natitingou	X-squared = 0.382, p-value = 0.8261	X-squared = 0.35864,p-value = 0.8358
Parakou	X-squared = 0.42966, p-value = 0.8067	X-squared = 0.83217,p-value = 0.6596
Result of the test of Shapiro - Wilk		
Station	Linear regression	Polynomial regression
Kandi	W = 0.95689, p-value = 0.7053	W = 0.94078, p-value = 0.4671
Natitingou	W = 0.96555, p-value = 0.8361	W = 0.95344, p-value = 0.6512
Parakou	W = 0.98584, p-value = 0.9967	W = 0.95313, p-value = 0.6464

The Breusch-Pagan test, as shown in **Table 5**, accepts the assumption of homogeneity of the residuals.

Figure 13 shows the least-squares regression line with the confidence interval on the scatterplot of the mean number of flashes as a function of the mean surface wind speed for the three stations. It shows the regression parameters, the coefficient of determination and the p-value.

The overall evaluation of the model shows that it is highly significant as the p-value is well below 1% for all stations. The interpretation of the coefficients begins by determining their significance. The coefficients are also highly significant except in Kandi for β_0. The p-value of the β_0 student t-test is 0.4034; 0.0148 and 0.00386 respectively at Kandi, Natitingou and Parakou stations. Similarly, the p-value of the student t-test of β_1 is 0.0032; 3.5e-05 and 2.29e-05 respectively at Kandi, Natitingou and Parakou stations. This coefficient is very significant at each station. Interpretation is therefore possible. This coefficient is positive at each station. When the average speed increases by 1m/s the average number of flashes increases by at least 8400. The quality of the model which is assessed from the determination coefficient R^2 is appreciable for two stations. The adjusted determination coefficient is 0.5214; 0.7839 and 0.7997 respectively in Kandi, Natitingou and Parakou. As the adjusted R^2 is not higher than 85%, no problem, especially endogeneity, can be raised. Thus the fit between the model and the observed data is very strong. At Kandi station, 56.13% of the variability in the mean number of flashes is explained by the mean surface wind speed; 80.19% at Natitingou and 81.64% at Parakou.

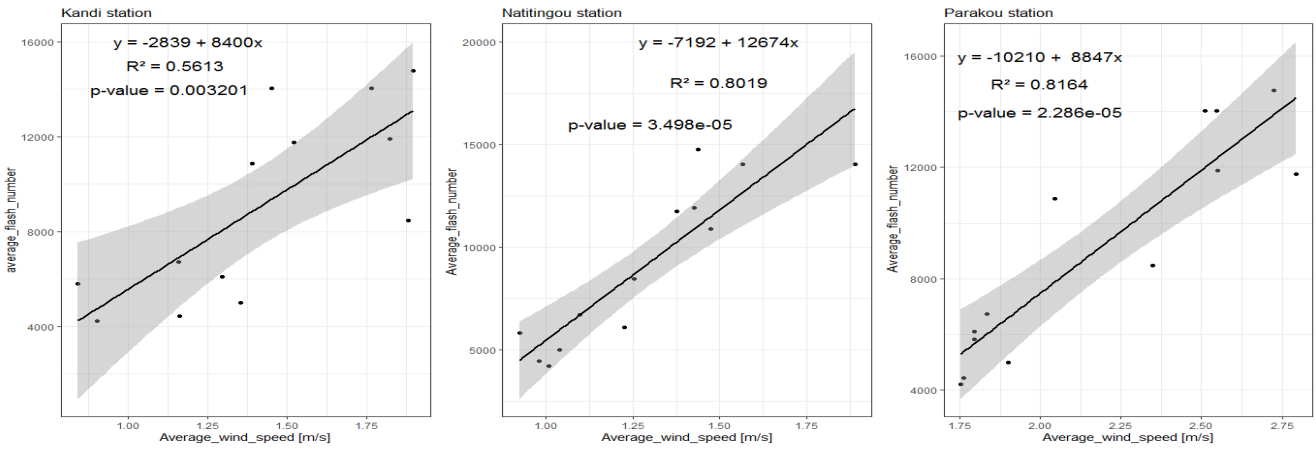

Figure 13. Representation of the least-squares regression line with the confidence interval on the scatterplot of the mean number of flashes as a function of the mean surface wind speed of the three synoptic stations in the study area.

3.3.2. Polynomial regression

The search for a better link suggests exploring a second model. This is how polynomial regression is examined. As in the case of linear regression, the analysis of the residuals is done in order to proceed with the interpretation of the coefficients. **Figure 14** shows the plots for the analysis of linearity, normality and homogeneity of the residuals. **Tables 4** and **5** provide the results of statistical tests on the assumptions of normality and homogeneity of the residuals. These assumptions cannot be rejected. **Figure 15** shows the fit curve with the confidence interval on the scatterplot of the mean number of

flashes as a function of the mean surface wind speed for the three stations. It shows the regression parameters, the coefficient of determination and the p-value. The overall evaluation of the model shows that it is highly significant, especially the p-value is much lower than 1% at two stations. The coefficients are not significant. The interpretation is therefore delicate. The adjusted determination coefficient is 0.4761; 0.8148 and 0.8134 respectively in Kandi, Natitingou and Parakou. Here again, the adjusted R^2 is not higher than 85%, no problem, notably endogeneity, can be raised. Thus, the fit between the model and the observed data is very strong.

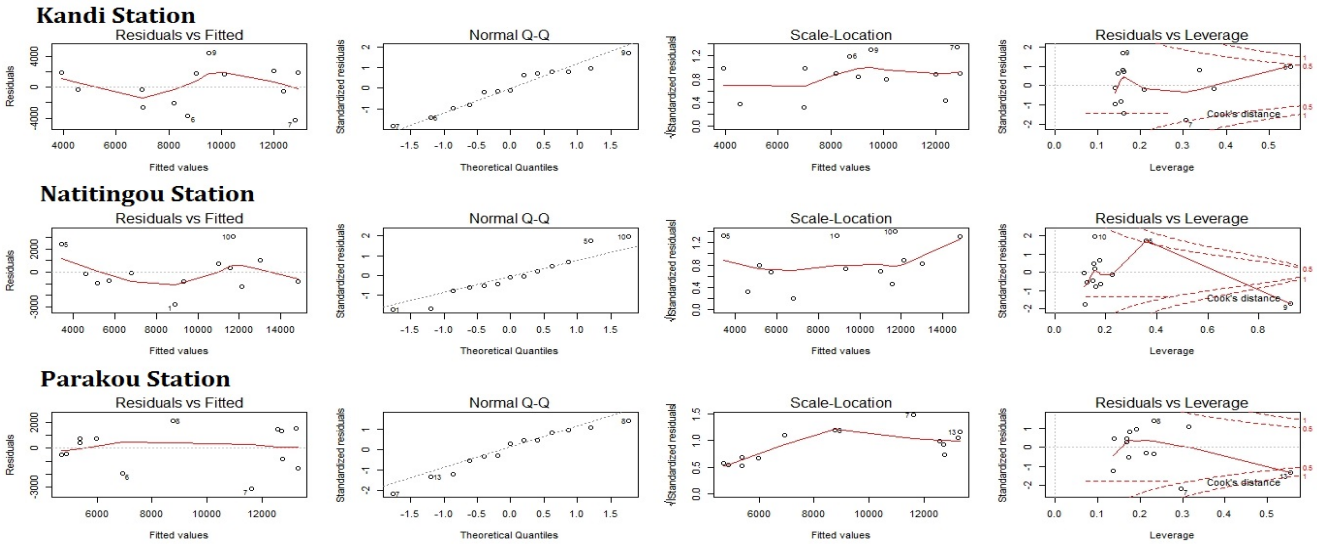

Figure 14. Normality and homogeneity of residues per station.

Table 5. Residue homogeneity test result.

Result of the Breusch-Pagan test		
Station	Linear regression	Polynomial regression
Kandi	Chisquare = 0.6494217, p = 0.42032	Chisquare = 0.5203733, p = 0.47068
Natitingou	Chisquare = 2.178342,p = 0.13997	Chisquare=0.0001117762,p=0.99156
Parakou	Chisquare = 0.9910194, p = 0.31949	Chisquare = 1.261368,p = 0.26139

At Kandi station, 56.34% of the variability in the mean number of flashes is explained by the mean surface wind speed; 84.57% at Natitingou and 84.45% at Parakou. It should be noted that the observations did not vary too much at Kandi station.

3.3.3. Comparison of the two models

In order to compare the adjustments of these two models, F-test is calculated. It gives respectively 0.048; 2.8369; 1.8035 to Kandi, Natitingou and Parakou with a respective p-value of 0.831; 0.123; 0.209. It

should be noted that all the p-values are higher than 0.05, the null hypothesis, which specifies that the adjustments of the two models are equal, is accepted. Thus the fit of the polynomial regression model of degree 2 is not significantly better than that of the linear regression model. However its coefficient of determination is higher.

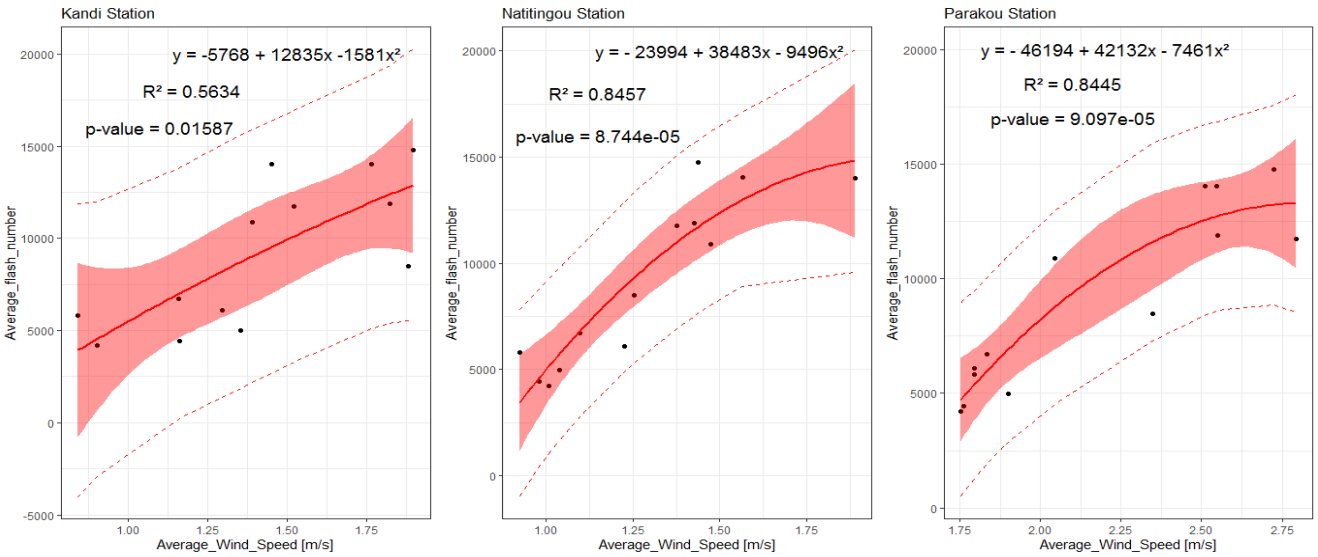

Figure 15. Adjustment curve with confidence and prediction intervals on the scatterplot of the mean number of flashes as a function of the mean surface wind speed of the three synoptic stations in the study area.

Conclusion

A concordance between the data collected by the synoptic stations is noted despite the record of missing values at the Natitingou station and a slight discrepancy at the Kandi station. Graphical analysis of the raw data suggested some relationship between the average number of lightning strikes and the average surface wind speed. This relationship is confirmed by statistical analysis of the data. Linear or polynomial regression models resulted in the same description of the relationship between the mean number of lightning strikes and the mean surface wind speed. However, the coefficient of determination is higher with polynomial regression. A correlation of nearly 90% is significantly established between the data. All other things being equal, when the mean surface wind speed increases by 1m/s, the mean number of lightning flashes increases by at least 8400. More than 80% of the variability in the mean number of flashes is explained by the mean surface wind speed. These results are in agreement with previous studies which have shown that thunderstorm activity is linked to the orography of the area and therefore to the uplift of the surface winds.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgments
The authors wish to thank Professor Pascal Ortega and the World Wide Lightning Location Network (http://wwlln.net), for providing the lightning location data used in this paper. Data. We would like to express our sincere gratitude to two anonymous reviewers who made very constructive comments on the manuscript and helped improve the final output.

References

[1] W.C. Skamarock, J.E. Dye, E. Defer, M.C. Barth, J.L. Stith, B.A. Ridley, *et al.* Observational-and modeling-based budget of lightning-produced NOx in a continental thunderstorm. *Journal of Geophysical Research: Atmospheres.* 108(D10) (2003).

[2] C. Barthe, J-P. Pinty, C. Mari, Lightning-produced NOx in an explicit electrical scheme tested in a Stratosphere-Troposphere Experiment: Radiation, Aerosols, and Ozone case study. *Journal of Geophysical Research: Atmospheres.* 112(D4) (2007).

[3] C. Barthe, M.C. Barth, Evaluation of a new lightning-produced NOx parameterization for cloud resolving models and its associated uncertainties. *Atmospheric Chemistry and Physics*. 8(16): 4691–4710 (2008).

[4] H. Höller, H-D. Betz, K. Schmidt, R.V. Calheiros, P. May, E. Houngninou, *et al.* LINET Lightning Characteristics Observed on 4 Different Continents. *In EGU General Assembly Conference Abstracts* p. 13610 (2009).

[5] Bucsela EJ, Pickering KE, Huntemann TL, Cohen RC, Perring A, Gleason JF, *et al.* Lightning-generated NOx seen by the Ozone Monitoring Instrument during NASA's Tropical Composition, Cloud and Climate Coupling Experiment (TC4). *Journal of Geophysical Research: Atmospheres.* 115(D10) (2010).

[6] Beirle S, Huntrieser H, Wagner T. Direct satellite observation of lightning-produced NO x. Atmospheric Chemistry and Physics. 10(22): 10965–10986 (2010).

[7] Virts KS, Thornton JA, Wallace JM, Hutchins ML, Holzworth RH, Jacobson AR. Daily and intraseasonal relationships between lightning and NO_2 over the Maritime Continent: LIGHTNING, NO_2 OVER MARITIME CONTINENT. Geophys Res Lett. 38(19): n/a-n/a (2011).

[8] Huntrieser H, Schlager H, Lichtenstern M, Stock P, Hamburger T, Höller H, *et al.* Mesoscale convective systems observed during AMMA and their impact on the NOx and O3 budget over West Africa. Atmospheric Chemistry and Physics. 11(6): 2503–2536 (2011).

[9] Wong J, Barth MC, Noone D. Evaluating a lightning parameterization based on cloud-top height

for mesoscale numerical model simulations. Geoscientific Model Development. 6(2): 429–443 (2013).

[10] Kang D, Wong D, Gilliam RC, Pleim JE, Mathur R. Assessment of Lightning Assimilation and Lightning NO in the WRF-CMAQ Modeling System Using WWLLN Lightning Flash Data. In 100th American Meteorological Society Annual Meeting. AMS (2020).

[11] Allen DJ, Pickering KE, Bucsela E, Krotkov N, Holzworth R. Lightning NOx production in the tropics during the boreal summer as determined using OMI NO2 Retrievals and WWLLN stroke data. J Geophys Res, Atmospheres, submitted. (2019).

[12] Allen DJ, Pickering KE, Bucsela E, Krotkov NA, Holzworth RH. Lightning NO x Production in the Midlatitudes and Tropics as Determined Using OMI NO 2 Retrievals and WWLLN Stroke Data (Invited Presentation). In 99th American Meteorological Society Annual Meeting. AMS (2019).

[13] Allen DJ, Pickering KE, Bucsela E, Krotkov N, Holzworth R. Lightning NOx Production in the Tropics as Determined Using OMI NO2 Retrievals and WWLLN Stroke Data. Journal of Geophysical Research: Atmospheres. 124(23): 13498–13518 (2019).

[14] Bucsela EJ, Pickering KE, Allen DJ, Holzworth RH, Krotkov NA. Midlatitude Lightning NOx Production Efficiency Inferred From OMI and WWLLN Data. Journal of Geophysical Research: Atmospheres. 124(23): 13475–13497 (2019).

[15] Lang TJ, Miller LJ, Weisman M, Rutledge SA, Barker III LJ, Bringi VN, *et al.* The severe thunderstorm electrification and precipitation study. Bulletin of the American Meteorological Society. 85(8): 1107–1126 (2004).

[16] Liu C, Zipser EJ, Cecil DJ, Nesbitt SW, Sherwood S. A cloud and precipitation feature database from nine years of TRMM observations. Journal of Applied Meteorology and Climatology. 47(10): 2712–2728 (2008).

[17] Pessi AT, Businger S. Relationships among lightning, precipitation, and hydrometeor characteristics over the North Pacific Ocean. *Journal of Applied Meteorology and Climatology*. 48(4): 833–848 (2009).

[18] Barthe C, Asencio N, Lafore J-P, Chong M, Campistron B, Cazenave F. Multi-scale analysis of the 25–27 July 2006 convective period over Niamey: Comparison between Doppler radar observations and simulations. Quarterly Journal of the Royal Meteorological Society. 136(S1): 190–208 (2010).

[19] Etienne HB, Joseph AA, Sounmaïla M, François GK, Hilaire K, Erik HT. Relation Entre Éclairs Nuage-Sol et Précipitations Pendant la Mousson de 2006 au Bénin. European Journal of Scientific Research. 115(1): pp.122-132 (2013).

[20] Adechinan AJ, Houngninou BE, Kougbeagbede H. Relationships between lightning and rainfall

intensities during rainy events in Benin. International Journal of Innovation and Applied Studies. 9(2): 765–776 (2014).

[21] Yang Y, Song D, Wang S, Li P, Xu Y. Characteristics of cloud-to-ground lightning and its relationship with climate change in Muli, Sichuan province, China. Natural hazards. 91(3): 1097–1112 (2018).

[22] Minobe S, Park JH, Virts KS. Diurnal cycles of precipitation and lightning in the tropics observed by TRMM3G68, GSMaP, LIS, and WWLLN. Journal of Climate. (2020) (2020).

[23] Lang TJ, Rutledge SA, Dye JE, Venticinque M, Laroche P, Defer E. Anomalously low negative cloud-to-ground lightning flash rates in intense convective storms observed during STERAO-A. Monthly weather review. 128(1): 160–173 (2000).

[24] Avila EE, Bürgesser RE, Castellano NE, Collier AB, Compagnucci RH, Hughes AR. Correlations between deep convection and lightning activity on a global scale. Journal of Atmospheric and Solar-Terrestrial Physics. 72(14–15): 1114–1121 (2010).

[25] Liu C, Cecil DJ, Zipser EJ, Kronfeld K, Robertson R. Relationships between lightning flash rates and radar reflectivity vertical structures in thunderstorms over the tropics and subtropics. Journal of Geophysical Research: Atmospheres. 117(D6) (2012).

[26] Reinhart B, Fuelberg H, Blakeslee R, Mach D, Heymsfield A, Bansemer A, *et al.* Understanding the Relationships between Lightning, Cloud Microphysics, and Airborne Radar-Derived Storm Structure during Hurricane Karl (2010). Monthly Weather Review. 142(2): 590–605 (2014).

[27] Price C, Asfur M. Inferred long term trends in lightning activity over Africa. Earth, planets and space. 58(9): 1197–1201 (2006).

[28] Pinto Jr O, Pinto I. On the sensitivity of cloud-to-ground lightning activity to surface air temperature changes at different timescales in São Paulo, Brazil. Journal of Geophysical Research: Atmospheres. 113(D20) (2008).

[29] Lenouo A, Monkam D, Mkankam Kamga F. Analyse des conditions météorologiques pour la sécurité aérienne à Douala. La Météorologie. (2009).

[30] Price C, Asfur M. Can lightning observations be used as an indicator of upper-tropospheric water vapor variability? Bulletin of the American Meteorological Society. 87(3): 291–298 (2006).

[31] Neto OP, Pinto IR, Pinto Jr O. The relationship between thunderstorm and solar activity for Brazil from 1951 to 2009. Journal of Atmospheric and Solar-Terrestrial Physics. 98: 12–21 (2013).

[32] Etienne HB, Joseph AA, François GK, Waïdi OM, Hilaire K. Relationships between lightning and insolation during monsoon season in Benin. Research Journal of Physical Sciences. 5: 5 (2017).

[33] Corbosiero KL, Molinari J. The Relationship between Storm Motion, Vertical Wind Shear, and

Convective Asymmetries in Tropical Cyclones. Journal of Atmospheric Sciences. 60(2): 366–376 (2003).

[34] Price C, Rind D. The impact of a 2$\times$ CO2 climate on lightning-caused fires. Journal of Climate. 7(10): 1484–1494 (1994).

[35] Sinha A, Toumi R. Tropospheric ozone, lightning, and climate change. Journal of Geophysical Research: Atmospheres. 102(D9): 10667–10672 (1997).

[36] Flannigan M, Bergeron Y, Engelmark O, Wotton B. Future wildfire in circumboreal forests in relation to global warming. Journal of Vegetation Science. 9(4): 469–476 (1998).

[37] Wotton M, Logan K, McAlpine R, others. Climate change and the future fire environment in Ontario: fire occurrence and fire management impacts. Climate Change Research Report-Ontario Forest Research Institute. (CCRR-01) (2005).

[38] Kochtubajda B, Flannigan M, Gyakum J, Stewart R, Logan K, Nguyen T-V. Lightning and fires in the Northwest Territories and responses to future climate change. Arctic. 211–221 (2006).

[39] Flannigan MD, Kochtubajda B, Logan KA. Forest fires and climate change in the Northwest Territories. In Cold Region Atmospheric and Hydrologic Studies. The Mackenzie GEWEX Experience. Springer pp. 403–417 (2008).

[40] Lutz JA, Van Wagtendonk JW, Thode AE, Miller JD, Franklin JF. Climate, lightning ignitions, and fire severity in Yosemite National Park, California, USA. International Journal of Wildland Fire. 18(7): 765–774 (2009).

[41] Arif A. Lightning Activity and Climate Change with NCEP and CRCM Data: Canada and Northern United States Regions. Éditions universitaires européennes (2012).

[42] Price CG. Lightning applications in weather and climate research. Surveys in Geophysics. 34(6): 755–767 (2013).

[43] Price CG. Lightning applications in weather and climate research. Surveys in Geophysics. 34(6): 755–767 (2013).

[44] Reeve N, Toumi R. Lightning activity as an indicator of climate change. Quarterly Journal of the Royal Meteorological Society. 125(555): 893–903 (1999).

[45] Fisher J, Hoole P, Pirapaharan K, Thirukumaran S, Hoole S. Cloud to ground and ground to cloud flashes in lightning protection: and future severe lightning and climate change. In 2014 International Conference on Lightning Protection (ICLP). IEEE pp. 440–445 (2014).

[46] Romps DM, Seeley JT, Vollaro D, Molinari J. Projected increase in lightning strikes in the United States due to global warming. Science. 346(6211): 851–854 (2014).

[47] Krause A, Kloster S, Wilkenskjeld S, Paeth H. The sensitivity of global wildfires to simulated past,

present, and future lightning frequency. Journal of Geophysical Research: Biogeosciences. 119(3): 312–322 (2014).

[48] Tippett MK, Allen JT, Gensini VA, Brooks HE. Climate and hazardous convective weather. Current Climate Change Reports. 1(2): 60–73 (2015).

[49] Williams E. Lightning and climate: A review. Atmospheric Research. 76(1–4): 272–287 (2005).

[50] Guo F, Chen C, Ju X. Response of lightning, NO and O 3 to climate anomaly in China. In 2014 International Conference on Lightning Protection (ICLP). IEEE pp. 551–554 (2014).

[51] Abatzoglou JT, Kolden CA, Balch JK, Bradley BA. Controls on interannual variability in lightning-caused fire activity in the western US. Environmental Research Letters. 11(4): 045005 (2016).

[52] Charba JP. Two to six hours severe local storm probabilities: an operational forecasting system. Mon Wea Rev. 107: 269–281 (1979).

[53] Showalter AK. A stability index for thunderstorm forecasting. Bull Amer Meteor Soc. 34: 250–252 (1953).

[54] Sénési S, Thepenier R-M. Indices d'instabilité et occurrence d'orage: Le cas de l'Île-de-France. La Météorologie. (1997).

[55] Agli H. Recherches sur la variabilité pluviométrique dans le bassin-versant du Niger au Bénin. Mémoire du DEA" climats et contraintes climatiques", Centre de Recherche de Climatologie, Dijon. (1995).

[56] Vissin E. Impact de la variabilité climatique et de la dynamique des états de surface sur les écoulements du bassin béninois du fleuve Niger, https://tel.archives-ouvertes.fr/tel-00456097/document, (2007).

[57] Lawin EA, Afouda A, Gosset M, Lebel TH. Caractéristiques évènementielles des pluies en zone soudanienne : apport des données a haute résolution ammacatch a l'analyse de la variabilité de la mousson ouestafricaine en climat soudanien. Annales des Sciences Agronomiques. 13(1) (2010).

[58] PERARD J. Orages au Bénin : modèle scientifique, modèle traditionnel. Orages au Bénin : modèle scientifique, modèle traditionnel. (14): 33–40 (1992).

[59] Rudlosky SD, Shea DT. Evaluating WWLLN performance relative to TRMM/LIS: EVALUATING WWLLN RELATIVE TO TRMM/LIS. Geophysical Research Letters. 40(10): 2344–2348 (2013).

[60] Holzworth RH, Bering EA, Kokorowski MF, Lay EH, Reddell B. Balloon observations of temporal variation in the global circuit compared to global lightning activity. Advances in Space Research. (2005).

[61] Hutchins ML, Holzworth RH, Virts KS, Wallace JM, Heckman S. Radiated VLF energy

differences of land and oceanic lightning: ENERGY DIFFERENCE OF LAND AND OCEANIC LIGHTNING. Geophysical Research Letters. 40(10): 2390–2394 (2013).

[62] Dowden RL, Brundell JB, Rodger CJ. VLF lightning location by time of group arrival (TOGA) at multiple sites. Journal of Atmospheric and Solar-Terrestrial Physics. 64(7): 817–830 (2002).

[63] Dowden RL, Holzworth RH, Rodger CJ, Lichtenberger J, Thomson NR, Jacobson AR, *et al.* World-wide lightning location using VLF propagation in the Earth-ionosphere waveguide. IEEE Antennas and Propagation Magazine. 50(5): 40–60 (2008).

[64] Bürgesser RE, Nicora MG, Avila EE. Characterization of the lightning activity of "Relámpago del Catatumbo." Journal of Atmospheric and Solar-Terrestrial Physics. 77: 241–247 (2012).

[65] Lay EH, Holzworth RH, Rodger CJ, Thomas JN, Pinto Jr O, Dowden RL. WWLL global lightning detection system: Regional validation study in Brazil. Geophysical Research Letters. 31(3) (2004).

[66] Rodger CJ, Brundell JB, Dowden RL. Location accuracy of VLF World-Wide Lightning Location (WWLL) network: Post-algorithm upgrade. Annales Geophysicae. 23(2): 277–290 (2005).

[67] Rodger CJ, Werner S, Brundell JB, Lay EH, Thomson NR, Holzworth RH, *et al.* Detection efficiency of the VLF World-Wide Lightning Location Network (WWLLN): initial case study. Annales Geophysicae. 24(12): 3197–3214 (2006).

[68] Rodger CJ, Brundell JB, Holzworth RH, Lay EH, Crosby NB, Huang T-Y, *et al.* Growing Detection Efficiency of the World Wide Lightning Location Network. In AIP Conference Proceedings. AIP: Corte (France) pp. 15–20 (2009).

[69] Jacobson AR, Holzworth R, Harlin J, Dowden R, Lay E. Performance assessment of the world wide lightning location network (WWLLN), using the Los Alamos sferic array (LASA) as ground truth. Journal of Atmospheric and Oceanic Technology. 23(8): 1082–1092 (2006).

[70] Abreu D, Chandan D, Holzworth RH, Strong K. A performance assessment of the World Wide Lightning Location Network (WWLLN) via comparison with the Canadian Lightning Detection Network (CLDN). Atmospheric Measurement Techniques. 3(4): 1143–1153 (2010).

[71] Abarca SF, Corbosiero KL, Galarneau Jr TJ. An evaluation of the worldwide lightning location network (WWLLN) using the national lightning detection network (NLDN) as ground truth. Journal of Geophysical Research: Atmospheres. 115(D18) (2010).