WMO Aeronautical Meteorology Scientific Conference 2017
6 - 10 November 2017
Centre International de Conférences - Météo-France - Toulouse - France

Session 1 – Science underpinning meteorological observations, forecasts, advisories and warnings
1.6 – Observation, nowcast and forecast of future needs
1.6.1 – Advances in observing methods and use of observations

Remote-Ground based observations Merging Method For Visibility and Cloud Ceiling Assessment During the Night Using Data-Mining Algorithms

Driss BARI, Direction de la Météorologie Nationale, Casablanca, Morocco
bari.driss@gmail.com

Abdelhak LEMKHENTER, Ecole CentraleSupelec, Paris, France
abdelhak.lemkhenter@student.ecp.fr

Speaker: Driss BARI

Abstract

The occurrence of adverse cloud ceiling and visibility conditions that restricted the flow of air traffic in major airport's terminals is one of the main causes of aircraft delays and is crucial for air traffic safety and economic issues. In this study, Data-mining methods are applied to ground-based observations and satellite data to develop automated algorithms for the diagnosis of visibility and low cloud ceiling, during the night, in areas where no local observations are available. To achieve this, a database of hourly records of satellite data and conventional meteorological parameters has been used. It covers the winter months from January 2014 to February 2017 for 15 meteorological synoptic stations in the north-western part of Morocco. Based on ensemble approach, the developed classification decision trees have been used for the separate detection of fog and low cloud ceiling using only satellite data while the regression decision trees have been used for estimating the visibility and low cloud ceiling using a combination of ground-based observations and remote ones. Results show that detection of both phenomena has percent correct and probability of detection above 70% with false alarm ratio below 30%. The performance evaluation of the continuous parameter's estimation indicates a mean absolute error of 675m (resp. 540 m) with a 0,96 (resp. 0,94) correlation and a root mean-square error of 1120m (resp. 1070m) for visibility (resp. low cloud ceiling).

1. Introduction

Fog and low cloud ceiling are two meteorological phenomena that result in reduced horizontal and vertical visibility. They have strong impact on the safety of people as well as on various economic (such as aviation, marine and road traffic) and military activities.

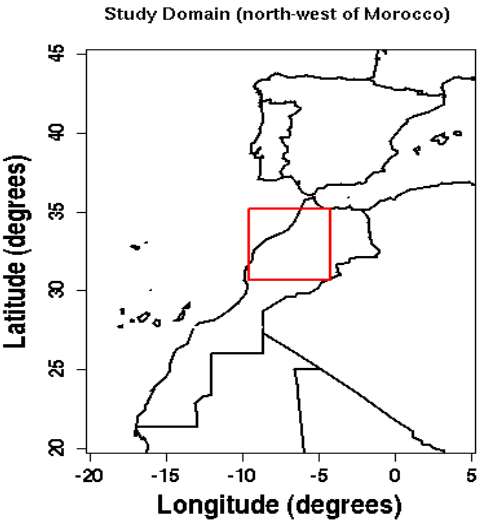
From the observational point of view, spatial detection remains very difficult where no local observations is available due to the low density of synoptic meteorological stations; so continuous data on local conditions is of a particular importance and is an essential requirement for this purpose. Thanks to their large coverage, high spatial resolution (3 km^2) and temporal frequency (15 min), geostationary satellite data have been suggested as a possible source of fog and low cloud detection. Nevertheless, derived satellite products have some limitations such as distinction between fog and low cloud, and fog identification when it is obscured by the clouds (Cermak and Bendix, 2007).

Hence, many efforts are still needed to improve the spatial detection of these two phenomena in order to minimize the uncertainties in terms of duration and localization. As an alternative, supervised machine-learning techniques are used to discover patterns in data and to develop associated classification and parameter estimation algorithms. These data-mining methods, used in a Knowledge Discovery from Databases (KDD) procedure, have been applied to the cloud-ceiling height assessment problem (Bankert *et al.* 2004).

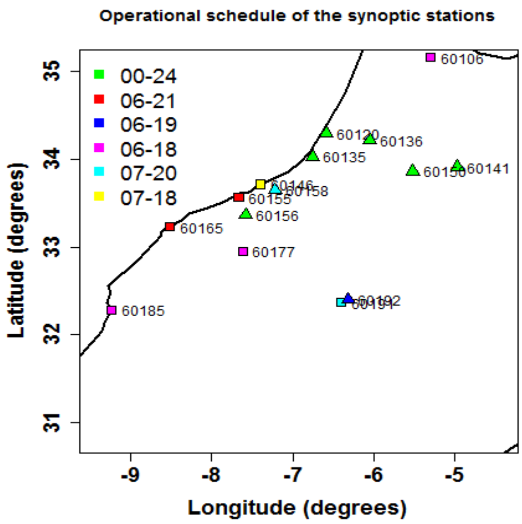
The main objective of this study is to evaluate the potential of Data Mining methods in separately detecting fog and low cloud ceiling from satellite data and to estimate the visibility and low cloud-ceiling height from ground-based standard meteorological parameters and/or satellite data. The study domain covers the northwestern part of Morocco (Fig. 1(a)). This region is a fog-prone area (Bari *et al.* 2016) and contains many airports.

2. Methodology

To achieve the main objective of this study, hourly data from *Meteosat Second Generation* (MSG) and conventional meteorological parameters, from 15 synoptic stations, at nighttime have been used (Fig. 1(b)). The created database covers the winter months (December, January and February) from January 2014 to February 2017.



(a)



(b)

Figure 1: (a) Map showing the study domain covering the northwestern part of Morocco. (b) Map showing the locations of the synoptic stations and their operating schedule. Triangles refer to airports.

Satellite data consist of all infrared sensor channels at the four pixels over land that are closest to the latitude-longitude of each synoptic station. In addition to brightness temperature (BT) extracted from each infrared channel, the mutual differences of temperature (BTD) between these channels have also been used. Thus, the following derived satellite output are used: BT-IR3.9, BT-IR8.7, BT-IR10.8, BT-IR12.0, BT-IR13.4, BTD-IR3.9-IR8.7, BTD-IR3.9-IR10.8, BTD-IR3.9-IR12.0, BTD-IR8.7-IR10.8, BTD-IR8.7-IR12.0, and BTD-IR10.8-IR12.0.

The used standard meteorological parameters are temperature, dew-point temperature, relative humidity, vapor pressure, total cloud coverage, lowest-level cloud coverage, lowest-level cloud base height, visibility, present weather, wind direction, wind speed, pressure, mean sea level pressure.

The KDD-produced algorithms consist of algorithms for detecting separately fog and low cloud ceiling based on classification decision trees, and algorithms for estimating visibility and cloud ceiling height based on regression decision trees. The classification algorithms used satellite data only while the algorithms of estimation used a combination of conventional meteorological parameters with satellite data. To develop the boosted trees, XGBoost (Extreme Gradient Boosting), which is a scalable machine learning system for tree boosting, has been used (Chen and Guestrin, 2016). In the tree ensemble methods, we are learning functions (trees) instead of learning numerical weights by additive training (boosting). To learn the set of functions used in the developed model, we minimize a regularized learning objective:

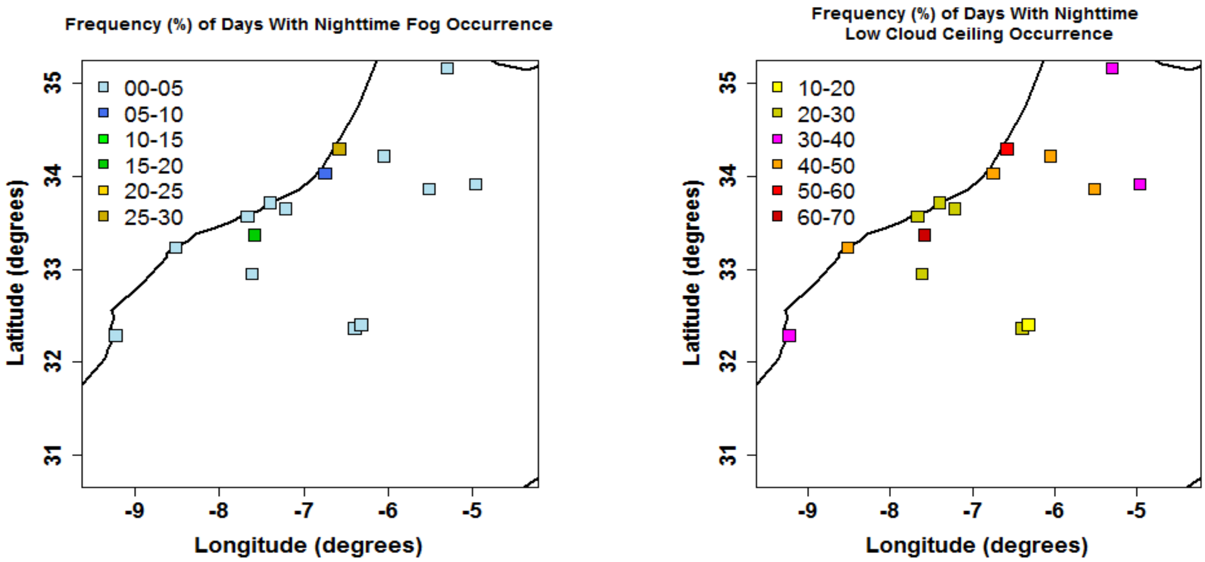
$$\text{Objective} = \text{Training Loss} + \text{Regularization}$$

Where training loss term measures how well model fit on training data and the additional regularization term measures complexity of the model and helps to smooth the final learnt weights to avoid over-fitting. This approach aims to have predictive and simple functions. In this work, XGBoost is used for supervised learning problems, where we use the training data (with multiple features) x_i to predict a target variable y_i. The training (75% of all data) and testing (25% of all data) sets have been created by a random split of all available data for all locations together and they verify the condition of equality between the numbers of occurrence and no-occurrence of the studied phenomenon. This is done to overcome the fact that the developed algorithms may converge on a solution by which the largest class is always forecast (Fabian *et al.* (2007), Bari and El Khlifi (2015)).

Regarding the detection of fog or low cloud ceiling, we constructed the target binary attribute based on a combination of some meteorological parameters. Thus, an observed fog is identified when visibility is below 1000 m and present weather ranges from 40 to 49 or is equal to 11 or 12. This is done to distinguish reduced visibility by fog from its reduction by another phenomenon such as heavy rain. On the other hand, an observed low cloud ceiling is defined as the lowest level that has at least a cloud coverage equal or greater than 4/8 with cloud base height below 1000 m.

3. Results

The study domain covers the northwestern part of Morocco, which contains 15 synoptic meteorological stations (Fig. 1(a)). During nighttime, only six stations, located in airports, are operating during the whole nighttime and the remaining stations operate up to 5 hours (Fig. 1(b)). Over the studied period, the low cloud ceiling (64% of total days, Fig. 2(b)) is more frequent than fog (21% of total days, Fig. 2(a)) over the region. From a geographical standpoint, fog is often rare (frequency <= 5%) per each station while low cloud ceiling frequency is greater than 20% and reached 64% for some synoptic stations.



(a) (b)

Figure 2: Climatology of days with (a) fog and (b) low cloud ceiling during nighttime. Frequency (%) is defined as the ratio of number of days with phenomenon during nighttime per total available days for each synoptic station.

To evaluate the potential of the KDD-produced algorithms of classification and estimation, the results of the data-mining experiments on the testing dataset are presented below. Frequency bias (FBIAS) will be used to measure the reliability of the developed models for diagnosis the two phenomena while percent correct (PC), probability of detection (POD) and false-alarm ratio (FAR) will be used to measure the accuracy of the algorithms of detection. The equitable threat score (ETS) and true skill score (TSS) are used as skill scores.

It is seen clearly from Table 1 that classification algorithms have a frequency bias of about 1 (the perfect score). In addition, classification of both phenomena was successful with percent correct and probability of detection above 70% and false alarm rate below 30%. Using only satellite data, it is found that the performance of detection over the entire study domain is slightly higher for fog (TSS=0.66 and ETS=0.49) than for low cloud ceiling (TSS=0.41 and ETS=0.26).

Table 1: Fog and Low Cloud ceiling classification performance statistics on the testing dataset

Phenomenon	FBIAS	PC	POD	FAR	TSS	ETS
Fog	1.01	0.83	0.83	0.17	0.66	0.49
Low cloud ceiling	1.02	0.71	0.72	0.30	0.41	0.26

To evaluate the generalization error of the developed model, the verification scores of the unified model, applied to individual stations, have been investigated (Table 2). Four airports are selected because they proved the high frequency of both phenomena. The results show some limitations of the developed model in terms of localization, in particular for low cloud ceiling detection. The worst results were obtained at Kenitra station (WMOID: 60120, POD=0.49, FAR=0.51 and ETS=0.08). For fog detection, the lowest probability of detection (=0.57) were observed at Sidi Slimane station (WMOID: 60136) while the worst TSS (0.3) were found for Kenitra station.

Table 2: Fog and Low Cloud ceiling classification performance statistics of the unified model applied to individual station, on the testing dataset.

Phenomenon	Fog				Low cloud ceiling			
WMO ID	**POD**	**FAR**	**ETS**	**TSS**	**POD**	**FAR**	**ETS**	**TSS**
60120	1.00	0.17	0.25	0.30	0.49	0.51	0.08	0.15
60135	0.75	0.19	0.47	0.63	0.61	0.35	0.20	0.33
60136	0.57	0.14	0.39	0.51	0.76	0.24	0.37	0.54
60156	0.79	0.14	0.27	0.45	0.84	0.25	0.21	0.32

To evaluate the performance of estimating the continuous parameter, correlation coefficient (CC) is used to measure the relationship of the algorithm output with observation, mean absolute error (MAE) and root-mean-square error (RMSE) are used to measure the accuracy of the algorithms.

Table 3 summarizes the visibility and low cloud ceiling height estimation performance statistics on the testing dataset. The lowest performance is related to estimation algorithms using only satellite data. This shows clearly the complexity of estimating both local meteorological parameters from remote observations. On the other hand, when using local meteorological parameters with or without satellite data, the performance of the developed algorithms is improved. Thus, the performance evaluation of the continuous parameter's estimation, on the testing set taking into account the conventional meteorological parameters, indicates a mean absolute error of 675m (resp. 540 m) with a 0,96 (resp. 0,94) correlation and a root mean-square error of 1170m (resp. 1070m) for visibility (resp. low cloud ceiling height).

Table 3: Performance statistics of visibility and height of low cloud ceiling estimation on the testing dataset

Parameter	Visibility			Cloud ceiling height		
Data source	**RMSE**	**MAE**	**CC**	**RMSE**	**MAE**	**CC**
Satellite + Meteo	1167.81	676.92	0.96	1071.27	542.21	0.94
Meteo	1170.57	675.24	0.96	1120.68	596.42	0.93
Satellite	2659.25	1892.36	0.78	2448.38	1648.28	0.93

4. Conclusions

Motivated by aeronautical requirements for more accurate assessment of visibility and cloud ceiling, an improved utilization of available observation data sources (local meteorological parameters and satellite data) to detect separately fog and low cloud ceiling and to estimate visibility and low cloud ceiling, using data mining methods, has been developed.

The results analysis demonstrate the potential of using Data mining methods to develop algorithms from ground-based and remote observations. The classification algorithms well detect the fog and low cloud ceiling with a slightly higher performance for fog detection. However, the developed model have shown some limitations in terms of localization. The performance of estimation algorithms is higher when local meteorological parameters are taking into account during the development procedure.

Since results have indicated that kDD-produced algorithms could be geographically dependent, our future research will concentrate on the enhancement of the predictors such as NWP model output and physiographic features (environmental conditions, vicinity to ocean, etc.).

References:

Bankert R. L., M. Hadjimichael, A. P. Kuciauskas, W. T. Thompson, and K. Richardson. **Remote cloud ceiling assessment using datamining methods**. *Journal of Applied Meteorology*, 43(12):1929–1946, 2004.

Bari D., T. Bergot, and M. El Khlifi. **Local meteorological and large-scale weather characteristics of fog over the grand Casablanca region, Morocco**. *Journal of Applied Meteorology and Climatology*, 55(8):1731–1745, 2016.

Bari D. and M. El Khlifi. **LVP conditions at Mohamed V airport, Morocco: Local characteristics and prediction using neural networks**. *International Journal of Basic and Applied Sciences*, 4(4):354, 2015.

Cermak J. and J. Bendix. **Dynamical nighttime fog/low stratus detection based on meteosat seviri data: A feasibility study**. *In Fog and Boundary Layer Clouds: Fog Visibility and Forecasting*, pages 1179–1192. Springer, 2007.

Chen T. and C. Guestrin. **Xgboost: A scalable tree boosting system**. *In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794. ACM, 2016.

Fabbian D., R. de Dear, and S. Lellyett. **Application of Artificial Neural Network forecasts to predict fog at Canberra International Airport**. *Weather and forecasting*, 22(2):372–381, 2007.