Scientific
Research
Publishing

Application of Machine Learning Methods on Climate Data and Commercial Microwave Link Attenuations for Estimating Meteorological Visibility in Dusty Condition

Wend Yam Serge Boris Ouedraogo[1], Sibiri Tiemounou[2], Moumouni Djibo[1], Ali Doumounia[1,3], Serge Roland Sanou[1,4], Moumouni Sawadogo[1], Idrissa Guira[5], François Zougmore[1]

[1]Laboratoire Matériaux et Environnement, Université Joseph KI-ZERBO, Ouagadougou, Burkina Faso
[2]Institut du Génie Informatique et Télécommunications, Ecole Polytechnique de Ouagadougou, Ouagadougou, Burkina Faso
[3]Ecole Normale Supérieure, Koudougou, Burkina Faso
[4]Autorité de Régulation des Communications Électroniques et des Postes, Ouagadougou, Burkina Faso
[5]Telecel Faso SA, Ouagadougou, Burkina Faso
Email: ouedboris@gmail.com

Abstract

Accurately measuring meteorological visibility is an important factor in road, sea, rail, and air transportation safety, especially under visibility-reducing weather events. This paper deals with the application of Machine Learning methods to estimate meteorological visibility in dusty conditions, from the power levels of commercial microwave links and weather data including temperature, dew point, wind speed, wind direction, and atmospheric pressure. Three well-known Machine Learning methods are investigated: Decision Trees, Random Forest, and Support Vector Machines. The correlation coefficient and the mean square error, between the visibility distances estimated by Machine Learning methods and those provided by Burkina Faso weather services are computed. Except for the SVM method, all the other methods give a correlation coefficient greater than 0.90. The Random Forest method presents the best result both in terms of correlation coefficient (0.97) and means square error (0.60). For this last method, the best variables that explain the model are selected by evaluating the weight of each variable in the model. The best performance is obtained by considering the attenuation of the microwave signal and the dew point.

Keywords

Meteorological Visibility, Commercial Microwave Link, Machine Learning, Climate Parameters, Dust, Attenuation

1. Introduction

It is undeniable that the reduction of meteorological visibility increases the risk of transportation accidents. Meteorological visibility is defined as the greatest distance in a given direction at which a prominent black object can be seen and recognized in daylight, or the greatest distance at which it could be seen and recognized at night if general lighting were brought up to the level of normal daylight [1].

Several works have already been conducted to study and estimate the meteorological visibility under different conditions. O. Kolawole *et al.* [2] used meteorological visibility information to estimate the availability of optical wireless communication links. N. Hautiére *et al.* [3] [4] proposed on-board vision systems, specifically cameras, to detect weather conditions and estimate the visibility distance in foggy conditions. By still using images, H. Chaabani *et al.* [5] proposed a deep learning method to estimate the meteorological visibility range, again in foggy conditions. Another approach for estimating meteorological visibility in the presence of fog has been explored by David *et al.* [6] [7]. They exploit the attenuation of commercial microwave links of mobile phone operators. In [8], L. Ortega *et al.* have used Machine Learning methods on weather data to classify meteorological visibility.

In West Africa, and particularly in Burkina Faso, fog is an extremely rare weather phenomenon, but some severe meteorological visibility reductions sometimes occur due to dust sheets. These dust sheets can remain for several days and are generally caused by the harmattan, a wind that sweeps across the Sahara, West Africa, and Central Africa, annually between November and March [9].

To our knowledge, very few works in the literature have addressed the problem of estimating meteorological visibility in dusty conditions. In recent work [10], we proposed a linear regression model for estimating the meteorological visibility in dusty conditions, using the attenuation of the microwave signal propagating between two telecommunication antennas. The present study seeks to improve the meteorological visibility estimation model, by taking into account, in addition to the attenuation of commercial microwave links (CML) signal, the following weather parameters: temperature, dew point, wind speed, wind direction, and atmospheric pressure.

For visibility prediction, Machine Learning methods are used. Note that, Machine learning (ML) methods stand for data analytics technique, that teaches computers to reason like a human should do naturally, *i.e.*, learn from experience. ML methods use optimization algorithms to "learn" information directly from data without relying on a predetermined equation as a model. The algorithms adaptively improve their performance as the number of samples available for learning increases. ML methods can be split into two types of techniques: supervised learning, which trains a model on known input and output data so that it can predict future outputs, and unsupervised learning, which finds hidden

patterns or intrinsic structures in input data. In addition, supervised learning algorithms can be divided into two groups: regression (when the target variable is quantitative) et classification (when the target variable is qualitative). Since the study aims to predict visibility (which is a quantitative variable) then the regression methods will be used.

This paper is organized as follows: After a description of the dataset, the different Machine Learning methods that will be used are briefly described. Then visibility prediction results, obtained during a harmattan season in Burkina Faso are presented, before concluding.

2. Materials and Methods

2.1. The Data

To predict the meteorological visibility, two kinds of data are used: the attenuations of the power level of a commercial microwave link, and weather data.

The attenuation values of the microwave signal come from the network of Telecel Faso, a mobile phone operator in Burkina Faso. The UNIVERSITE-NABIYAAR link, shown in **Figure 1** is considered. This link is 1.45 km long, transmitting at 13.143 GHz, vertically polarized, and located 2.2 km from the Ouagadougou international airport.

When a microwave signal propagates between a transmitting and a receiving antenna, it is gradually attenuated due to the free space propagation [11], atmospheric gases and associated effects [12]. If one knows the transmission frequency (f[Ghz]) and the link length (d[km]), the attenuation due to free space propagation is constant and is determined by Equation (1) [11].

$$L_F\left(\text{dB}\right) = 32.4 + 20\log_{10}\left(d_{\text{km}}\right) + 20\log_{10}\left(f_{\text{Ghz}}\right) \tag{1}$$

The attenuation due to atmospheric gases and associated effects depends on the pressure, the temperature, and the water vapor concentration [12].
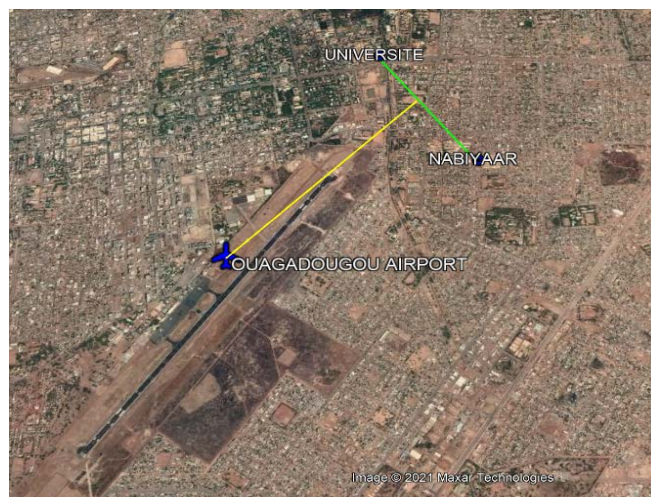
Figure 1. Localization of Ouagadougou international airport and Université-NabiYaar CML.

The sum of attenuations due to free space propagation, atmospheric gases and associated effects, is denoted baseline and noted $b(t)$. The baseline can time varies, according to the time variation of the weather parameters over the link.

Moreover, if the microwave propagating between 2 antennas encounters a weather event (rain, snow, fog, dust, etc.), it undergoes additional attenuation due to this event. This study considers the case of the presence of a dense dust sheet on the microwave link, illustrated in **Figure 2**.

Let's note by T_x, the transmitted power level, typically constant in the absence of automatic gain control, and $R_x(t)$ the received power level. The attenuation of the microwave signal, $a(t)$, is:

$$a(t) = T_x - R_x(t) = b(t) + \delta(t) \tag{2}$$

where $b(t)$ is the baseline and $\delta(t)$ is the additional attenuation due to dust.

In this study, the emitted and power received levels (respectively T_x and $R_x(t)$) were collected by an acquisition system installed at the Laboratoire Matériaux et Environnement (LA.ME) of Université Joseph KI-ZERBO (Burkina Faso), over the period from November 2019 to April 2020. This system records the power levels of Telecel Faso's commercial links with a sampling period of one minute. These data are then downsampled at a one-hour period. For each dust event, the baseline is manually determined and then additional attenuation due to the dust is derived.

The second king of data used in this study is the weather parameters including temperature, dew point, wind speed, wind direction, atmospheric pressure, and meteorological visibility measurement provided by Burkina Faso weather services. These data are recorded at a one-hour sampling period at the Ouagadougou international airport and are available at [13].

Table 1 shows the attenuations due to dust for the UNIVERSITE-NABIYAAR link, as well as the weather parameters for the dusty event of November 15, 2019, between 4:00 at 11:00.
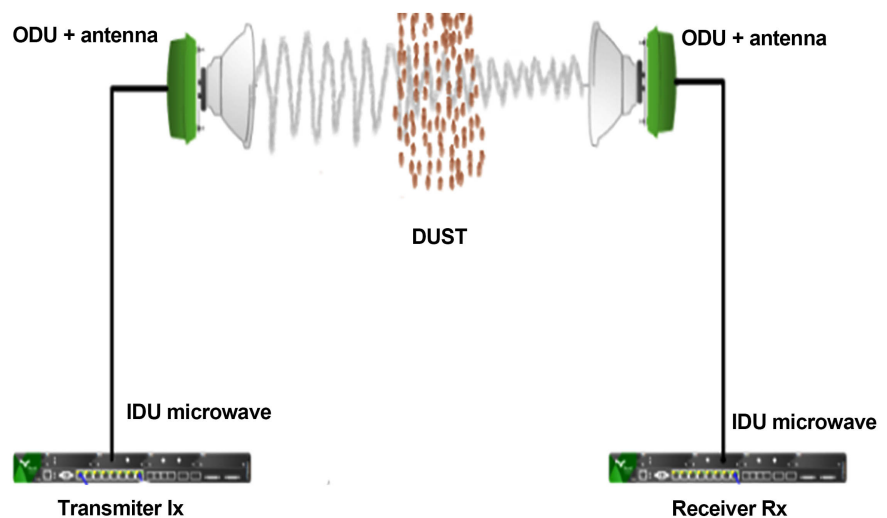
Figure 2. Attenuation of the CML signal in the presence of dust.

Table 1. CML attenuations and weather data for November 15, 2019, between 4:00 at 11:00.

Time	Temperature (˚C)	Dew point (˚C)	Wind speed (km/h)	Wind direction (˚)	Atmospheric pressure (mbar)	Measured meteorological visibility (km)	Attenuation (dB)
4 h	20	15	3.7	320	1010	8	0.27
5 h	20	15	5.56	320	1010	8	0.27
6 h	20	15	3.7	330	1011	6	0.28
7 h	22	15	1.85	330	1011	2.5	2.27
8 h	27	10	9.26	20	1012	2.5	2.27
9 h	30	9	11.1	40	1012	4.5	2.76
10 h	33	7	13	50	1012	8	0.27
11 h	35	6	16.7	40	1012	≥10	0.001

2.2. Machine Learning Methods

Supervised Machine Learning methods can be mainly split into three groups:

- **Decision tree algorithms** which build a model of decisions based on actual values of attributes in the data.

- **Ensemble methods** which stand for machine learning technique that combines several base models in order to produce one optimal predictive model.

- **Instance-based algorithms** that build up a training dataset and compare new data to the dataset using a similarity measure in order to find the best match and make a prediction.

In this study, well-known and most used methods have been considered for each group: Classification And Regression Trees (CART) [14], Random Forest [15], and Support Vector Machines [16]. Note that all these methods can be used for classification as well as regression problems.

The CART algorithm is a supervised learning algorithm that uses a flowchart, like a tree structure, to show the predictions that result from a series of feature-based splits. It starts with a root node and ends with a decision made by leaves. The tree is built by the recursive partition of each node according to an attribute maximizing the homogeneity gain.

The random forest is an improved version of decision trees [15]. This algorithm builds hundreds or even thousands of more or less decorrelated trees.

As far as concerned the Support Vector Machines [16], it aims at finding the best decision boundary (also called hyperplane) that can segregate n-dimensional space into classes so that one can easily put the new data point in the correct category.

3. Results and Discussion

3.1. ML Steps Description

For building the database, CML attenuations and weather parameters of 17 days

with dusty events are used.

Before applying the three Machine Learning algorithms, the data were previously normalized using formula (3):

$$X_{norm} = \frac{X - m}{\sigma} \tag{3}$$

where X represents a given descriptive variable, m and σ are the mean and standard deviation of this variable, respectively. X_{norm} is the normalized variable.

Then, the data were randomly split into two groups: training data (80%) and test data (20%). The different Machine Learning methods were applied to training data and then the obtained models are validated on the test data. **Figure 3** illustrates the processing steps.

3.2. Looking for the Hyperparameters

To apply ML methods to the dataset, libraries provided by Python are used. To find the model for the dataset, one needs to determine the hyperparameters for each method during the learning step. To do so, for each method, different values are tested for each parameter, thus hyperparameters correspond to those which give the best prediction performance. **Table 2** shows the results of this stage.

3.3. Results and Discussions

The 3 Machine Learning methods described below are successively applied to the data sets. For each method, a comparison of the measured and predicted visibility distances is performed by computing the Pearson correlation coefficient (ρ) [17] between these two variables, as well as the Mean Square Error (MSE) [17]. **Table 3** presents the performances of the 3 models.
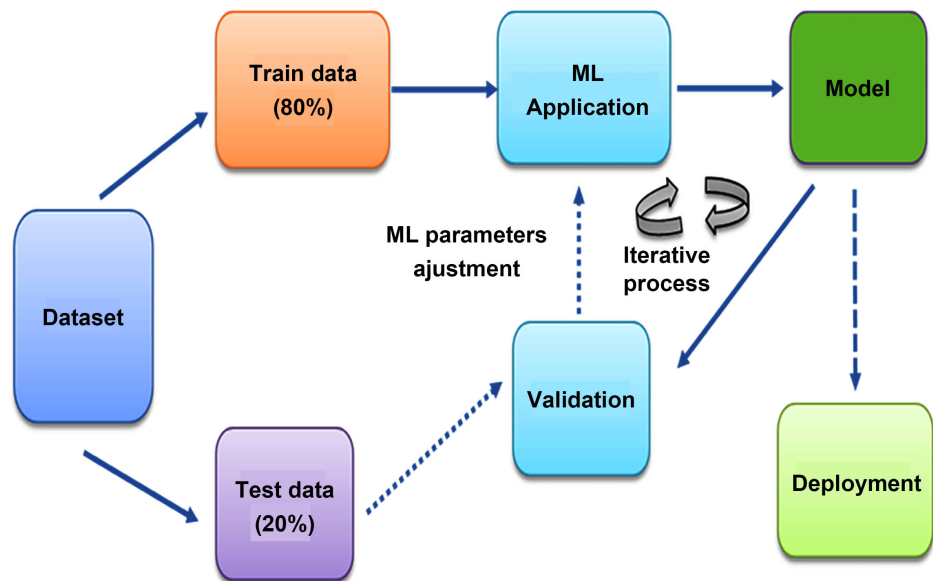
Figure 3. ML methods application steps.

By considering the correlation coefficient, the Random Forest method has the best result with a correlation coefficient of 0.97. It is followed by CART (0.94) and SVM (0.84). The random Forest method has again the best result in terms of MSE (0.60). It is still followed in order by CART (0.99) and SVM (2.65).

The Random Forest method is therefore the one that presents the best results and will be used for the rest of the study.

Another aspect of this study is to select the best variables that explain the obtained model. To do this, for the Random Forest method, the weight of each variable in the meteorological visibility prediction model is evaluated by the "Forward Selection" method [18]. This method tests the effect of adding a new variable in the prediction accuracy, using a regression criterion. A variable is selected if it significantly improves the performance of the model. The process is repeated until no remaining variable significantly improves the performance of the model. Here, the criterion for selecting the variables is the correlation coefficient between the measured and the predicted visibility distance. Table 4 shows the result of the variable selection method.

Table 2. Description of the different parameters for each method.

ML methods	Hyperparameters
CART	Homogeneity gain = Mean Square Error
	Maximum depth of the tree = 4
Random Forest	Homogeneity gain = Mean Square Error
	Maximum depth of the tree = 6
	Number of trees = 10
SVM	Kernel = Radial Basis Function
	Regularization parameter = 6.1

Table 3. Performance of the 3 ML models for weather visibility prediction.

Method	ρ	MSE
CART	0.94	0.99
Random Forest	0.97	0.60
SVM	0.84	2.65

Table 4. Results of Forward Selection method.

Number of variables	Variables names	ρ
1	CML attenuation	0.94
2	CML attenuation, Dew point	0.97
3	CML attenuation, Dew point, Wind speed	0.95
4	CML attenuation, Dew point, Wind speed, Atmospheric pressure	0.95
5	CML attenuation, Dew point, Wind speed, Atmospheric pressure, temperature	0.95
6	CML attenuation, Dew point, Wind speed, Atmospheric pressure, temperature, Wind direction	0.95

As one can see in Table 4, the CML attenuations alone yield a correlation coefficient of 0.94. This leads to believe that with this parameter alone, one can properly estimate the meteorological visibility. By taking into account the dew point, we obtain a correlation coefficient of 0.97, that is to say, an improvement of 0.03 representing the contribution of the dew point on the accuracy of the prediction. The accuracy of the model degrades when the other variables are taken into account (see Table 4).

4. Conclusions

In West Africa and particularly in Burkina Faso, the harmattan may cause each year, strong meteorological visibility reduction, with the suspension of dust for several days. This leads to safety risks in transportation. In this work, Machine Learning methods are used in order to predict meteorological visibility in dusty conditions, from the attenuation of the signal of a commercial microwave link provided by the mobile phone operator Telecel Faso, and weather data (temperature, dew point, wind speed, wind direction, atmospheric pressure). Among the 3 Machine Learning methods that have been tested, Random Forest gives the best performance in terms of correlation coefficient (0.97) and Mean Square Error (0.60), between recorded and estimated visibility distance. By using the Forward Selection method, we show that with only the microwave signal attenuations, the correlation coefficient between recorded and estimated visibility distance is 0.94. The combination of attenuation and dew point gives the best visibility prediction results.

Future work will investigate dust automatic detection methods, as well as automatic estimation of microwave attenuation baseline.

Acknowledgements

Conflicts of Interest

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

References

[1] Hautiére, N. (2005) Détection des conditions de visibilité et estimation de la distance de visibilité par vision embarquée. Ph.D. thesis, Université Jean Monnet, Saint-Etienne.

[2] Kolawole, O., Afullo, T. and Mosalaosi, M. (2019) Estimation of Optical Wireless Communication Link Availability Using Meteorological Visibility Data for Major Locations in South Africa. 2019 *Photonics & Electromagnetics Research Symposium*, Rome, 17-20 June 2019, 319-325.

https://doi.org/10.1109/PIERS-Spring46901.2019.9017842

[3] Hautiére, N., Tarel, J.P., Lavenant, J. and Aubert, D. (2009) Automatic Fog Detection and Estimation of Visibility Distance through Use of an Onboard Camera. *Machine Vision and Applications*, **17**, 8-20.
https://doi.org/10.1007/s00138-005-0011-1

[4] Hautiére, N., Babari, R., Dumont, E., Brémond, R. and Paparoditis, N. (2011) Estimating Meteorological Visibility Using Cameras: A Probabilistic Model-Driven Approach. In: Kimmel, R., Klette, R. and Sugimoto, A., Eds., *Computer Vision—ACCV* 2010, Lecture Notes in Computer Science, Vol. 6495, Springer, Berlin, 243-254.
https://doi.org/10.1007/978-3-642-19282-1_20

[5] Chaabani, H., Werghi, N., Kamoun, F., Taha, B., Outay, F. and Yasar, A. (2018) Estimating Meteorological Visibility Range under Foggy Weather Conditions: A Deep Learning Approach. *Procedia Computer Science*, **141**, 478-483.
https://doi.org/10.1016/j.procs.2018.10.139

[6] David, N., Pinhas, A. and Hagit, M. (2013) The Potential of Commercial Microwave Networks to Monitor Dense Fog-Feasibility Study. *Journal of Geophysical Research*: *Atmospheres*, **118**, 11750-11761. https://doi.org/10.1002/2013JD020346

[7] David, N., Sendik, O., Messer, H. and Pinhas, A. (2015) Cellular Network Infrastructure: The Future of Fog Monitoring? *Bulletin of the American Meteorological Society*, **96**, 1687-1698. https://doi.org/10.1175/BAMS-D-13-00292.1

[8] Ortega, L., Otero, L.D. and Otero, C. (2019) Application of Machine Learning Algorithms for Visibility Classification. 2019 *IEEE International Systems Conference*, Orlando, 8-11 April 2019, 1-5. https://doi.org/10.1109/SYSCON.2019.8836910

[9] Schwanghart, W. and Schütt, B. (2008) Meteorological Causes of Harmattan Dust in West Africa. *Geomorphology*, **95**, 412-428.
https://doi.org/10.1016/j.geomorph.2007.07.002

[10] Djibo, M., Serge Boris Ouedraogo, W., Doumounia, A., Sanou, S., Sawadogo, M., Guira, I. and Zougmoré, F. (2021) Estimation de la visibilité météorologique à l'aide des liens micro-ondes commerciaux de telecommunications. *Journal de Physique de la Soaphys*, **3**, C21A03-1-C21A03-4. https://doi.org/10.46411/jpsoaphys.2021.01.03

[11] ITU-R (2019) Calculation of Free-Space Attenuation. Recommendation P.525-4, Electronic Publication, Geneva.

[12] ITU-R (2019) Attenuation by Atmospheric Gases and Related Effects. Recommendation P.676-12, Electronic Publication, Geneva.

[13] https://fr.weatherspark.com/h/td/147795/Météo-historique-à-Aéroport-international-de-Ouagadougou-Burkina-Faso

[14] Breiman, L., Friedman, J., Stone, C.J. and Olshen, R.A. (1984) Classification and Regression Trees. Chapman and Hall/CRC, London, 368 p.

[15] Breiman, L. (2001) Random Forests. *Machine Learning*, **45**, 5-32.
https://doi.org/10.1023/A:1010933404324

[16] Steinwart, I. and Christmann, A. (2008) Support Vector Machines. Springer, New York.

[17] James, G., Witten, D., Hastie, T. and Tibshirani, R. (2021) An Introduction to Statistical Learning: With Applications in R, 2nd Edition, Springer, Berlin.
https://doi.org/10.1007/978-1-0716-1418-1

[18] Hocking, R.R. (1976) The Analysis and Selection of Variables in Linear Regression. *Biometrics*, **32**, 1-49. https://doi.org/10.2307/2529336