

Automated multi-classifier recognition of atmospheric turbulent structures obtained by Doppler lidar

Anton Sokolov^{1,*}, *Egor Dmitriev*², *Ioannis Cheliotis*¹, *Hervé Delbarre*¹, *Elsa Dieudonne*¹, *Patrick Augustin*¹, and *Marc Fourmentin*¹

¹University of Littoral Cote d'Opale, Laboratory for Physico-Chemistry of the Atmosphere, Dunkerque, France

²Institute of Numerical Mathematics of Russian Academy of Sciences, Moscow, Russia

Abstract. We present algorithms and results of automated processing of LiDAR measurements obtained during VEGILOT measuring campaign in Paris in autumn 2014 in order to study horizontal turbulent atmospheric regimes on urban scales. To process images obtained by horizontal atmospheric scanning using Doppler LiDAR, the method is proposed based on texture analysis and classification using supervised machine learning algorithms. The results of the parallel classification by various classifiers were combined using the majority voting strategy. The obtained estimates of accuracy demonstrate the efficiency of the proposed method for solving the problem of remote sensing of regional-scale turbulent patterns in the atmosphere.

1 Introduction

Atmospheric turbulence is a key meteorological characteristic, responsible for the dispersion of the air pollution and the cloud formation. When wind speed measurements are available with sufficient spatial and temporal resolution, the estimation of turbulence parameters becomes possible. As an example, the ground-based Doppler wind lidar Leosphere WLS100 allows remote measurements of radial wind speed profiles with the temporal resolution of about one profile by second and the spatial resolution of 50 meters.

The measurement database used in this study is based on data of VEGILOT measurement campaign [1], which was held in Paris in September – October 2014. This campaign was aimed at studying urban atmospheric dynamics, air pollution and turbulent regimes from lidar measurements. More information about VEGILOT campaign could be found in [1].

Horizontal wind turbulent patterns was calculated based on horizontal radial wind scans (see [2, 3]). Typical examples of these images are presented at Fig.1. Three classes of local atmospheric patterns were introduced: Thermals, Rolls and Streaks. An additional fourth class 'Others' contains patterns that could not be classified as these three turbulence types.

* Corresponding author: anton.sokolov@univ-littoral.fr

Each type of turbulence structure forms a specific cloud pattern and could be observed on satellite images.

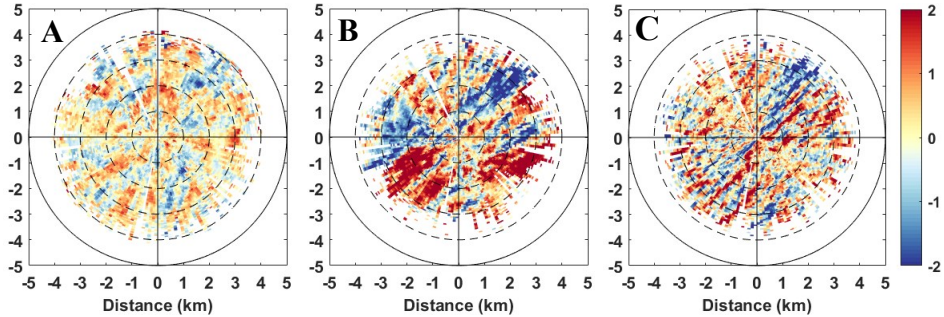


Fig. 1. Examples of turbulent structure patterns: A) Thermals observed on October 10 at 1111 UTC, B) Rolls observed on September 11 at 1100 UTC and C) Streaks observed on September 11 at 1957 UTC.

2 Description of Algorithms

A few thousands of images was obtained for the two-month campaign. The following supervise machine learning SML algorithms was applied to classify local atmospheric patterns:

- Parzen – Rozenblatt window (PRW);
- K-nearest-neighbors (KNN), a few number of neighbors were tested: $K=1$, $K=3$, and $K=5$;
- Error-correcting output codes with support vector machine (SVM);
- Quadratic discriminant analysis (QDA).

An expert classified 150 patterns to construct the training set [2]. The in-situ meteorological and satellite data were used in addition to lidar measurements. Since SML algorithms require numerical values characterizing patterns, Haralick texture features were applied for image classification [4]. Below we describe briefly how the features were calculated, see [2] for more details.

The following four statistics were calculated: Contrast, Homogeneity, Correlation and Energy. Each of those characteristics was calculated at different distances and directions of adjacency. The neighbors ranging from first until 30th for each statistic allow generating $30 \times 4 = 120$ angular functions [2, 4]. Each of these functions was characterized by the following three properties: amplitude (maximum value minus minimum value), integral, symmetry.

Three additional features were added to those $30 \times 4 \times 3 = 360$ textural features, namely the time of the scan, the average wind speed and the cosine fit error [2, 3]. Thus, 363 features (predictors) were calculated for each turbulent image.

To avoid the curse of dimensionality, the cross-validation stepwise forward selection method was implemented, since the dimension of feature vector largely exceeds the number of patterns in the learning ensemble. The number of optimal features selected [2, 5] vary between 2 and 20.

The SML classifiers were combined in one multiple classification (MC) algorithm using the majority voting strategy [6].

3 Results and Discussion

The total overall accuracy (TA) score for each classification technique was presented in the table 1. It shows the percentage of correctly classified images based on cross validation.

We can see that among applied standard SML algorithms the QDA has the best TA performance, following by SVM and KNN. The PRW algorithm has the lowest TA. It shows that SML algorithms, that produce more complex decision boundaries are less accurate for the given classification problem. We can also note the significant TA improvement for the proposed MC algorithm.

Table 1. Total overall accuracy for applied SML algorithms.

SML Algorithm	PRW	1NN	3NN	5NN	SVM	QDA	MC
TA, %	82.7	90	88.7	89.3	90.7	92	94.7

A confusion matrix for MC is presented in Fig. 2A. All types of turbulent patterns are identified with good accuracy. After the learning step, in the classification step, the MC algorithm was applied to a complete lidar dataset (test set) of 4557 patterns. Streaks was detected in 23% of cases, Rolls in 10% and Thermals in 17%.

In Fig. 2B, the distribution of turbulence types by the time UTC is shown. As expected, Streaks are generally observed during the nighttime, while Thermals and Rolls are detected in the daytime (see [2] for more details).

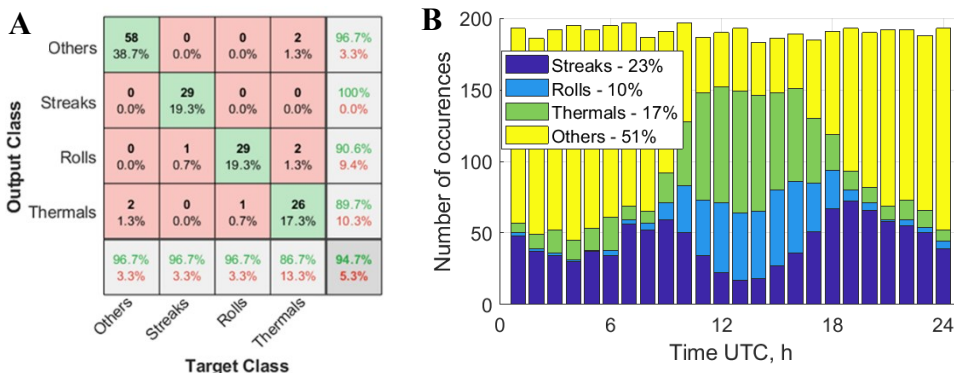


Fig. 2. A) Confusion Matrix for MC algorithm. B) Histogram of resulting distribution of turbulent classes by time.

4 Conclusions and Perspectives

On the basis of the results of large lidar dataset processing, we can conclude that the proposed method allows efficient solving the problem of turbulent regimes classification. The comparison of SML algorithms shows that the accuracy of relatively simple QDA is better than accuracy of other SML algorithms that construct more complicated decision boundaries (PRW, KNN and SVM). An important increase of the total accuracy could be achieved by combining a few SML algorithms in MC system. This technique was successfully applied for the study of atmospheric turbulence in Paris region in the autumn.

In the future studies, we are going to optimize the procedure of classifiers' combination and to increase the number of Haralick texture features. The proposed algorithm will be applied to characterize the atmospheric turbulence in coastal areas.

This work is a contribution to the CPER research project IREN and CLIMIBIO. The authors thank the French Ministère de l'Enseignement Supérieur et de la Recherche, the Hauts-de-France Region and the European Funds for Regional Economic Development for their financial support to this project. The work is supported by the CaPPA project. The CaPPA project (Chemical and Physical Properties of the Atmosphere) is funded by the French National Research Agency (ANR) through the PIA (Programme d'Investissement d'Avenir) under contract "ANR- 11-LABX-0005-01" and by the Regional Council Nord-Pas de Calais and the European Funds for Regional Economic Development (FEDER). The work was supported by the Moscow Center for Fundamental and Applied Mathematics (Agreement 075-15-2019-1624 with the Ministry of Education and Science of the Russian Federation) and Russian Foundation for Basic Research (RFBR, project № 20-07-00370). The aerosol lidar belongs to the air quality observatory, QUALAIR (<http://qualair.aero.jussieu.fr>), located at the Université Pierre et Marie Curie. Experiments presented in this paper were carried out using the CALCULCO computing platform, supported by SCoSI/ULCO (Service COmmun du Système d'Information de l'Université du Littoral Côte d'Opale).

References

1. A. Klein, F. Ravetta, J.L. Thomas, G. Ancellet, P. Augustin, R. Wilson, E. Dieudonné, M. Fourmentin, H. Delbarre, J. Pelon. Influence of vertical mixing and nighttime transport on surface ozone variability in the morning in Paris and the surrounding region. *Atmos. Environ.* **197**, 92-102. (2019) <https://doi.org/10.1016/j.atmosenv.2018.10.009>
2. I. Cheliotis, E. Dieudonné, H. Delbarre, A. Sokolov, E. Dmitriev, P. Augustin, M. Fourmentin. Detecting turbulent structures on single Doppler lidar large datasets: an automated classification method for horizontal scans, accepted for publication to *Atmos. Meas. Tech. journal*, 2020. <https://doi.org/10.5194/amt-2020-82>
3. W.L. Eberhard, R.E. Cupp, K.R. Healy. Doppler Lidar Measurement of Profiles of Turbulence and Momentum Flux. *Journal of Atmospheric and Oceanic Technology*, **6(5)**, 809-819. (1989) [https://doi.org/10.1175/1520-0426\(1989\)6\(5\)<809-819>2-B](https://doi.org/10.1175/1520-0426(1989)6(5)<809-819>2-B)
4. R.M. Haralick, I. Dinstein, K. Shanmugam. Textural Features for Image Classification. *IEEE Transactions on Systems, Man and Cybernetics*, **SMC-3(6)**, 610-621. (1973). <https://doi.org/10.1109/TSMC.1973.4309314>
5. A. Sokolov, E. Dmitriev, C. Gengembre, H. Delbarre. Automated classification of regional meteorological events in a coastal area using in-situ measurements. *J. Atmos. Ocean. Technol.* **JTECH-D-19-0120.1**. (2020) <https://doi.org/10.1175/JTECH-D-19-0120.1>
6. R. Ranawana and V. Palade. 'Multi-Classifer Systems: Review and a Roadmap for Developers'. *IJHIS*, **3, 1**, 35-61 (2006).