Remote sensing analysis of small scale dynamic phenomena in the Atmospheric Boundary Layer

Ioannis Cheliotis

Thèse de Doctorat

Mention: Physique
Spécialité: Milieux dilués et optiques fondamentales

présentée à ***l'Ecole Doctorale en Sciences Technologie et Santé (ED 585)***

de l'Université du Littoral Côte d'Opale

par

Ioannis CHELIOTIS

pour obtenir le grade de Docteur de l'Université du Littoral Côte d'Opale

Remote sensing analysis of small scale dynamic phenomena in the Atmospheric Boundary Layer

Soutenue le 17/02/2021, après avis des rapporteurs, devant le jury d'examen :

M. P. DUBUISSON, Professeur, Université de Lille	**Président**
M. R. WILSON, Maître de Conférences HDR, Université de Versailles Saint-Quentin-en-Yvelines	**Rapporteur**
M^me^ F. LOHOU, Maître de Conférences HDR, Université Toulouse III Paul Sabatier Nom de l'Université	**Rapporteur**
M. A. SOKOLOV, Maître de Conférences, Université du Littoral Côte d'Opale	**Examinateur**
M. G. ROUSSEL, Professeur, Université du Littoral Côte d'Opale	**Examinateur**
M. H. DELBARRE Professeur, Université du Littoral Côte d'Opale	**Directeur de thèse**
M^me^ E. DIEUDONNE, Maître de Conférences, Université du Littoral Côte d'Opale	**Co-encadrant**

Abstract

The aim of this thesis project was to study the coherent turbulent structures (convective rolls & streaks) and more generally the medium to large fluctuations in the atmospheric boundary layer using the observations recorded by a single Doppler (wind) lidar during a 2-month campaign in Paris, France. An innovative method was developed in order to classify automatically the radial wind speed patterns visible on the quasi-horizontal lidar scans, based on texture analysis parameters and supervised machine learning algorithms. A 150-case training ensemble was built using ancillary data (satellite pictures and weather observations) to ascertain the manual classification into four types: rolls, thermals, streaks and “others”. The performance of the classification process was assessed on the training ensemble using the 10-fold cross-validation method. A very satisfying 9% error was obtained for the Quadratic Discriminant Analysis algorithm, using only 5 texture analysis parameters classifiers. This process was then applied to classify the whole dataset (4577 lidar scans) and the results showed that the classified structures respected a plausible diurnal cycle and were associated with the meteorological parameters as expected by the theoretical knowledge. The size of the coherent structures in the direction transverse to the mean wind were estimated from the wind spectrums on a four-day case study. They ranged from 400 to 800 m for the cases classified as streaks, and from 1.3 to 2.0 km for the cases classified as rolls. These results pave the way for future long-term studies providing statistical insight on the frequency of occurrence of the different structure types, their physical properties, and their impact on pollutants’ concentrations.

Keywords: atmospheric turbulence, coherent structures, boundary layer meteorology, texture analysis, supervised machine learning, image classification

Abstract in French

Le but de ce projet de thèse était d'étudier les structures turbulentes cohérentes (rouleaux de convection et traînées) et plus généralement les fluctuations de taille moyenne à grande dans la couche limite atmosphérique en utilisant les observations recueillies par un unique lidar Doppler (lidar vent) pendant une campagne de deux mois à Paris (France). Une méthode innovante a été développée pour classer de façon automatisée les motifs visibles sur les balayages quasi-horizontaux du lidar, basée sur des paramètres de texture et des algorithmes d'apprentissage supervisé. Un ensemble d'entrainement de 150 cas a été construit en utilisant des informations auxiliaires (images satellite, observations météorologiques) pour vérifier la classification établie manuellement entre quatre types : rouleaux, thermiques, trainées et « autres ». La performance du processus de classification a été évaluée par la méthode de validation croisée à 10 blocs. Une erreur très satisfaisante de 9% a été obtenue avec un algorithme d'analyse discriminante quadratique et en utilisant seulement 5 paramètres de texture comme classificateurs. Ce processus a ensuite été appliqué à l'ensemble des données (4577 balayages lidar). Dans la classification obtenue, les type de structures suivent un cycle diurne plausible et sont associées avec les paramètres météorologiques d'une façon logique au regard des considérations théoriques. La taille des structures cohérentes dans la direction perpendiculaire au vent dominant a été estimée en utilisant les spectres du vent pour un cas d'étude de quatre jours. Elle allait de 400 à 800 m pour les cas classés comme traînées, et de 1,3 à 2,0 km pour cas classés comme rouleaux. Ces résultats ouvrent la voie à de futures études de long terme qui fourniront une vision statistique de la fréquence d'occurrence des différents types de structures, de leurs propriétés et de leur impact sur les concentrations de polluants.

Mots clés en français : turbulence atmosphérique, structure cohérentes, météorologie de la couche limite, analyse de texture, apprentissage supervisé, classification d'images

Dedication

I dedicate this manuscript to my parents Anastasios Cheliotis and Panagiota Karouni and to my sister Metaxia Chelioti. Their psychological and economical support throughout my studies have given me the courage to follow a scientific career. Thank you for all your help.

Acknowledgements

I would like to kindly thank Hervé Delbarre and Elsa Dieudonné of the Laboratoire de Physico-Chimie de l'Atmosphère, Université du Littoral Côte d'Opale (LPCA-ULCO) (Dunkerque, France) for giving me the opportunity to work on this topic. Since my bachelor degree year I was very interested in the Boundary Layer Meteorology field. I am grateful for collaborating with such high level scientists in a topic that has drawn the attention of the scientific community in prestigious conferences and peer-reviewed journals. Furthermore, I would like to thank them for their feedback throughout my Ph.D. studies that ensure the quality of this work. I appreciate the time and effort they put in order to address every issue I had regarding studying and living in France.

I would also like to thank Anton Sokolov of the LPCA-ULCO and Egor Dmitriev of the Marchuk Institute of Numerical Mathematics Russian Academy of Sciences (Moscow, Russia) for their valuable assistance in this work. Their contribution in the analysis of the results was decisive for the completion of this work. Moreover, they provide me with all the necessary information regarding a field that I am not familiar with (image classification), thus showing me how beneficial the image classification could be in atmospheric science.

I am very thankful to Hervé Delbarre, Patrick Augustin and Marc Fourmentin of the LPCA-ULCO for installing and monitoring the Doppler wind lidar in the field campaign and to François Ravetta, Jacques Pelon, Gilles Plattner and Amelie Klein of the LATMOS, Sorbonne University (Paris, France) for organizing and carrying out the VEGILOT campaign. The quality of the observations led to a smooth analysis of the data and eventually in very interesting results.

I would like to thank my colleagues and friends from LPCA-ULCO for helping me to adapt in a new environment in a new country, especially Eric Fertein and his family, Muriel Chaghouri, Khaoula Karroum, Lamia Moufarrej, Tinehinane Bounazef, Rebecca El Khawaja and Soulemane-Halif Ngagine. I should like to express my appreciation to my very good friends and colleagues Konstantinos Christakos of the University of Bergen (Bergen, Norway) and George Varlas of the Hellenic Centre for Marine Research (Anavyssos, Greece) for all the support throughout my Ph.D. studies and for their great ideas for collaborative studies that we have carried out all these years. Let me also express my gratitude to my good friends Aljoša Slameršak, Xabier Pedruzo-Bagazgoitia and Ilias Apostolakos for supporting me and encouraging me to pursue a Ph.D. degree.

I would like to warmly thank my parents Anastasios Cheliotis and Panagiota Karouni and my sister Metaxia Cheliotis. It would have been impossible to pursue a Ph.D. degree without their psychological and economical support.

This work is a contribution to the CPER research project IRenE 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).

I acknowledge the use of imagery provided by services from NASA's Global Imagery Browse Services (GIBS), part of NASA's Earth Observing System Data and Information System (EOSDIS).

The aerosol lidar belongs to the air quality observatory, QUALAIR (http://qualair.aero.jussieu.fr), located at the Université Pierre et Marie Curie, Paris France.

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).

Table of Contents

Preface

Boundary layer meteorology is a field that has always excited me. I find the understanding of weather phenomena and atmospheric processes in the part of the atmosphere that has direct impact on human life particularly interesting. For this reason, during my Bachelor degree in physics and Master degree in climate studies I followed several theoretical, practical and numerical weather simulation courses in Boundary Layer Meteorology. I wanted to deepen my knowledge and expertise in a related topic and this Ph.D. project offered me the opportunity to do so by working on the coherent structures in the atmospheric boundary layer. Furthermore, I had the privilege to gain a hands-on experience in large datasets recorded by a Doppler wind lidar. I extended my knowledge in programming as I developed several algorithms for the analysis of the results throughout my Ph.D. thesis.

I hope that this study can stimulate interest in the scientific community to perform long term studies for coherent structures and likewise understand their behavior to a greater extent. A long term study that correlates the structures with the pollution dispersion could be significantly beneficial for the scientific community as well as the general public.

Extensive Summary in French

Chapitre 1 – Physique de la couche limite & classification des images

Couche limite atmosphérique et turbulence

La troposphère est la partie la plus basse de l'atmosphère qui s'étend de la surface de la Terre à environ 10 km d'altitude aux latitudes moyennes. En plus d'être la partie de l'atmosphère où vivent les humains, c'est aussi le lieu où se produisent la grande majorité des événements météorologiques et où se développent les nuages. Seule une fraction de la troposphère, à savoir la couche limite atmosphérique (ABL en anglais[1]), peut être directement affectée par la surface de la terre, *via* des phénomènes tels que frottement, le transfert de chaleur, l'émission de polluants, l'évaporation et la transpiration des végétaux etc. L'ABL représente la couche la plus basse de l'atmosphère et varie en profondeur entre plusieurs dizaines de mètres et quelques kilomètres selon le lieu et le moment. La couche située au-dessus de l'ABL est connue sous le nom d'atmosphère libre. La principale distinction entre les deux est la variation diurne de la température de l'air qui est évidente dans l'ABL, mais pas dans l'atmosphère libre. Au fur et à mesure que le sol réagit au rayonnement solaire (ou à son absence) en devenant plus chaud ou plus froid, l'air près de la surface sera affecté par des processus de transport. Il est possible que ces processus influencent l'ensemble de la troposphère. Cependant, au-dessus de l'ABL, le changement sera lent en comparaison, ce qui permet de distinguer les deux parties de la troposphère.

La compréhension des processus atmosphériques au sein de l'ABL est vitale pour divers domaines de recherche. Que ce soit le transport et la dispersion des polluants atmosphériques ou leur dépôt sec à la surface de la Terre, le développement de nuages bas ou un fort cisaillement du vent (par exemple des jets de basses couches ou LLJ[2]) qui affectent les décollages et les atterrissages dans le secteur de l'aviation, tous ont lieu dans l'ABL. En outre, la précision des prévisions météorologiques et des prévisions climatiques dépend énormément de la représentation réaliste des processus atmosphériques de l'ABL dans les modèles numériques de prévision météorologique et dans les modèles climatiques respectivement. Par conséquent, il est important de simuler avec succès l'écoulement dans ABL afin de prédire le transport, le mélange et la dispersion, la transformation et l'interaction avec le sol et la végétation des variables météorologiques et atmosphériques au sein de l'ABL. À cet égard, les variables météorologiques et

[1] Dans ce résumé, tous les acronymes sont indiqués en anglais afin de permettre au lecteur de faire la correspondance avec le corps du manuscrit de thèse.

[2] LLJ: Low-Level Jet

atmosphériques les plus importantes sont le vent, la température de l'air, l'humidité de l'air et les composés atmosphériques inertes et réactifs.

L'ABL peut être divisée en sous-catégories en ce qui concerne le développement de son cycle diurne. Pendant la journée, le chauffage de surface est le moteur du développement de l'ABL. Les instabilités thermiques, générées par le chauffage de surface, créent des conditions convectives qui donnent le nom de couche limite convective (CBL) ou couche mixte (ML). L'ABL est séparée de l'atmosphère libre par la zone d'entraînement où les échanges entre l'ABL et l'atmosphère libre se produisent pendant la journée. D'autre part, le refroidissement radiatif pendant la nuit, crée une couche d'air stable près de la surface conduisant ainsi à une couche limite stable (SBL), également connue sous le nom de couche limite nocturne (NBL). Les périodes de passage de la SBL à la CBL et vice versa sont respectivement appelées transitions du matin et du soir. Une couche de stabilité neutre persiste au-dessus de la SBL pendant la nuit, qui n'est pas affectée par les propriétés liées à la surface. Celle-ci est connue sous le nom de couche résiduelle (RL).

Le comportement des polluants à l'intérieur de l'ABL varie selon les conditions de stabilité, donc diffère selon la sous-catégorie d'ABL considérée. Les instabilités intenses de la CBL peuvent disperser les polluants loin de la source et nettoyer l'atmosphère près de la surface. Cependant, dans les conditions stables de la SBL, les polluants peuvent être piégés près de la surface et éventuellement s'accumuler en grandes concentrations. Le transport à l'échelle régionale ou continentale des polluants dans l'atmosphère est dû au champ de vent synoptique, tandis que les mouvements turbulents provoquent leur dispersion dans toutes les directions et leur dépôt. L'évolution des concentrations des polluants est la résultante de leur transport ainsi que les taux d'émission des sources. Ces informations sont essentielles pour déterminer les zones propices à l'installation de sources potentielles de pollution ainsi que pour prévoir les éventuelles conséquences à long terme de leur utilisation. De cette manière, plusieurs modèles numériques de dispersion sont développés pour la prédiction de la distribution spatiale et temporelle des contaminants rejetés dans l'atmosphère.

Structures turbulentes cohérentes

L'influence du frottement et du chauffage par la surface, l'humidité, les polluants etc., tout cela est rapidement et efficacement transmis à l'ensemble de l'ABL par le mécanisme de transfert ou de mélange turbulent. La quantité de mouvement, la chaleur et les espèces chimiques peuvent également être transférés dans la direction inverse, vers le bas, par le même mécanisme. Bien que les écoulements turbulents soient

caractérisés par une imprévisibilité élevée, des structures cohérentes peuvent toujours se former dans une atmosphère turbulente. Les structures cohérentes se produisent lorsque les tourbillons sont organisés selon un schéma quasi-périodique dans les deux dimensions horizontales. Cette zone organisée s'étend sur une grande surface (c'est-à-dire qu'elle compte un grand nombre de motifs adjacents). En outre, une structure cohérente doit conserver sa forme pendant une période de temps suffisante pour les calculs de statistiques moyennées dans le temps. Les structures cohérentes dont traite ce travail sont les stries turbulentes et les rouleaux de convection.

Les stries turbulentes sont des bandes étroites, alignées avec la direction horizontale du vent, dans lesquelles alternent des zones de vents horizontal plus fort et plus faible. L'espacement entre les stries est généralement de l'ordre de quelques centaines de mètres ; les stries sont en général localisées dans la SBL mais elles atteignent parfois la couche mélangée. Leur cycle de vie est plutôt court : elles durent quelques dizaines de minutes avant de se régénérer. Le principal facteur de formation de stries est le développement d'un cisaillement de vent élevé entre la surface et le bas de la couche mélangée (ou la couche résiduelle le nuit). En effet, lorsque des tourbillons se forment, l'air ascendant provient d'altitudes où le vent est moins rapide (et inversement), ce qui crée au niveau du sommet de la SBL des zones de vent plus faible (ou plus fort) qui ont naturellement tendance à s'aligner. Cela se produit en particulier sous un jet de basses-couches, c'est-à-dire un maximum de vent se produisant dans les premières centaines de mètres de l'atmosphère.

Les rouleaux de convection sont également des motifs allongés alignés avec la direction horizontale du vent et ils se développent dans la couche mélangée convective. Les variations spatiales du flux thermique issu de la couche de surface, le cisaillement du vent à basse altitude et l'homogénéité de la surface favorisent le développement et le maintien des rouleaux. A l'opposé des stries, les rouleaux se forment lorsque l'instabilité thermique est forte, c'est-à-dire lorsque la flottabilité est le paramètre dominant pour la formation de tourbillons, par rapport au cisaillement du vent. La rotation des grands tourbillons entraîne des mouvements ascendants et descendants dans l'ABL. Lorsque des cumulus de beau temps se forment au sommet des zones d'ascendance des rouleaux, on observe des allées de nuages séparée de bandes de ciel clair qui permettent de détecter visuellement les rouleaux, notamment sur les images satellites. Cette étude se concentre sur la détection et l'identification des structures cohérentes dans l'ABL.

Des instruments de télédétection, tels que le lidar, peuvent être utilisés mesurer la vitesse du vent le long du faisceau laser (vent radial) puis reconstruire le profil des composantes horizontale et verticale du vent en combinant plusieurs directions de visée obliques. À partir d'un balayage horizontal, c'est-à-dire d'une mesure du vent radial dans

un disque autour de l’instrument, il est possible d'estimer la vitesse et la direction du vent horizontal moyen, ainsi que les fluctuations turbulentes du vent radial. En effet, si le champ de vent était homogène dans l’espace entourant l’instrument, le vent radial mesuré le long d’un « anneau » (à une distance fixée de l’instrument) serait une fonction sinusoïdale de la direction du faisceau. Les écarts à cette fonction que l’on observe correspondent aux fluctuations turbulentes du vent radial. Cette méthode est connue sous le nom de d'affichage azimutal de la vitesse (VAD). Cette méthode présente l’avantage de permettre de visualiser tout le champ des fluctuations et de reconnaître certains motifs ressemblant aux structures cohérentes.

Chapitre 2 – Données et méthodologie

Campagne et instrumentation

Afin d'étudier les structures cohérentes, les balayages horizontaux réalisés par un lidar Doppler (lidar vent) installé au sommet d'une tour de 75 m de haut au centre de Paris (campus Jussieu) pendant une campagne de deux mois (04/09/2014-06/11/2014) ont été utilisées. Des profils verticaux du vent, des aérosols et de l'ozone ont également été mesurés durant cette campagne. Les instruments utilisés étaient le lidar Doppler Leosphere WLS100, le lidar aérosol CAML[3] de la société CIMEL et la version modifiée du Lidar aéroporté pour l'ozone troposphérique (ALTO). Pour cette étude, les observations enregistrées par le lidar ozone n'ont pas été utilisées.

Paris est un endroit intéressant pour les observations, car bien qu'il s'agisse d'une mégalopole de presque 12 millions d'habitants, la hauteur des bâtiments en centre-ville ne dépasse pas 50 m, à quelques exceptions près, en raison de la réglementation locale. Par conséquent, un faisceau émis depuis un bâtiment de 75 m de hauteur, même pour des balayages quasi horizontaux, n'est pas interrompu par d'autres bâtiments. La portée maximale théorique du lidar vent était de 5 km pour toutes les méthodes de balayage avec une zone aveugle de 100 m. La zone couverte par les balayages horizontaux était entièrement urbanisée, avec peu d'espaces verts et une portion de la Seine. La résolution spatiale était de 50 m, il était donc possible d'étudier des fluctuations turbulentes associées à des structure cohérentes de taille moyenne à large (mlf-cs[4]), mais pas de petites structures turbulentes. Les balayages près de la surface étaient quasi horizontaux, mais l'angle d'élévation de 1° entraînait malgré tout une différence de hauteur de 87 m entre la position du lidar au centre et l'altitude du faisceau au niveau les bords de la zone de balayage. Cependant, la portée effective était plutôt de l'ordre de 3 à 4 km (celle-ci est limitée par le niveau de signal) et l'altitude du faisceau par rapport au sol ne variait pas de plus de 20 à 30 m pour les 3 premiers km autour du lidar.

Quatre types de balayage ont été réalisées par le lidar Doppler au cours de la campagne. Le profil vertical du vent était mesuré par la méthode DBS[5] (une combinaison de 4 faisceaux inclinés + 1 vertical) ; le profil de la variance du vent vertical était estimée

[3] CAML : Cloud and Aerosol Micro-Lidar

[4] mlf-cs : medium to large fluctuations coherent structures

[5] DBS : Doppler Beam Swinging

grâce à des séries de tirs verticaux (LOS[6] au zénith) ; les balayages verticaux ou RHI[7] permettaient notamment de visualiser le sommet de la NBL ; enfin, les balayages horizontaux ou PPI[8] ont servi à observer les structures turbulentes.. Il y a eu plus de 10000 balayages au total pendant la campagne, dont 4577 étaient horizontaux. Les observations du lidar aérosols, quant à elles, ont été utilisées pour estimer l'épaisseur de l'ABL au cours de la journée. Ce lidar était installé sur le toit d'un bâtiment adjacent à la tour sur laquelle le lidar vent était installé, mais à seulement 20 m d'altitude. L'instrument a fonctionné en continu pendant la période 04/09/2014-06/11/2014 et a fourni des observations moyennées sur des périodes de 10 minutes.

Type de motifs observés sur les balayages

Une fois la méthode VAD appliquée à l'ensemble des 4577 balayages horizontaux, l'inspection visuelle des mlf-cs a montré que certains d'entre eux contenaient des motifs allongés bien définis avec des bandes de vent radial turbulent alternativement positives et négatives. Il était également à noter que pendant la nuit, ces motifs allongés étaient étroits (quelques centaines de mètres de largeur) alors que pendant la journée, ils étaient plus larges (de l'ordre de 1 à 2 km de largeur). D'autres mlf-cs enregistrés au cours de la journée présentaient de grands zones fermés de valeurs positives et négatives (de 1 à 2 km de diamètre).

Les motifs allongés étroits observés durant la nuit étaient alignés avec la direction du vent moyen et ressemblaient aux stries décrites dans la littérature. Par ailleurs, les profils verticaux de la vitesse horizontale du vent montraient de dans de nombreux cas, ces motifs allongés nocturnes se produisaient dans des conditions de fort cisaillement du vent près de la surface, ce qui confirmait qu'il s'agissait bien de stries (cette catégorie sera donc nommé « stries » dans la classification). Les motifs allongés plus larges observés pendant la journée étaient également alignés avec la direction moyenne du vent et ressemblaient à des rouleaux de convection. Pour confirmer la présence de rouleaux dans la couche mélangée, les images fournies par les instruments satellites MODIS[9] ont été utilisées pour détecter les allées de nuages. Les satellites Terra et Aqua à bord desquels volent les instruments MODIS ne passent au-dessus de Paris que vers 1100 et 1300 UTC, mais de larges motifs allongés ont été observés à d'autres moments de la journée ; ceux-ci ont été attribués à des rouleaux (catégorie « rouleaux » dans la classification) à

[6] LOS : Line Of Sight

[7] RHI : Range-Height Indicator

[8] PPI : Plan Position Indicator

[9] MODIS : Moderate resolution Imaging Spectroradiometer

condition qu'ils appartiennent à une série de balayages consécutifs montrant tous la présence de rouleaux, dont l'un au moins coïncide avec une image MODIS montrant des allées de nuages.

Les grands motifs fermés qui peuvent être observés pendant la journée correspondent à des cellules de convection thermique non alignées en rouleaux et disposées de façon aléatoire (catégorie « thermiques » dans la classification). Pour le confirmer, le nombre d'heures d'ensoleillement quotidien et les valeurs horaires du rayonnement solaire enregistrés par la station Météo-France de Paris-Montsouris ont été utilisées. Les grands motifs fermés se produisaient les jours où les valeurs de rayonnement solaire étaient supérieures aux moyennes mensuelles et avec un ensoleillement persistant. Les images satellitaires des instruments MODIS (lorsqu'elles sont disponibles) ont confirmé que ce type de motif coïncidait souvent avec la présence de cumuli de beau temps non disposés en allées de nuages.

Enfin, de nombreux balayages ne présentaient aucun des trois types de motifs définis précédemment. Il s'agissait soit de motifs chaotiques dans les champs mlf-cs, à une échelle plus petite que les modèles thermiques, soit le plus souvent de cas où la méthode VAD n'avait pas été appliquée avec succès en raison d'un champ de vent radial non symétrique. Pendant les 2 mois de la campagne, les cas de ce type se sont produits principalement par temps calme (vent moyen < 2 $m \cdot s^{-1}$) ou par vent faible (2-4 $m \cdot s^{-1}$), lorsque le vent n'a pas une direction bien définie. Ce type de motif a été regroupé dans la catégorie « autres » de la classification.

Classification automatisée des types de motifs

Malgré la similarité des motifs observés, la classification du type de motif dans un si grand ensemble de données avec des milliers d'observations est une tâche fastidieuse, voire impossible. À cette fin, une nouvelle méthode automatisée a été développée pour l'identification des quatre types de motifs mentionnés ci-dessus. La méthodologie est basée sur les paramètres d'analyse de texture des fluctuations du champ de vent radial. Chaque champ de vent radial est associé à un ensemble de matrices de cooccurrences (CM) calculées pour les différentes dispositions de paires de points de grille en termes de distance et d'orientation. Pour chaque CM, plusieurs paramètres d'analyse de texture peuvent être estimés ; ici c'est l'homogénéité, le contraste, la corrélation et l'énergie de la CM qui ont été utilisés. Ensuite, ces paramètres peuvent être fournis à un algorithme d'apprentissage automatique qui va déterminer lesquels sont les plus significatifs pour la classification des types de motifs. L'algorithme peut être utilisé de façon non supervisée, c'est-à-dire directement pour l'ensemble des balayages, auquel cas il déterminera les

différentes classes de motifs sans aucune implication du chercheur. Il est aussi possible de sélectionner un sous-ensemble spécifique avec des cas préalablement classés par le chercheur (ensemble d'entraînement) et de le fournir à l'algorithme qui cherchera à reproduire la classification fournie. Cette dernière technique, dite apprentissage supervisé, a été choisie pour cette étude car chaque type de motif est associé à des caractéristiques physiques spécifiques.

Construction de l'ensemble d'entraînement

Dans cette étude, l'ensemble d'entraînement qui a été construit se composait de 30 cas de stries, 30 cas de rouleaux, 30 cas de thermiques et 60 cas appartenant à la catégorie «autres» sur un total de 4577 balayages horizontaux. Les 30 cas de stries sélectionnés correspondent à des motifs étroits et allongés (avec une dimension transverse de quelques centaines de mètres) observés de façon concomitante avec un fort cisaillement du vent près de la surface (sous un jet). Afin de s'assurer que les motifs étroits et allongés étaient principalement dus au fort cisaillement du vent et non aux thermiques, les 30 cas de stries sélectionnés ont été choisis pendant la nuit, en particulier entre 1800 UTC et 2200 UTC. Après 2200 UTC, le vent était généralement plus faible, conduisant à un champ de vent radial non symétrique sans structures cohérentes et pour lequel la méthode VAD ne fonctionnait pas ou mal. La moitié des 30 cas sélectionnés proviennent des 15 premiers jours de la campagne, les cas restants étant répartis sur le reste des deux mois.

En ce qui concerne les rouleaux de convection, seule la présence de allées de nuages peut confirmer leur formation, aussi les 30 cas sélectionnés étaient soit concomitants avec des allées de nuages visibles sur les images satellites de MODIS Aqua ou Terra, soit ils appartenaient à des séries continues de balayages consécutifs présentant des motifs similaires (motifs allongés de largeur supérieures à 1 km) dont l'un balayages coïncidait avec l'observation d'allées de nuages. Pendant la campagne, il y a eu 8 jours pour lesquels les images satellite montrent des allées de nuages ; ainsi sur les 30 balayages de type « rouleaux » inclus dans l'ensemble d'entraînement, 6 se sont produit début septembre, 9 mi-septembre et 15 la mi-octobre. Tous ces balayages ont été enregistrés entre 1030 UTC et 1400 UTC.

Les 30 balayages de type « thermiques » de l'ensemble d'entrainement proviennent de 23 jours différents répartis sur toute la campagne entre 900 UTC et 1500 UTC, avec un cas vers 1700 UTC. La sélection était basée sur la présence de grands motifs fermés (zones homogènes dans le champ mlf-cs d'environ 1 km ou plus) observés pendant des périodes avec des valeurs de rayonnement solaire horaire beaucoup plus élevées que la moyenne mensuelle affichée par le système d'information géographique photovoltaïque.

Enfin, le type « autres » était représenté par 60 cas, soit deux fois plus que chacun des autres types. La raison de ce choix est la sensibilité des algorithmes à l'ensemble d'apprentissage. Si l'ensemble d'entraînement n'est pas équilibré en fonction du résultat attendu de la classification, cela peut résulter sur de mauvaises performances. On s'attendait à ce que le jeu de données complet contienne une majorité significative du type « autres » en raison des nombreux épisodes de vents faibles observés au cours de la campagne. Par conséquent, il a été décidé d'inclure deux fois plus de cas du type « autres ». Afin de représenter les différents sous-type du groupe « autres », les champs de vent radial étaient non symétriques pour 53 des 60 cas tandis que pour les 7 derniers, les champs de vent radiaux étaient symétriques, mais il n'y avait pas de motif défini. Parmi les 53 cas avec un champ de vent radial non symétrique, 46 se sont produits dans des conditions de vent calme ou faible : en revanche, les 7 cas sans motif défini se sont tous produits dans des conditions de vent modéré (6-8 $m \cdot s^{-1}$). Les balayages du type « autres » sont répartis sur les deux mois de la campagne, la majorité étant nocturne, avec cependant 10 balayages enregistrés entre 0800 UTC et 1600 UTC.

Calcul des matrices de cohérence et des paramètres de texture

Avant d'estimer les CM et par la suite les paramètres d'analyse de texture, il était nécessaire de faire un prétraitement des champs de fluctuation du vent radial. Pour commencer, une rotation a été appliqué à chaque champs mlf-cs pour que le vent moyen semble toujours provenir du nord. En effet, on souhaitait que les structures soient classées indépendamment de leur orientation (alignée avec le vent moyen pour les stries et rouleaux). Ensuite, comme les points de mesure du vent radial sont localisés en coordonnées polaires (direction du faisceau laser et distance le long du faisceau), il a fallu interpoler les observations sur une grille régulière en coordonnées cartésiennes, ce qui est nécessaire afin de trouver les paires de pixels voisins. Pour terminer, il fallait répartir les valeurs de vitesse radiale du vent dans un nombre réduit de classes de vitesse. La priorité était de séparer les valeurs positives et négatives de la vitesse radiale du vent, car elles correspondent à différentes directions des fluctuations du vent, donc différentes zones dans la structure turbulente. Pour ce faire, les valeurs des fluctuations du vent ont été divisées en 8 classes : la première classe contenait toutes les valeurs supérieures à +0,5 $m \cdot s^{-1}$, les six classes suivantes étaient équitablement réparties entre +0,5 $m \cdot s^{-1}$ et −0,5 $m \cdot s^{-1}$ et la dernière classe contenait toutes les valeurs inférieures à −0,5 $m \cdot s^{-1}$. Ce prétraitement a amélioré les motifs et les a rendus plus visibles même à l'œil nu.

Une fois le prétraitement effectué, les matrices de cohérence ont été calculées pour chaque balayage, pour les paires de pixels du premier au trentième voisin (ordre n de 1 à 30 soit 50 à 1500 m de distance réelle) et pour l'ensemble des orientations possibles des

paires (azimut φ de −90° soit une paire nord→sud, à +90° soit une paire sud→nord). Ensuite, afin de mettre en évidence les différences entre les CM calculées pour les différents types de motifs, les quatre paramètres d'analyse de texture ont été calculés à partir de chaque CM de l'ensemble. Par conséquent, il était possible de tracer la variation des paramètres d'analyse de texture en fonction de l'azimut φ de la paire, pour un ordre n donné. Les courbes des cas de rouleaux, et surtout des cas de stries, montrent un pic proéminent à 0° pour certaines valeurs de n, par comparaison aux courbes plus lisses des cas de thermiques et des cas du type « autres ». En effet, l'orientation 0° correspond aux paires de pixels orientées ouest→est (perpendiculaires au vent moyen après rotation des champs) pour lesquelles des valeurs similaires de la fluctuation du vent radial peuvent être observées lorsque n correspond à une période des stries ou des rouleaux. Au final, cela signifie que l'amplitude de la courbe des paramètres de texture (différence maximum-minimum) doit varier selon le type de structures observé. Un motif parfaitement aligné avec le vent moyen doit également être symétrique autour du 0°, alors que pour les motifs aléatoires comme ceux des types thermiques et « autres », cela n'est pas nécessairement le cas. Le paramètre de symétrie ainsi que la somme des points (intégrale de la courbe) ont donc également été calculés, en plus de l'amplitude. Le calcul de ces « paramètres de courbe » ne fait pas partie du processus standard de classification des images décrit dans la littérature. Cette étape a été ajoutée afin de réduire le nombre de paramètres fournis à l'algorithme de classification automatisée. En effet, un trop grand nombre de paramètres peut créer un problème de multi-dimensionnalité dans l'algorithme, ce qui diminue ses performances.

Pour l'ensemble d'apprentissage, les trois paramètres de courbe pour les quatre paramètres d'analyse de texture ont été calculés, et ce pour les voisins d'ordre 1 à 30, ce qui porte le nombre de classificateurs à 360. Trois paramètres physiques ont également été inclus : la vitesse moyenne du vent et l'erreur quadratique moyenne (RMSE) obtenue lors de la régression de la méthode VAD, ainsi que l'heure UTC (qui est proche de l'heure solaire à Paris). Les 363 classificateurs calculés pour chaque balayage de l'ensemble d'entrainement ont été fournis à l'algorithme d'analyse discriminante quadratique (QDA) suivant un algorithme glouton de sélection progressive. L'erreur de classification a été estimée sur la base de la validation croisée en 10 blocs : l'algorithme est entraîné en utilisant seulement 90% de l'ensemble d'apprentissage, tandis que les 10% restants (15 cas) sont utilisés pour tester le résultat de la classification en le comparant au résultat attendu. Ce processus est répété 10 fois afin que les 150 cas de l'ensemble d'entraînement soient tous utilisés tour à tour pour validation. Dans le même processus, l'algorithme sélectionne les classificateurs qui minimisent l'erreur (nombre de cas incorrectement classifiés).

Chapitre 3 - Résultats

Classification de l’ensemble d’entraînement et performance de l’algorithme de tri

L'erreur obtenue était légèrement supérieure à 9% avec les cinq classificateurs sélectionnés par la méthode QDA car ils minimisent l'erreur. Ces classificateurs correspondent à des distances variées dans l’espace réel – 2ème voisin (100 m), 8ème voisin (400 m) 18ème voisin (900 m) etc. – ceci est particulièrement important pour distinguer les grands motifs des petits. Il est également confirmé que les amplitudes des courbes d'homogénéité et de contraste constituent des classificateurs significatifs pour les structures, ce qui signifie que les pics proéminents des courbes différencient effectivement les types de motifs. Les intégrales des courbes de contraste et de corrélation ont également été sélectionnées par l'algorithme. En effet, le type « autres » se caractérise par des motifs chaotiques tandis que le type « thermiques » se caractérise par de grandes zones homogènes, ce qui se traduit par des différences dans les intégrales des courbes pour des paramètres tels que le contraste et la corrélation, en particulier pour les distances moyennes (8ème voisin / 400 m) à grandes (18ème voisin / 900 m). Il est à noter que la symétrie de la courbe d'homogénéité a été sélectionnée par l'algorithme. Cela confirme l'importance d'aligner tous les champs mlf-cs dans la même direction. L'un des résultats les plus frappants, cependant, est qu'aucun des trois paramètres physiques n'a été sélectionné par l'algorithme, malgré le fait que l'heure de la journée pourrait permettre de faire la distinction entre les structures nocturnes (stries) et diurnes (thermiques, rouleaux).

L'algorithme QDA a donné les meilleurs résultats pour la classification des structures de type « stries » : il a classé correctement 29 des 30 cas de stries inclus dans l'ensemble d'entraînement, le dernier étant incorrectement classé parmi les rouleaux. La deuxième meilleure performance concerne la catégorie « autres », pour laquelle 58 des 60 cas ont été classés correctement, les deux derniers étant classés à tort parmi les thermiques. Les performances de l'algorithme pour les cas de rouleaux étaient également bonnes, avec 27 cas correctement classées et 3 cas classées à tort comme « thermiques ». La performance là moins satisfaisante concerne les thermiques, même si 24 cas ont été correctement classés. Parmi les 6 cas incorrectement classés, 4 cas ont été à tort classés comme des rouleaux et 2 cas comme du type « autres ». Il est important de noter que différentes options de l’algorithme de tri, d'autres paramètres d'analyse de texture ou un ensemble d'apprentissage plus large pourraient encore améliorer les résultats de la classification. Un algorithme de plus haut niveau qui révèlerait la meilleure combinaison de paramètres minimisant l'erreur serait particulièrement utile.

Classification des deux mois de données, conditions d'occurrence des structures

Les cinq classificateurs sélectionnés par l'algorithme lors du processus d'entraînement ont été utilisés pour classifier l'ensemble des 4577 balayages horizontaux enregistrés pendant la campagne. L'algorithme a classé 46% des balayages dans la catégorie « autres », 25% comme stries, 9% comme rouleaux (les structures cohérentes totalisent ainsi 34%) et 20% comme thermiques. Bien que les classificateurs sélectionnés n'incluent pas l'heure de la journée, le cycle diurnes moyen des structures observées au cours de la période de deux mois concorde avec les attendus théoriques. En particulier, la grande majorité des cas de rouleaux (87%) et de thermiques (77%) ont été trouvés pendant la journée, entre 0800 UTC et 1800 UTC comme attendu puisque ces structures se développent dans une atmosphère convective. A l'opposé, la plupart des cas de stries (66%) et des cas de type « autres » (79%) ont été trouvés pendant la nuit, de 1800 à 800 UTC. Cela était attendu pour les stries car des jets nocturnes se sont produits pendant 20 des 63 nuits de la campagne, de sorte que des conditions étaient fréquemment favorables à la formation de structures de type stries. Concernant les cas de type « autres », 75% d'entre eux sont associés à des valeurs moyennes de vent inférieures à 4 $m \cdot s^{-1}$, ce qui correspond au fait que le type « autre » correspond le plus souvent à des champs de vent asymétriques observés par vent faible. En revanche, les cas classés comme stries et comme rouleaux étaient principalement associés à des valeurs de vent moyen modérées, entre 4 et 10 $m \cdot s^{-1}$, comme c'est le cas dans la littérature.

La méthodologie de classification automatique a rendu l'analyse statistique des conditions d'occurrence des différents types de structures pendant les deux mois rapide, facile et fluide. En particulier, il était possible de sélectionner uniquement les dates auxquelles l'épaisseur de l'ABL pouvait être estimée et d'associer ce paramètre aux différents types de structures. À noter que la précision sur l'épaisseur de l'ABL n'était pas essentielle pour les besoins de cette étude, une valeur approximative était suffisante. La hauteur de l'ABL a été estimée en utilisant les faisceaux LOS verticaux des observations du lidar Doppler pendant la nuit et les observations du lidar aérosol pendant la journée. Pendant la nuit, la méthodologie repose sur l'écart type du vent vertical (σ_w) calculé à partir des multiples tirs LOS enregistrés pendant des périodes de 30 secondes. La valeur de σ_w est plus élevée dans la couche limite du fait de la turbulence, passe par un maximum au sommet de la couche limite, dans la zone d'entraînement, puis décroit au passage dans la couche résiduelle. La hauteur ABL a donc été estimée grâce au maximum de σ_w. Cette méthode fonctionne pendant la nuit, car les aérosols dans la couche résiduelle fournissent un niveau de signal suffisant pour que le lidar Doppler mesure le vent au-dessus de l'ABL. Pendant la journée, il n'y a la plupart du temps aucune

mesure de vent au-dessus de l'ABL car la troposphère libre est trop propre. Par conséquent, le profil de variance du vent vertical est incomplet et ne peut pas être utilisé pour estimer la hauteur de l'ABL.

La survenue de LLJ nocturnes peut également compliquer la détermination de l'épaisseur d'ABL. En effet, on définit alors le sommet de l'ABL comme la frontière inférieure du cœur de jet (altitude du maximum de vent). Cependant, le fort cisaillement du vent peut générer des turbulences et éventuellement augmenter la hauteur de l'ABL. Pour l'étude actuelle, la hauteur ABL n'a été estimée que via les valeurs d'écart-type du vent vertical pendant la nuit, même lorsque des jets se sont produits, car la précision dans l'estimation de la hauteur ABL n'était pas une priorité pour cette étude.

De jour, ce sont les observations de lidar d'aérosol qui ont été utilisées pour déterminer l'épaisseur de la couche limite. La concentration en aérosols étant nettement plus élevée dans l'ABL, le signal lidar corrigé de l'atténuation géométrique (Pr^2) diminue fortement au passage de l'ABL vers la troposphère libre. Le minimum de la dérivée première du Pr^2 a ainsi été utilisé pour l'estimation de la hauteur d'ABL. À l'inverse du lidar Doppler, les observations du lidar aérosols ne peuvent pas être utilisées pour l'estimation de la hauteur d'ABL pendant la nuit, ceci en raison de la zone aveugle du lidar CAML, qui est de l'ordre de 200 m. Par ailleurs, pour les deux types de lidar, l'estimation de la hauteur ABL n'est possible qu'en ciel clair. En effet, les nuages bas, le brouillard, les précipitations ou tout autre phénomène météorologique qui absorbe la lumière du laser « éteint » le signal des lidars rend impossible l'estimation de la hauteur d'ABL.

La hauteur ABL a pu être estimée avec précision pour la majorité de la période de 24 heures pendant 18 des 64 jours de la campagne. Ces 18 jours ont été sélectionnés afin d'examiner la relation entre la hauteur ABL et les types de structures observés. 94% des balayages classés comme « rouleaux » et 80% de ceux classés comme « thermiques » se sont produits dans une ABL en cours de développement (de 350 à 1000 m d'épaisseur) ou déjà bien développée (1000 à 2000 m d'épaisseur). Au contraire, 90% des balayages classés comme « stries » et 86% de ceux classés comme « autres » ont été associés à une ABL fine (moins de 350 m d'épaisseur).

Étude de cas, largeur des structures allongées

Une étude de cas de 4 jours, avec formation d'un LLJ tous les soirs, a montré que les structures de type stries apparaissent en même temps que le LLJ. Au cours des matinées, au fur et à mesure que le LLJ disparait et que les instabilités thermiques se développent, les stries sont remplacées par des structures de type thermiques ou rouleaux. La différence de taille horizontale entre stries et rouleaux a été confirmée par

l'estimation de la période spatiale dans la direction perpendiculaire au vent moyen (direction transverse). Cette grandeur est estimée grâce au spectre de la vitesse radiale du vent dans la direction transverse, calculée par Transformée de Fourier Rapide (FFT). Plusieurs études ont montré que ce spectre présente des pics distincts correspondant aux périodicités des structures. Afin d'accroitre le domaine spatial couvert, donc les longueurs d'ondes accessibles, les deux faisceaux opposés du balayage quasi-horizontal et situé dans la direction transverse sont combinés avant de calculer la FFT. Ces deux faisceaux sont enregistrés avec 90 secondes d'écart, il faut donc supposer que les structures gardent les mêmes propriétés au fur et à mesure de leur défilement au-dessus du lidar, c'est-à-dire qu'elles sont homogènes dans la direction du vent moyen (direction longitudinale) sur une distance de l'ordre de 200 à 700 m (90 s multiplié par la vitesse moyenne du vent, qui est d'environ 2 à 8 $m \cdot s^{-1}$). Cette distance est du même ordre de grandeur que les périodicités visibles dans la direction transverse, il est donc raisonnable de supposer que les structures allongées ne varient pas sur une telle distance dans la longitudinale.

L'utilisation d'une seule paire de faisceaux produit des spectres très bruités, de sorte que les spectres des cinq paires de faisceaux les plus proches de la direction transverse ont été calculés, puis moyennés. Avant d'appliquer la FFT, le champ de vent radial a été converti des coordonnées polaires en coordonnées cartésiennes et interpolé sur une grille régulière afin que tous les points de la grille aient un espacement égal à 50 m dans la direction transverse. Les périodicités observées (comprenant une bande positive et une bande négative dans le mlf-cs) varient entre 400 et 800 m pour les balayages classés comme stries, et entre 1,3 et 2 km pour les balayages classés comme rouleaux. À noter que les spectres étaient limités à 300 m du côté des courtes longueurs d'onde courtes. En effet, même si la résolution du lidar (50 m) devrait permettre de descendre jusqu'à 100 m, la partie du spectre correspondant aux longueurs d'onde comprises entre 100 et 300 m est trop bruitée pour pouvoir y distinguer les pics associés aux stries.

Il est compliqué de comparer les valeurs de périodicités obtenues ici avec la littérature. En effet, dans le cas des rouleaux, on a observé leur taille au niveau de la base de la structure, dans le bas de la couche de mélange. En revanche, les autres études, notamment celles basées sur des données aéroportées, ont observé les rouleaux au milieu de la couche mélangée. Malgré cela, les spectres ont clairement montré des tailles différentes pour les rouleaux et les stries. Par conséquent, les tailles horizontales pourraient être un paramètre pour valider la classification. Les tailles horizontales ont été estimées uniquement pour les structures cohérentes car les modèles aléatoires des thermiques ne produisent pas de pics clairs dans les spectres.

Conclusions générales

La classification automatisée des mlf-cs fournit un moyen cohérent d'identifier et de classer les structures. Il peut être implémenté dans n'importe quel autre ensemble de données indépendamment de la date, de la durée et du lieu tant que les modèles des champs mlf-cs sont similaires. Les résultats indiquent une bonne concordance entre les structures classées et les paramètres physiques. Les résultats pourraient encore être améliorés en optimisant la sélection des paramètres d'analyse de texture et/ou de l'algorithme d'apprentissage automatique supervisé. Néanmoins, cela crée des possibilités passionnantes pour de nouvelles études.

Jusqu'à présent, les observations lidar du vent n'ont pas été utilisées pour une analyse statistique des propriétés physiques des structures en raison des limites de l'identification visuelle des structures dans les grands ensembles de données. La classification automatisée peut aider de telles études. De plus, elle peut être combiné avec des modèles LES[10] afin d'examiner l'impact des structures sur les concentrations de polluants. Enfin, comme le lidar est capable d'estimer le profil des paramètres turbulents (flux de moment et énergie cinétique turbulente) en utilisant une configuration de balayage appropriée, il serait possible d'alterner entre des balayages horizontaux, donc des champs mlf-cs, et des mesures des paramètres turbulents. De cette manière, la possibilité d'observer des sous-catégories plus fines de structures pourrait être examinée.

[10] LES : Large Eddy Simulation (simulation à grand tourbillons)

Introduction

The atmospheric boundary layer is the part of the atmosphere containing the turbulent motions, which are particularly interesting in meteorology for their chaotic behaviour and their unpredictability. Nevertheless, in the turbulent atmosphere, it is still possible to observe some coherent structures. These coherent structures are characterized by patterns of alternating low and high wind speed areas. They are important structures as they can affect pollutants' concentrations, aircraft landing and takeoff or wind turbine operations. It is possible to identify such structures in radial wind observations provided by Doppler lidar or radar scans. However, lidar or radar studies so far are limited to particular cases for short periods of time, as the analysis of large wind lidar datasets with regards to the coherent structures can be a cumbersome process. Furthermore, the visual identification of the structures is not consistent as it depends on the researcher, thus making the participation of multiple researchers necessary in order to ensure the validity of the identification.

The limitations caused by the visual identification can be overcome by an algorithm that automatizes the classification of the structures based on their patterns. For the current study, an innovative method was developed to identify and classify the patterns based on a single Doppler wind lidar (Windcube WLS100). The methodology relied on the image classification principle and it was based exclusively on texture analysis parameters. A supervised machine learning algorithm revealed the texture analysis parameters that could classify with higher accuracy a training ensemble of coherent structures as well as other types of patterns. Subsequently, a large dataset of scans recorded in Paris for a two-month campaign was classified based on these texture analysis parameters. By analyzing the physical properties of the classified structures it was evident that the different patterns were characterized by similar physical properties.

The successful development of an automated classification of coherent structures facilitates the analysis of large datasets while providing consistency. The automated classification is not dependent on the subjective judgement of the individual researcher. Moreover, it is not dependent on the location nor on the period of the year. As long as the patterns in the radial wind fields are similar, the methodology is applicable. This research can pave the path for unique studies that are currently hard or impossible to perform. For instance, a statistical analysis of the physical characteristics of the structures can be carried out. Furthermore, a study examining the correlation between the structures and pollutants' concentrations for long periods is feasible. Any study that requires large datasets and is so far rare or impossible to perform can be achieved.

In the current manuscript, a step by step presentation of the automated classification methodology development for medium-to-large fluctuations and coherent structures is demonstrated. Additionally, the state of the art regarding wind lidar observations with respect to coherent structures is presented along with all the relevant information regarding texture analysis and supervised machine learning as means to classify images. The performance of the classification in the two-month dataset is showcased and the results are explained and discussed. More analytically, the manuscript is divided in four chapters as follows:

Chapter 1 contains all the theoretical background information related to this study. Atmospheric boundary layer, coherent structures, wind lidar observations and the different scanning techniques are discussed along with illustrations in order to provide a better understanding for each term. Additionally, all the essential information regarding image classification are included in Chapter 1. More particularly the different steps for the image classification process are presented via an ideal case. The texture analysis parameters and supervised machine learning techniques that are relevant to this study are described as well.

In Chapter 2 the methodology as well as the means to obtain the observations are displayed in full details. First by introducing the experimental set-up and the area under observation and consequently, by describing the implemented methods for the measurement instruments, namely wind and aerosol lidar, along with their ability to record parameters useful for this study. The aim of this study was to identify and classify coherent structures and therefore an analytical description of how to estimate medium-to-large fluctuations and coherent structures fields based on radial wind observations is featured in this chapter. Then the methodology is introduced with arguments for each step and choices that were made. Chapter 2 also includes the description of the estimation for two important physical parameters in this study, namely the atmospheric boundary layer height and the horizontal size of the structures.

The results are showcased in Chapter 3 in the form of journal articles. The first paper presents the automated classification methodology and its' results for a training ensemble of 150 cases. This manuscript has been published in the peer-reviewed journal *Atmospheric Measurement Techniques* in December 2020. The second article presents the implementation of the classification in the whole ensemble (4577 cases) of the two-month campaign in Paris. The physical properties of the structures are displayed as well. This paper is pending submission in a peer-reviewed journal specialized in atmospheric physics. An introduction precedes each paper to provide insights about the particular study. Additionally, a discussion part follows each paper in order to pinpoint the strongest

remarks and indicate possible suggestions for future studies or improvements of the results.

Finally, all the major points of the study are summarized in the conclusions including arguments regarding the improvement of the methodology and possible applications for future studies.

1. Boundary Layer Physics & Image Classification

1.1 Fundamental concepts

The troposphere is the lowest part of the atmosphere extending from the Earths' surface to approximately 10 km in altitude in mid-latitudes, 9 km in high latitudes and 16 km in the tropics on average. Aside from being the part of the atmosphere where humans are living, it is also the home for the vast majority of the weather events and clouds' development (Shonk, 2013). Notwithstanding only a fraction of the troposphere, namely the atmospheric boundary layer (ABL), can be affected directly by the earths' surface, with incidents such as friction, heat transfer, pollutants' emission, evaporation and transpiration etc. The ABL occupies the lowest layer of the atmosphere and ranges in depth from dozens of meters to few kilometers depending on time and space. The layer above the ABL is known as the free atmosphere. The main distinction between the two parts is the diurnal variation of air temperature which is evident in the ABL but not in the free atmosphere. As the grounds responds to the solar radiation by becoming warmer or cooler, the air near the surface will be affected via the transport processes. It is possible that the surface characteristics will influence the entire troposphere. However, above the ABL, the change will be slow in comparison and thus the two parts of the troposphere can be distinguished (Stull, 1988). The current study concentrates exclusively on the atmospheric processes within the ABL.

1.1.1 The atmospheric boundary layer

As stated in Section 1.1, the ABL is defined as the part of the atmosphere that is affected by the earth's surface. The understanding of the atmospheric processes within the ABL is vital for various research fields. The transport and dispersion of atmospheric pollutants or their dry deposition on the Earths' surface, the development of low clouds or severe wind shear (e.g. low-level jets) that affect the take-offs and the landings in the aviation sector, all take place in the ABL. Furthermore, the accuracy of the weather forecasts and climate predictions depend immensely on the realistic representation of the ABL atmospheric processes in the numerical weather prediction and climate models respectively (Garratt, 1994). More particular, it is important to simulate successfully the ABL flow in order to predict the transport, mixing and dispersion, transformation and interaction with soil and vegetation of the meteorological and atmospheric variables within the ABL. In this regard the most important meteorological and atmospheric variables are the wind, air temperature, air moisture and the inert and reactive atmospheric compounds.

1.1.1.1 Diurnal cycle and internal layering

The ABL depth is characterized by a strong diurnal evolution over continental regions. The main driving processes of the diurnal variations of the ABL are the heat

exchanges in the earths' surface, namely the shortwave and longwave radiative fluxes, the sensible and latent heat fluxes and the ground energy fluxes (Garratt, 1994; Pal and Haeffelin, 2015). These fluxes dictate the surface temperature and as a result, they affect the air temperature near the surface. Additionally, the meteorological conditions, soil moisture and the surface roughness can also influence the diurnal variation of the ABL over a continental region (Davis et al., 2020). On the contrary, over ocean regions which are defined by large heat capacity and as a result the air temperature near the surface undergo slow changes, the diurnal variation is insignificant. Moreover the surface roughness is explicitly related to the waves and is one of the smallest in nature surfaces, hence its' impact is also negligible (Larsen, 2013). It is noteworthy that over coastal regions, the sea-breeze circulations along with the atmospheric and surface conditions and the orography determine the ABL depth (Anurose et al., 2018; De Tomasi et al., 2011; Miller, 2003).

The ABL can be divided in subcategories with regards to its' diurnal cycle development. During the daytime, the surface heating drives the development of the ABL. The thermal instabilities, generated by the surface heating, create convective conditions that give the name convective boundary layer (CBL) or mixed layer (ML). The ABL is separated from the free atmosphere by the entrainment zone where exchanges between the ABL and the free atmosphere occur during the day. On the other hand, the radiative cooling during the night time, creates a stability in the atmosphere near the surface thus leading to a stable boundary layer (SBL), also known as nocturnal boundary layer (NBL) (Fernando and Weil, 2010). The shifting periods from the SBL to the CBL and vice versa are known as morning and evening transitions respectively. A neutrally stratified layer remains above the SBL during the night that is not being affect by the surface related properties, known as the residual layer (RL) (Fochesatto et al., 2001). The characteristics of the RL, such as mean state and concentration variables, remain initially the same as the ones of the preceded ML (Stull, 1988). There are not many exchanges between the residual layer and the free atmosphere, hence the layer that separates the two is known as the capping inversion. Figure 1 showcases the diurnal evolution of the ABL.

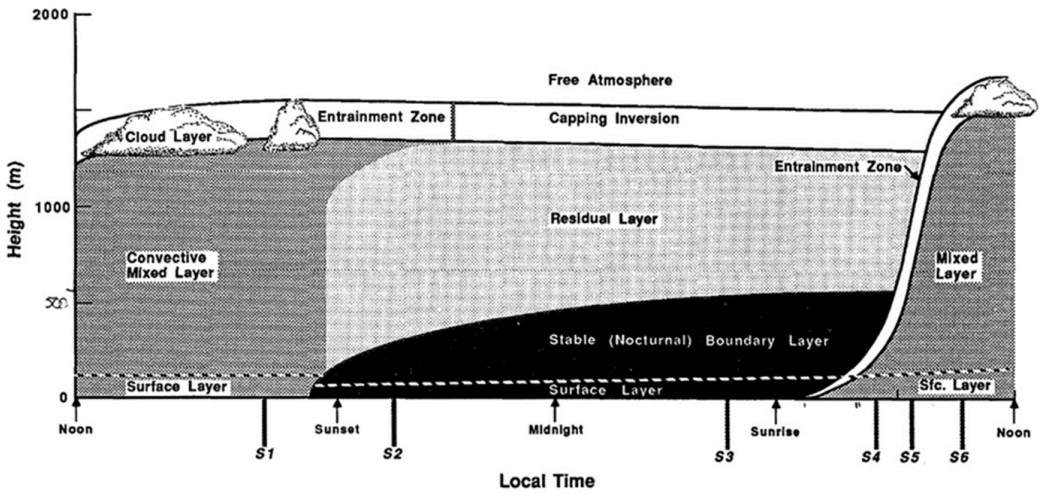


Figure 1: Diurnal evolution of the ABL. Adapted from Stull (1988).

1.1.1.2 Impact on pollutants' dispersion

Since the different subcategories of the ABL are defined by diverse stability conditions, consequently the behaviour of the pollutants varies within the ABL. The intense instabilities in the CBL can disperse the pollutants away from the source and clear the atmosphere near the surface. However, under the stable conditions in the SBL, the pollutants can be trapped near the surface and eventually accumulate into large concentrations (Russell et al., 1974; Sportisse, 2010). The role of the RL can be also important in the concentration of the pollutants. As pollutants remain in the RL during the night, they can be entrained near the surface on the following day (Kim et al., 2007). It is also noteworthy that although pollutants are able to spread rapidly upon entering into the atmosphere, their removal is a rather slow process. More specifically, pollutants are removed from the atmosphere explicitly by naturally occurring processes, such as dry deposition on surfaces, precipitation, radioactive decay and chemical reactions. These removal processes can be harmful for the environment as well as for human health. The products of chemical reactions, e.g. photochemical oxidants can be catastrophic for the atmosphere and combined with phenomena such as acid deposition and acid rain can cause serious issues on human health (Johnson and Siccama, 1983). The movement of pollutants in the atmosphere is driven by the synoptic wind field, the turbulent motions (see Section 1.1.2) causing dispersion in all directions and the deposition (Bolin and Persson, 1975). Their movement along with the emission rates of the pollutants' source and the atmospheric conditions define the evolution of pollutants' concentrations. This is essential information in order to determine areas suitable for the installation of potential sources of pollution as well as to predict possible long-term consequences of their utilization. In this manner several dispersion models are developed for the mathematical

description and prediction of the spatial and temporal distribution of contaminants released in the atmosphere (Heinsohn and Kabel, 1998).

1.1.1.3 Measuring the ABL height

For the aforementioned reasons, the urgency to estimate accurately the ABL height is apparent. Fortunately, there are various meteorological parameters to exploit in order to estimate the ABL. Most notably the temperature, the humidity and the aerosol or molecules' concentrations. More particularly, in the top of the ABL there is a definite temperature inversion, meaning that the temperature is increasing with height at that level (Kotthaus and Grimmond, 2018). In this case the ABL height is considered at the base of the temperature inversion (Ao et al., 2012; Seibert et al., 2000). Additionally, potential temperature can provide an estimation of the ABL height, with the maximum vertical gradient level corresponding to the ABL height (Oke, 1988). Regarding humidity, a significant decrease is observed in the absolute and relative humidity between the ABL and the free atmosphere (Guo et al., 2011). For both the specific and the relative humidity, the ABL height is estimated as the level where the minimum of their vertical gradients is found (Ao et al., 2012; Chandra et al., 2014). Aerosol or pollutant concentrations can also reveal the level of the ABL height as a clear separation between the high aerosol or pollutants concentrations inside the ABL and the clean free atmosphere can be observed (Cimini et al., 2013; Pal et al., 2010).

Such parameters can be observed by in-situ and/or remote sensing methods, using different experimental techniques and analysis methods employing sodar, lidar, wind-profiling radar, tethersonde and radiosonde observations (Seibert et al., 2000). Overall, radiosonde observations serve as the superior technique for the determination of boundary-layer height using profiles of temperature and specific humidity (Schmid and Niyogi, 2012). However, radiosonde observations are expensive so they are practically limited to fixed time intervals, usually 12 or 6 hours, 3 hours at best (Seidel et al., 2012; Wang and Wang, 2016; Zhang et al., 2014). The evolution of the convective boundary layer (CBL) can be rapid during the forenoon period (De Tomasi et al., 2011; Tucker et al., 2009), hence such observations may have severe limitations for capturing rapid CBL development. Furthermore, the determination of the daytime CBL height using continuous sodar observations is often constrained by the upper height limit of observations with sufficient signal-to-noise ratio (mostly <800 m height) (Lokoshchenko, 2002).

On the other hand, continuous profiles of backscattered signal strength and structure constant observed using wind-profiling radars or lidars are good tracers of the ABL height and its diurnal variation (Bianco et al., 2011). Lidar observations can provide the mixed-layer height (Chen et al., 2001; De Tomasi et al., 2011), but in general, the lidar-

derived mixed layer height is slightly greater than the thermodynamic boundary-layer height as thermals can overshoot to greater heights than the top of the adiabatic temperature profile (Caicedo et al., 2017). Another potential tool in order to determine the ABL height and its diurnal evolution during both cloudy and clear-sky conditions, is the continuous observation of temperature and humidity profiles derived from a microwave radiometer profiler (MRP) (Cimini et al., 2006; Coen et al., 2014; Solheim et al., 1998). High temporal resolution of such observations enables determination of the rapid growth of the daytime CBL as well as its decay (Renju et al. 2017). In general, the height of the CBL can reach 5 km in low-latitude deserts during a summer day and only 500 m over the ocean (Mehta et al., 2017). Concerning the SBL, its' height is usually below 500 m (Garratt, 1994).

1.1.2 The turbulent atmosphere

The influence of surface friction, heating etc. is quickly and efficiently transmitted to the entire ABL through the mechanism of turbulent transfer or mixing. Momentum, heat and mass can also be transferred downward to the surface through the ABL by the same mechanism (Arya, 2001). Atmospheric turbulent flows consist of three-dimensional, rotational motions characterized by high unpredictability. Despite of their unpredictability, turbulent flows can be analysed as a sequence of coherent patterns of velocity, vorticity and pressure known as eddies. Eddies vary over a wide range of intensity, as well as length and time scales. Strong eddies with size equivalent to the depth of the boundary layer are associated with intense wind shear, complex topography and thermals. On the other hand, the smallest eddies have diameters of a few millimetres and are particularly weak due to the molecular viscosity's impact. The moving air parcels have kinetic energy with the term related to the turbulent motions known as the turbulent kinetic energy (TKE). The different eddies' sizes are associated with different TKE values. By analysing the TKE in the spectral space, there are three major spectral regions that correspond to the different eddies' sizes (see Figure 2). The first region corresponds to the large eddies with long periods characterized by the maximum TKE values, representing the energy production. This region is referred as the energy containing region. The second region showcases a decrease of the TKE, however the energy does not dissipate but rather passes from big to small eddies and it is known as the inertial region. Finally, there is the dissipation range representing the smallest scale in atmospheric sciences. As the name implies it is the region where the TKE dissipates to internal energy (Harnby et al., 1985).

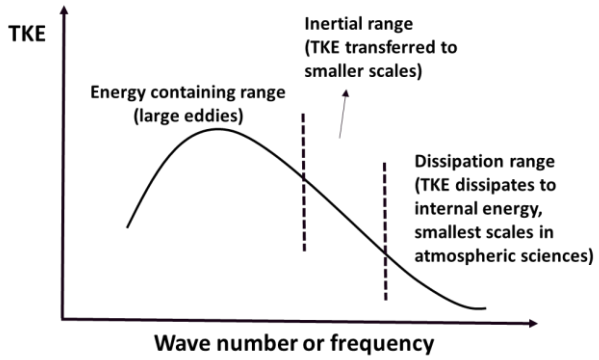


Figure 2: The TKE spectrum showcasing the different ranges with respect to the atmospheric scales.

The urban areas present interesting topographies with regard to development of turbulence. Turbulence is there enhanced by surface roughness due to the complexity of the topography, along with the potential effect in the atmospheric instability as a result of the urban heat island effect (Roth, 2007). It is important to note that, in this manuscript, the focus is only on medium-to-large fluctuations and coherent structures (mlf-cs from few hundred meters to 2 km) and not the small scale turbulence (below 100 m). Additionally, this study examines the coherent structures in the atmosphere. Structures associated with wall turbulence, such as hairpins or packets, are also referred as coherent structures in studies at laboratory scales (Adrian, 2007; Hutchins and Marusic, 2007), but they were not examined during this study.

1.1.2.1 Coherent structures

Although turbulent flows are characterized by high unpredictability, coherent structures can still be formed within a turbulent atmosphere (Tur and Levich, 1992). Coherent structures occur when the eddies are organized in a quasi-periodic pattern in the two horizontal dimensions. This organized area extends over a large area (i.e. a large number of adjacent patterns). Furthermore, a coherent structure must maintain its form for a time period sufficient for time-averaged statistics calculations (Hussain, 1983). Young et al. (2002) have presented in their review the most typical types of coherent structures in details, namely the turbulent streaks, the rolls and the gravity waves.

Turbulent streaks are defined as narrow stripes, aligned with the horizontal wind direction, alternating from high to low horizontal winds. The spacing between the streaks is typically in the magnitude of hundreds of meters, with the vertical expansion sometimes reaching the mixed layer. Their lifecycle is rather short, reaching several tens of minutes before regenerating (Drobinski and Foster, 2003). The major factor for the formation of streaks is the high wind shear development between the surface layer and lower mixed

layer. The turbulent motions driven by the wind shear instead of buoyancy tends to create these streak patterns (Moeng and Sullivan, 1994). The wind shear near the boundary layer top generates small scale turbulent motions which create these alterations between high and low wind values, as the ascending motions diminish the horizontal wind speed and the descending motions accelerate it (Khanna and Brasseur, 1998). Figure 3 illustrates an example of the airs motions during streaks formation.

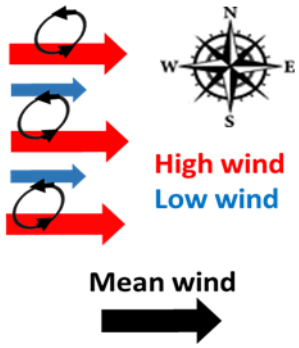


Figure 3: Air motions in the surface layer during streaks occurrence with the black cycles representing eddies.

The streaks can be formed during stable and unstable conditions, but the most favourable conditions for their formation are under neutral or near-neutral stratifications (Khanna and Brasseur, 1998). It is possible to visually detect the streaks formation when they are accompanied by low clouds or fog events (MacDonald et al., 2020; Nakanishi, 2000). In the review of Young et al., (2002) a picture of streak formation during a steam fog event over the Lake Michigan, United States of America, when the patterns are visible, is displayed. It is noteworthy that this type of elongated streak patterns have been observed in laboratory shear flows as well (Kline et al., 1967).

Rolls are also elongated patterns aligned with the horizontal wind direction. They develop in the mixed layer extending from the surface to the capping inversion (LeMone, 1973). The spatial variations of the surface-layer heat flux, the low-level wind shear and the surface homogeneity favour the development and maintenance of the rolls (Weckwerth and Parsons, 2006). Opposite to the streaks, the rolls are formed during strong thermal instabilities when the buoyancy is the dominant parameters compared to the wind shear (Moeng and Sullivan, 1994). The rotation of the large eddies results in ascending and descending motions in the ABL. Figure 4 portrays an example of the air motions during rolls formation.

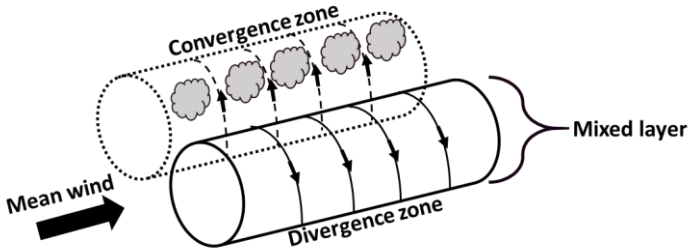


Figure 4: Air motions during rolls formation.

The depiction in Figure 4 corresponds to convective rolls when the saturated atmosphere leads to clouds formation. These cases are particularly interesting as the clouds are formed in rows separated by clear sky areas, a phenomenon known as cloud streets (Lohou et al., 1998). In that manner it is possible to visually detect the rolls cases, especially on satellite images. In Figure 5, an example of cloud streets over Paris and its' surrounding area as recorded by the MODIS[11] instrument aboard the Terra satellite (true colour reconstructed image) is displayed.

Figure 5: Cloud streets over Paris and its' surrounding area as recorder by the satellite true color image of MODIS Terra on 11/09/2014 at 11:10 UTC.

Rolls may also form however under relatively low humidity conditions impeding cloud formation. This phenomenon is known as "dry streets" (Kuettner, 1959, 1971). Such formations in the atmosphere were observed by analysing the behaviour of birds' movement (Woodcock and Wyman, 1947). Furthermore, glider pilots have experienced the presence of rolls during clear sky days and thereafter radar observations have confirmed the existence of "dry streets" (Konrad, 1968). Apart from being much larger structures than streaks, rolls also have a much more extensive lifecycle ranging from hours to even days (Drobinski and Foster, 2003). Young et al., (2002) further distinguished the rolls in two types with regards to their size. Ascending areas that are only one thermal wide correspond to narrow mixed-layer rolls, visually identified as "string of pearls"

[11] MODIS: Moderate Resolution Imaging Spectroradiometer

(Weckwerth et al., 1999) while multiple grouped thermals consist in wide mixed-layer rolls, visually identified as a “band of froth” (Brümmer, 1999).

When the instability in the mixed layer is strong and the synoptic wind is weak, convection does not organize into rolls but into cells (Etling and Brown, 1993). Two types of cells are observed. The first type is the open cell characterized by clear air enclosed by clouds and it is detected during unstable conditions in the surface layer. The other type is the closed cell characterized by cloudy areas surrounded by clear air and it is detected during stable conditions in the surface layer (Lenschow and Agee, 1976). Since a low level wind shear favours the formation of both rolls and streaks, these structures often coexist (Farrell and Ioannou, 2012).

Finally, the gravity waves appear between stable layers of atmosphere with different density. When the equilibrium is disturbed, waves propagation is caused by the efforts of gravity or buoyancy to restore the balance(Fritts and Alexander, 2003). The gravity waves forced from the free troposphere or from within the mixed layer are able to develop coherent structures in the mixed layer during convective conditions (Balaji and Clark, 1988). According to theory, convection waves showcase three propagation behaviours which are: trapping in the capping inversion, trapping in the troposphere, and propagation into the stratosphere (Young et al., 2002). Persistent rolls over a large distance can eventually fuse into larger structures due to interactions with gravity waves. The large number of parameters involved in this mechanism complicates its’ understanding (Etling and Brown, 1993). The spatial periodicity of the rolls is an indication as to whether the rolls are imposed by gravity waves or not. Spatial periodicity values equivalent to the Brünt-Väisälä frequency multiplied by the plane velocity presumably imply the existence of gravity waves (Bernard-Trottolo et al., 2004).

In this work, only the turbulent streaks and rolls structures were considered based on the description of Young et al. (2002). The main aspects of these structures are summarized in Table 1. The gravity waves structures were out of the scope of this study, as the physical parameters available during the campaign for the characterization of the structures were limited. Similarly, the cells structures were not considered in this work.

Table 1: Characteristic physical aspects of the rolls and streaks structures.

	Rolls	Streaks
Height	**Mixed Layer**	**Near the surface**
Horizontal spacing	**Few km to few 10 km**	**Few 100 m**
Duration	**Hours**	**Tens of minutes**
Orientation	**Aligned with the synoptic wind**	**Aligned with the synoptic wind**
Wind conditions	**High or moderate**	**High or moderate**

1.1.2.2 Coherent structures and air pollution dispersion

As already exposed in Section 1.1.1.2, atmospheric instability and turbulence plays a major role in the dispersion of pollutants. The coherent structures and more particularly rolls can also affect the pollutants' concentrations. Ferrare et al. (1991) observed elongated patterns of ascending and descending motions by utilizing aerosol lidar observations: at the top of the updrafts the aerosol backscatter signal was enhanced. Those observations are also supported by a modelling study by Sandeepan et al. (2013), which used simulations from the Weather and Research Forecasting (WRF) model coupled with the particle trajectory model FLEXPART (Stohl et al., 2005). The results show that the pollutants' concentrations alternate between high and low concentration in the updraft and downdraft areas of the coherent structures. Since the orientation of the structures is aligned with the wind direction, the horizontal dispersion of the pollutants is also determined (Zilitinkevich, 1995). Various other studies have also acknowledged the significance of the coherent structures in the pollutants' concentrations (Aouizerats et al., 2011; Gisch et al., 2018; Soldati, 2005). It is therefore particularly important to observe and analyse the coherent structures events in a reliable and consistent way.

1.2 Observing coherent turbulent structures

Up to this point several studies focusing on coherent structures with regards to their characteristics have been addressed. Fortunately, there are various measurement techniques in order to observe the coherent structures. The first person to indicate that turbulence measurements can be derived from wind velocity measurements was Taylor (1938). His hypothesis, known as frozen turbulence, suggests that an eddy can be considered to remain the same as it passes by a sensor. However, this is not true for all cases and therefore Taylor's simplifications should be considered only for those cases where an eddy develops for a longer time span than the one the eddy requires to pass by a sensor (Powell and Elderkin, 1974). One of the most prevalent technique is by using aircraft measurements. This is a significantly advantageous technique, as the aircraft not

only provides atmospheric measurements in different heights of the ABL, but also visual identification of the structures; for instance, in Brümmer et al. (1985) a photo taken from the aircraft showcases the cloud streets formations. Another remarkable benefit of using aircraft measurements is the ability to obtain observations in areas where other techniques are difficult or even impossible to apply such as over the sea (e. g. Martin and Bakan, 1991). Despite the significant benefits of using aircraft measurements, it is not an accessible technique for the vast majority of researchers and furthermore, it is not possible to carry out this method for a long period of time. Moreover, the aircraft provides measurements only in one direction of the structure and performing measurements in two directions at different altitudes takes several hours. In that regard observations from meteorological towers can provide continuous measurements for long time periods. For example the study of Barthlott et al. (2007) covered a 10-month period of observations for the area of Paris identifying 36% coherent structures cases over the total study period. However, meteorological towers are limited in height, thus making essential the use of supplementary measurements in the forms of radiosonde ascents, tethered balloon soundings etc. (Smedman, 1991). Furthermore, the meteorological towers are installed in specific locations, most often rural areas, hence limiting the study areas.

1.2.1 A history of remote sensing observations

Remote sensing methods provide the benefits of continuous measurements and the ability to install the instrument in a plethora of places. Moreover, the remote sensing instruments are able to scan large areas in a short period of time (Kunkel et al., 1980). They can provide measurements even over offshore regions with a radar or lidar installed in a ship (e. g. Achtert et al., 2015; Lund et al., 2012). Cloud streets are even visible via satellite observations (Figure 5). Some of the references previously cited in Sections 1.1.2.1 and 1.1.2.2 refer to radar and lidar observations with regards to coherent structures. The aim of the majority of these studies is to examine the rolls structures in the mixed layer (e.g. Eymard and Weill, 1988; Kelly, 1982; Lin et al., 2008; Rabin et al., 1982). However, the added benefit of instruments such as radar and lidar is the capability to perform horizontal scans near the surface and therefore detect the streaks structures in the surface layer. The earliest study regarding the observation of coherent structures near the surface using a lidar was made by (Weckwerth et al., 1997a). The study of Drobinski et al. (2004) is also a characteristic example where the coherent structures were visible in the horizontal scans near the surface. Another added benefit is the convenience to switch between different scanning methods and as result use horizontal or vertical scans, vertical line-of-sight beams measurements or a combination of the above. In this way, one can combine vertical and horizontal observations in order to estimate the vertical and horizontal scales of the structures (e. g. Lothon et al., 2006).

Kropfli and Kohn, (1978) were the first to implement a dual-Doppler radar to study horizontal rolls structures. The utilization of two radars allows the retrieval of the wind speed fields in two and three dimensions with high spatial and temporal resolution. Therefore, Kropfli and Kohn were able to unveil the wind field in the three dimensions. Subsequently the dual-Doppler method expanded to the lidars as they operate using the same principle as radars (Calhoun et al., 2006; Newsom et al., 2005; Xia et al., 2008). Newsom et al. (2008) presented a qualitative analysis of the wind components' behaviour under stable, unstable and neutral conditions achieved by using dual-Doppler observations. The dual-Doppler lidar can be combined with *in-situ* measurements, as in Iwai et al. (2008a). Their study combines dual-Doppler lidar observations with measurements recorded by a sensor mounted on a helicopter. This measurement combination provided the ability to observe simultaneously rolls and near surface streaks.

It is clear that the dual-Doppler lidar method has many advantages. However, it is still limited to one horizontal and the vertical wind components (Davies et al., 2005). Mann et al. (2008) were the first to use three wind lidars that intersect in a point and so turbulent statistics for all wind components can be provided. Since the lidars measure in a point we do not have to assume horizontal homogeneity, hence this method can be applied for a complex terrain (Klein et al., 2015). Nonetheless, this is a significantly expensive method and therefore, only a handful of studies with lidars have been performed to date. Despite the exceptional benefits of radar and lidar observations, the studies thus far refer to short periods of time and only for particular case studies. An example of such a study is the one of Träumner et al. (2015) where the data recorded by two lidars were combined with the subjective identification of the structures by individual researchers. This methodology complicates the analysis of a large dataset and furthermore it is challenging to retain consistency. Finally, it is important to mention that several studies combining and benefiting from both in-situ and remote sensing observations have been carried out in order to study coherent structures (Bernard-Trottolo et al., 2004; Drobinski et al., 1998; Lohou et al., 1998). On the other hand, it is noteworthy that there is no study with a combination of wind lidars with a water vapour lidar or a temperature lidar.

1.2.2 Lidar scanning methods

For convenience, in the rest of this manuscript, the different scanning methods will be described only for the wind lidar which was used during this study. All the methods, however, could be applied for a wind radar. The wind lidar emits a laser pulse in the atmosphere and receives the light backscattered by the particles (e.g. dust, water droplets, aerosol etc.). Subsequently, the Fourier transform is used to determine the backscattered signal Doppler frequency shift from which, eventually, it is possible to estimate the winds

along the beam direction. The particles move in the same direction and at the same speed as wind (Cariou et al., 2007). So the measured parameter is the radial wind speed u_r in the beam direction. Figure 6 illustrates a single line of sight (LOS) beam with radial wind speed u_r along with the relative spatial parameters.

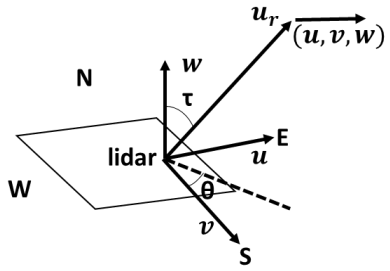


Figure 6: Schematic of a single LOS beam with radial wind speed u_r, where u is the zonal and v the meridional components of the targets' horizontal motion and w is the target's vertical velocity along with the relative spatial parameters azimuth angle θ, the angle between the vertical axis and the beam orientation τ and the cardinal directions.

The u_r can be analysed in its' zonal (West-East) and meridional (South-North) components by the following expression:

$$u_r = u \cdot \cos\theta \cdot \sin\tau + v \cdot \sin\theta \cdot \sin\tau + w \cdot \cos\tau + \epsilon \qquad (1)$$

where θ is the azimuth, τ is the angle between the vertical axis and the beam orientation, u is the zonal and v the meridional components of the targets' horizontal motion and w is the target's vertical velocity. The term ϵ in Eq. 1 refers to the lidar measurement error. It is mainly caused by the imperfect frequency control and the random movement of the scattering particles used as targets (Doviak and Zrnic, 1993).

1.2.2.1 Conical scans

Lhermitte, (1962) and Browning and Wexler, (1968) implemented a method for the estimation of the horizontal wind speed and direction based on wind radar observations, which works also for wind lidar observations. The method known as velocity azimuth display (VAD) or plan position indicator (PPI) method utilizes conical scans for azimuth angles from 0° to 360° and for a steady elevation angle. An example of a PPI conical scan representation is portrayed in Figure 7.

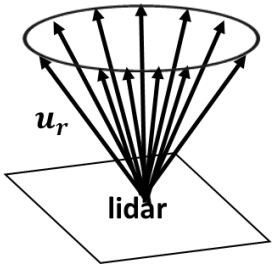


Figure 7: Visual representation of a PPI conical scan.

Figure 8 showcases an example of quasi-horizontal PPI scan (1°elevation) recorded in Paris during the VEGILOT campaign. The positive values (red colour) correspond to particles moving away from the lidar and the negative values (blue colour) particles moving towards the lidar. Therefore, in this example the wind is from the northeast.

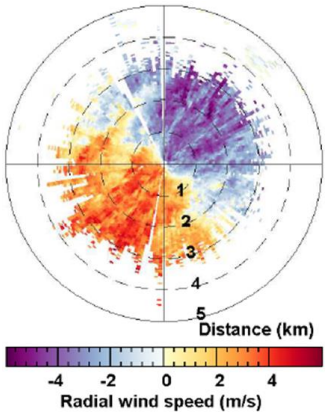


Figure 8: Radial wind speed field from a PPI quasi-horizontal scan recorded during the VEGILOT campaign in Paris at 09:42 on 08/09/2014. The positive values correspond to particles moving away from and the negative to particles moving towards the lidar.

If we assume horizontal homogeneity so that the wind components depend only on altitude and time, and not on the azimuth and range (Eberhard et al., 1989), the radial wind speed measured at a given distance (along a “ring”) will follow a cosine function on the azimuth (Figure 9). Therefore, the data are fitted using the following function:

$$f(\theta) = a + b \cdot cos(\theta - \theta_{max}) \qquad (2)$$

where θ is the azimuth angle of the beam, b is the mean wind speed, θ_{max} is the direction the wind is going to, and a is the offset associated with the vertical component of the wind, i.e. the term $w \cdot \cos\tau$ in Equation (1) (Newman et al., 2016). For a horizontal scan with elevation angle φ=0° (τ=90°), the offset a is zero.

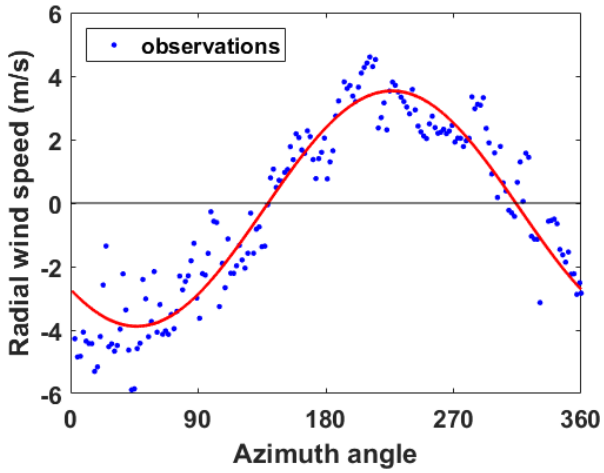


Figure 9: Radial wind speed (blue dots) as a function of the azimuth angle fitted by a cosine function, wind lidar observations recorded during VEGILOT campaign in Paris at 09:42 on 08/09/2014.

The difference between the radial wind u_r and the fitted curve $f(\theta)$, equation (2) where θ_{max} ~ 50° and b ~ 3,5 m·s^{-1}, represents the fluctuations of the radial wind speed u_r'. If this method is applied to all the "rings" at various distances from the lidar, it is then possible to plot the u_r' field (Figure 10), which basically represents the mlf-cs field for the entire scanning area.

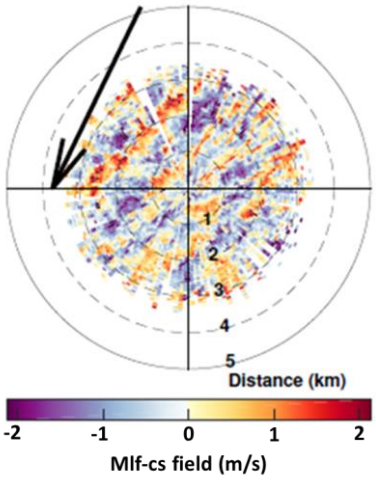


Figure 10: Mlf-cs field with the wind direction (black arrow) during the VEGILOT campaign at 09:42 on 08/09/2014.

Similar patterns of coherent structures near the surface were observed by Drobinski et al. (2004) studying a rural area in United States of America. On the other hand, Lohou et al. (1998) have demonstrated the ability to identify convective rolls and cells by combining 20 radar PPI scans, in elevation angles ranging from 0.5° to 50°, with airborne

observations at different altitudes. The duration of a PPI scan may vary from 1 to 3 minutes depending on the azimuth resolution. As the wind field have to remain homogeneous during the scanning period, complex terrains/flows makes the VAD method inapplicable (Sathe and Mann, 2013).

1.2.2.2 Wind vertical profiling

If the goal is only to estimate the 3 wind components, and not observed the mlf-cs fields, a faster scanning method is preferable to assure the homogeneity of the wind field In the Doppler beam swinging (DBS) method (Strauch et al., 1984), u_r measurements are made only in the north, south, west, east horizontal and vertical directions (Figure 11), so they depend only on one of the horizontal component of the wind:

$$u_{rV} = w \tag{3}$$

$$u_{rE} = u \cdot cos\varphi + w \cdot sin\varphi \tag{4}$$

$$u_{rS} = -v \cdot cos\varphi + w \cdot sin\varphi \tag{5}$$

$$u_{rN} = v \cdot cos\varphi + w \cdot sin\varphi \tag{6}$$

$$u_{rW} = -u \cdot cos\varphi + w \cdot sin\varphi \tag{7}$$

where the radial wind are labelled according to the cardinal direction and V for vertical and u, v and w are the three component of the wind vector, as in Equation (1) and φ is the elevation angle from the ground. Then a linear combination of these measurements provides the zonal and meridional wind components at different heights, as well as the vertical component of the wind. This is exceptionally useful for the identification of high wind shear or even low level jet (llj) cases (Cheliotis et al., 2020). This technique can be also applied using only three beams (e.g. north, east and vertical).

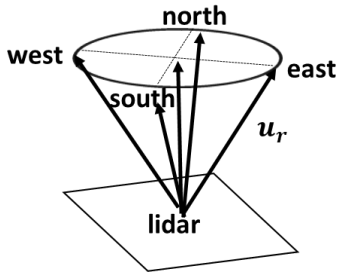


Figure 11: Visual representation of a DBS scan, from Weitkamp, (2006).

The measurements of the u and v can lead to the estimation of the total horizontal wind V_h via the following expression:

$$V_h = \sqrt{u^2 + v^2} \qquad (8)$$

If the V_h is computed for several heights of the DBS scans, the vertical profile of the wind shear is revealed. This is significantly useful, as the high wind shear near the surface plays a major role in streaks formation (see Sec. 1.1.2.1).

1.2.2.3 Vertical scans

If one of the goals is to visualize the vertical structures in the ABL, the range height indicator (RHI) method is more appropriate (Gal-Chen et al., 1992). In this method two consecutive verticals scans in the North-South and East-West directions are carried out. The elevation angle rotates from 0° to 180° (Figure 12). It is important to assume horizontal homogeneity also for the RHI method.

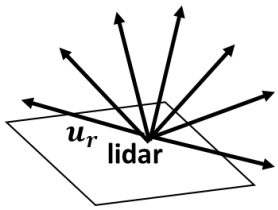


Figure 12: Visual representation of a RHI scan.

An example of the visualization of structures in the vertical plane is presented in Figure 13. The horizontal wind component $v = \frac{u_r}{cos\,\varphi}$ is estimated and plotted for the whole vertical scan in the North-South direction. The changes in the sign of v imply the presence of strong fluctuations. Similar patterns were observed by Drobinski et al. (2004) in RHI observations.

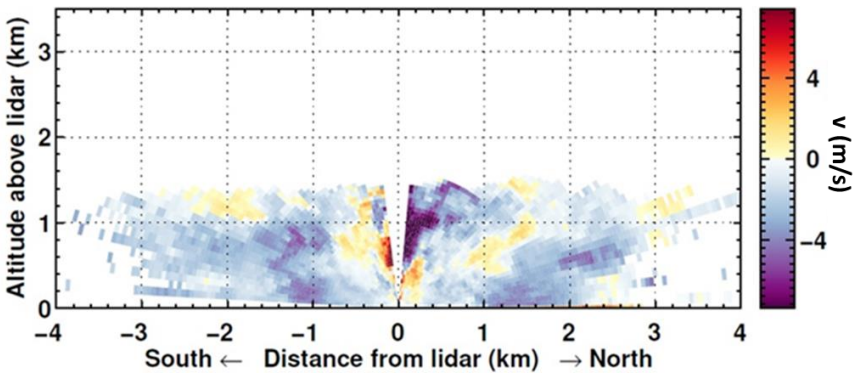


Figure 13: The horizontal wind component from a RHI scan during the VEGILOT campaign on 08/09/2014 at 13:23. The altitude is given as “above lidar” because the latter was installed on a tower (see Sec. 2.1).

If the RHI scan is not aligned with the streamwise and cross-stream direction relative to the mean wind, there will be cross-contamination in the estimation of the horizontal wind components (Gal-Chen et al., 1992). Furthermore, the RHI method can take up more than two minutes to perform both vertical scans. For this reason, many researchers prefer to do an arc scanning with a low elevation angle (e.g. 0°-20°) to achieve a higher repetition frequency of the observations. Examples of this approach can be found in studies regarding the estimation of turbulence parameters (e. g. Banta et al., 2006; Pichugina et al., 2008) as well as for turbulent structures (e. g. Drobinski et al., 2004). In particular Banta et al. (2006) and Pichugina et al. (2008) were implementing PPI scans for 20-30 minutes for the retrieval of wind direction and then adjusting the orientation of the RHI scan accordingly for the estimation of wind variances so the scan was always aligned with wind direction. Due to the short time period between the scans, it is thus possible to observe the evolution of the rolls in time.

1.2.2.4 Turbulent fluxes vertical profiling

In the search of the most efficient and rapid method to obtain information about the wind components observed from a lidar, Sathe et al. (2015) developed the six-beam technique. Through several testing, they detected the optimal combination of azimuth angles and number of beams in order to estimate the wind components with a minimum error. The result of the optimization process indicated that the optimal configuration will be a modified DBS scan with five beam directions equally spaced in azimuth (instead of four). Regarding the elevation angle they selected 45° as this is the threshold for considering horizontal homogeneity of turbulence. This limit is arbitrary and could be higher or lower according to the homogeneity of the terrain. The validation of their method in relation to a sonic anemometer showed a good agreement overall. The best results were observed during stable conditions. More studies have been carried out recently to examine the

performance of this method compared to the other established scanning methods. Newman et al., (2016) compared the PPI, the DBS and the six-beam methods with regards to the atmospheric turbulence measurements. The six-beam technique seems to reduce some of the errors caused by the PPI and DBS methods. However, under unstable conditions for two different type of lidars utilizing the six-beam method, the observations overestimate significantly the wind variances. Bonin et al. (2017) compared the six-beam method with the PPI and the low elevation RHI methods. The comparison was made for TKE, turbulence intensity and shear velocity. The lidar observations were validated with sonic anemometers observations at six different heights. The six-beam method shows more accurate measurements for TKE and turbulence intensity compared to PPI and RHI. All methods were unable to systematically give accurate measurements for shear velocity. Since this is a new method further validation of the six-beam method should be carried out.

1.3 Automated image Classification

A critical part of this study was the feasibility to identify and classify patterns of given images. This is not a complicated job for humans when the number of images is small. However, it becomes cumbersome or even impossible for a large number of images and therefore it is essential to use another consistent and reliable approach. Fortunately, computers are able to interpret and categorize images by applying a suitable algorithm. Once properly trained, such algorithms can separate images based on objects like guitars, plates, cups even in complex backgrounds (Kuznetsova et al., 2018). Image classification by computers have spread to many fields, for instance to detect breast cancer in early stages (e. g. Deniz et al., 2018; Holli et al., 2010).

In meteorology, the classification of satellite images based on cloud types is one of the most researched topic (e. g. Azimi-Sadjadi and Zekavat, n.d.; Gu et al., 1989). An accurate classification of the cloud types is a very useful tool for forecasters or nowcasters who are interested in the interpretation of cloud images (Bankert, 1994). Furthermore, the image classification can be utilized for weather circulation patterns (e. g. Philippopoulos et al., 2014) or weather events (e. g. Sokolov et al., 2020) in order to better understand the association between weather patterns and meteorological parameters or to estimate statistical parameters respectively. Even more complicated images, such as forest coverages by satellite observations, can be classified by computers and thus avoiding the mapping of a region by taking aerial photographs (Kayitakire et al., 2006). Closer to the subject of this thesis, (Wang et al., 2020) developed an automated detection method for rolls in the marine boundary layer, based on the images from a synthetic aperture radar.

There are two ways to perform an image classification. The first one is based on building a training ensemble, i.e. a set of typical images that need to be classified. Then, an ensemble of relevant parameters is computed for each image in the training ensemble, designed to highlight the characteristic features (textures) of the image types. Each image is at this stage represented by a vector of parameters, i.e. a point in a N-dimension space (N being the number of texture analysis parameters). Finally, a machine learning algorithm is applied to examine which of the parameters allow to best cluster the points and minimize the error for the ensemble (Schowengerdt, 2007). Consequently, the selected parameters can be used for the classification of a larger ensemble. This method is known as the supervised classification. On the contrary, the second way does not require a training ensemble or pre-defined image types: the texture analysis parameters are directly provided to the algorithm that will classify the images and define the image categories. This method is named unsupervised classification (Olaode et al., 2014). In the current study, the aim was to classify the patterns of the mlf-cs fields by providing a training ensemble built on the patterns as well as some striking atmospheric physics for each structure. Therefore, the supervised machine learning method was preferred.

1.3.1 Texture analysis parameters

The first step of the image classification is to simplify the image data for the computer to process them. For the lidar images that will be used in this study, the turbulent radial wind speed can be represented by a gray-scale with decimal numbers for instance from −2 to +2 m·s. Then, these numbers can be grouped into a few bins, usually 4 or 8 is enough, with the boundaries of the bins depending on the details that need to be highlighted from the images (Partio et al., 2002). It is for instance possible to define more bins around the center of the scale (if there are little outliers and the central shades hold more information), or conversely more bins near the edges of the scale (if the extreme values hold more information).

For a specific distance and angle between two pixels of the image (e.g. first neighbour in y direction, Figure 14b), a matrix can be constructed, named gray level co-occurrence matrix (GLCM or CM) (Haralick et al., 1973). In the CM, the cell at the intersection of column i and row j contains the number of neighbor pixel pairs in the image whose colors belong to the bins i and j. Figure 14a presents an ideal case with a simplified wind field in a 24 by 24 pixel image with values divided in 8 bins between −2 and +2 m·s. Figure 14c presents the CM for the adjacent neighbours in the y direction for this ideal case. The CM in this case is an 8 by 8 matrix as there are 8 wind bins / colors. Because the wind field in this case is composed of vertical stripes, the adjacent pairs in the y direction have always the same color and therefore only the diagonal cells of the CM have

non-zero values. For a given color, there are 3 stripes on the image, each containing 24 pixels, which means 3×23=69 pairs of pixels.

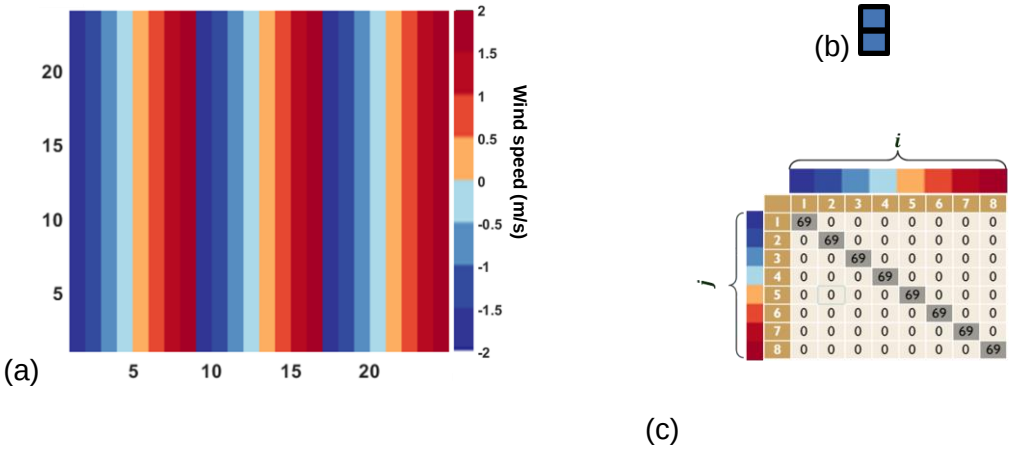


Figure 14: Example of an ideal case for the estimation of the CM: (a) Ideal wind field of 24 by 24 pixels (x and y axis), (b) example of adjacent in the y direction neighbour pair and (c) the corresponding CM for this type of pairs.

Likewise, if we examine the pair of pixels corresponding to the 8th horizontal neighbour (Figure 15a), as they are separated by one period, the color value is always the same and there will only be non-zero values in the diagonal of the CM (Figure 15b). On the contrary, for the 4th horizontal neighbour (Figure 15c), the pairs are separated by half a period so the values will be distributed in the sides of the CM (Figure 15d).

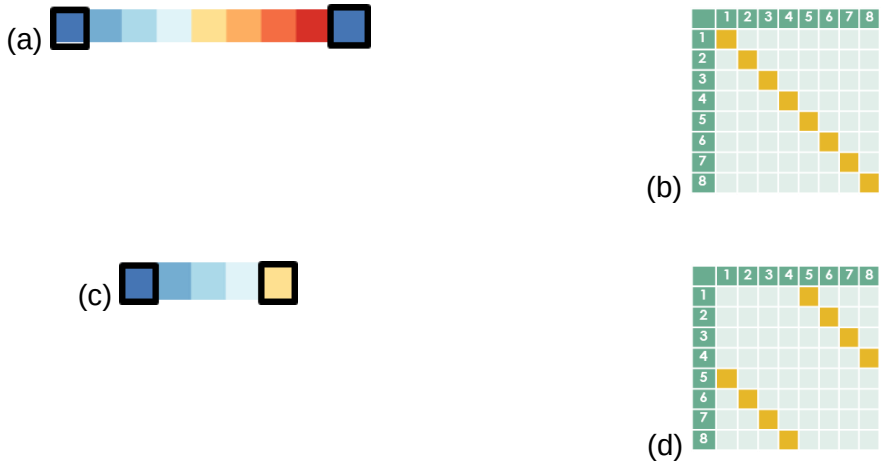


Figure 15: (a) Example of horizontal pair 8 grid cells away with (b) a scheme of the corresponding CM (non-zero cells are highlighted in orange). (c) Example of horizontal pair 4 grids cells away with (d) a scheme of the corresponding CM.

The CM can be calculated for any distance (neighbour order) n and any pair orientation, measured by an azimuth angle φ (Figure 16). In other words, any field can provide a $n \times \varphi$ number of CM.

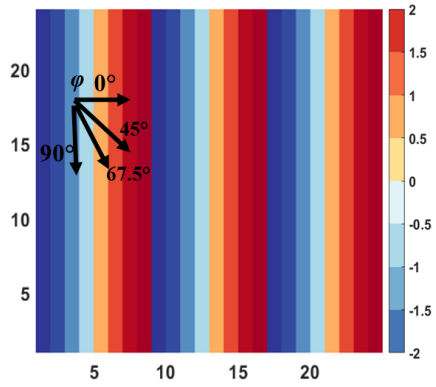


Figure 16: Possible orientations φ for a pair of cells.

Haralick et al., (1973) presented 14 parameters computed from the CM and able to reveal similarities or differences between the CM. The four most commonly used and characterized as the most important parameters are the contrast, correlation, homogeneity and energy (e. g. Clausi and Zhao, 2002; Haralick, 1979; Srivastava et al., 2018). These parameters were estimated for the current study as well. Contrast is related to the variations of the CM with the value 0 representing constant textures. Correlation refers to how well correlated are the grid points of the field with the values ranging from -1 (perfect negative correlation) to 0 (uncorrelated) and to 1 (perfect positive correlation). Homogeneity is related to the homogeneity of the image with the values ranging from 0 to 1 (diagonal CM). Energy refers to the uniformity of an image with values ranging from 0 to 1 (constant field) (Haralick et al., 1973; Yang et al., 2012). The mathematical expressions of the aforementioned parameters are presented in the following Equations:

$$\textit{Homogeneity: } Hom(\varphi, n) = \sum_{i,j} \frac{p(i,j)}{1+|i-j|} \quad (9)$$

$$\textit{Contrast: } Con(\varphi, n) = \sum_{i,j} p(i,j)|i-j|^2 \quad (10)$$

$$\textit{Correlation: } Cor(\varphi, n) = \sum_{i,j} \frac{(i-\mu_i)(j-\mu_j)p(i,j)}{\sigma_i \sigma_j} \quad (11)$$

$$\textit{Energy: } En(\varphi, n) = \sum_{i,j} p(i,j)^2 \quad (12)$$

$p(i,j) = \frac{CM(i,j)}{\sum_{i,j} CM(i,j)}$ is the normalized cell value for the i,j position in the CM. μ_i and μ_j are the marginal expectations: $\mu_i = \sum_i \sum_j i \cdot p(i,j)$ and $\mu_j = \sum_i \sum_j j \cdot p(i,j)$. σ_i and σ_j are the marginal standard deviations: $\sigma_i = \sqrt{\sum_i \sum_j (i - \mu_i)^2 \cdot p(i,j)}$ and $\sigma_j = \sqrt{\sum_i \sum_j (j - \mu_j)^2 \cdot p(i,j)}$.

Once the training ensemble is built, the texture analysis parameters can be computed. The next step is to examine if there is an algorithm that is able to group the images/fields as expected.

1.3.2 Supervised machine learning

As mentioned previously, the supervised machine learning approach was applied in this study for the classification of images (Bonamente, 2017; James et al., 2000; Kubat, 2017). Several machine learning algorithms can be used for image classification, among which three were utilized during this study (and are among the most frequently used). The main principle, in all three algorithms, is that an image is represented by a vector of texture analysis parameters. For example, assuming the images are represented by two texture analysis parameters, the machine learning algorithm will cluster the images in two dimensions, based on the values of the texture analysis parameters. This is possible if the points belonging to each class are gathered into well-separated clouds of points.

Any new image introduced in the algorithm will be classified in a cloud of points based on the principle of the given algorithm. The simplest of the three algorithms is the K-Nearest Neighbour (KNN) which classifies an image based on the class attributed to its nearest neighbour(s) in the texture analysis parameters space (Cover and Hart, 1967). Depending on how many neighbours are selected for comparison, the results of the classification may differ and therefore the number of neighbours must be carefully chosen. Figure 17 illustrates how the KNN method works. If a new image is introduced in the green circle location and the selected method is the one neighbour KNN, the new image will be classified as its nearest neighbour, i.e. in class 1 (blue square). However, if the three neighbour KNN method is used, the new image will be classified as the majority of the three closest neighbours, i.e. in class 2 (red triangle).

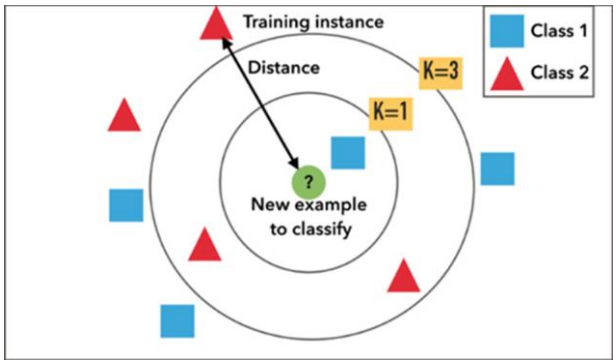


Figure 17: Principle of the KNN method for two classes and either 1 or 3 neighbours. The position of the images (blue squares or red triangles) in the plane correspond to their coordinates in a two-parameter space.

Another relative simple algorithm is the error-correcting output codes-support vector machine (ECOC-SVM) (Vapnik, 2000). In this method, the different classes are separated by an imaginary line, and every time a new case is introduced, the algorithm examines on which side of the line it resides. A simplified example with only two classes is displayed in Figure 18.

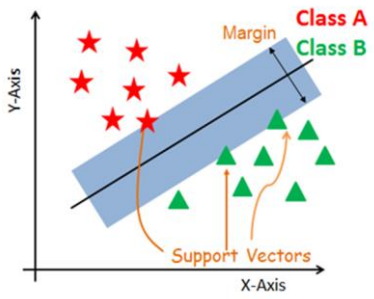
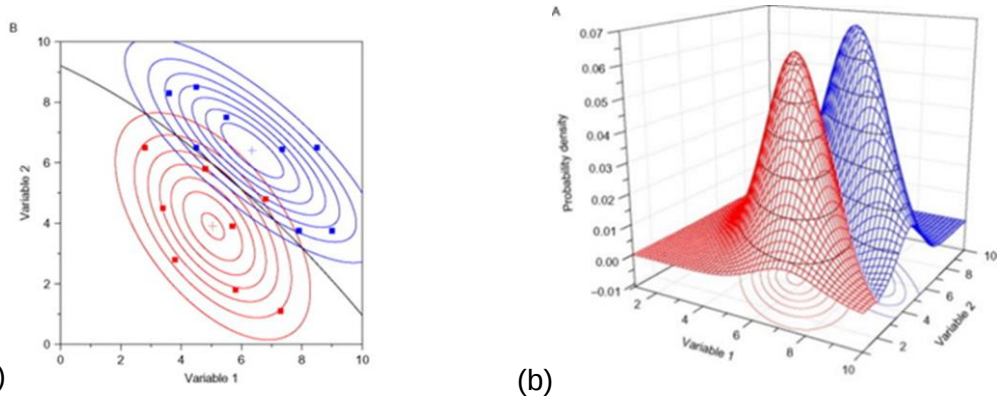


Figure 18: Principle of the ECOC-SVM algorithm for two classes and with two relevant texture analysis parameters for the classification.

The third and final algorithm was also the one which showed the best performance to classify the lidar images and which results were published (Sec. 3). It is the quadratic discriminant analysis (QDA) or normal Bayesian algorithm (Hastie et al., 2009). In this method, the probability density function (PDF) of each class is estimated in the texture analysis parameter space (Figure 19) and a new image is assigned to the most probable class at its coordinates in the parameter space (Kubat, 2017).



(a) (b)

Figure 19: Principle of the QDA algorithm for two classes and with two relevant texture analysis parameters for the classification. (a) Position of the images (red and blue dots) in the two-parameter space with the probability density function (PDF) contours (red and blue lines) and the separation line between the areas where each class is more probable than the other (black line). (b) The PDF represented as a surface plot.

If a new case is introduced in order to be classified, then the probability to fall in either class will be estimated. The case is eventually classified in the class where it has higher probability.

1.4 Image classification for the mlf-cs fields

Despite the repetitiveness of the mlf-cs fields patterns, so far there has not been any attempt to automatize the identification and classification processes of the mlf-cs. In particular, the elongated shape of the coherent structures is distinctive compared to random patterns. In that regard, an automated methodology for the mlf-cs fields was developed based on the image classification principle. This method can simplify and provide consistency in the analysis of large datasets (e.g. (Yagi et al., 2015)). Furthermore, the identification of structures for long periods of time, can lead to unique studies, with a stronger statistical weight, regarding either the physical parameters of the structures or the correlation between the structures and pollutants' concentrations. In Section 2, the detailed development of the methodology is showcased along with its' performance for a training ensemble.

2. Materials & methods

2.1 Experimental set-up

In the framework of the VEGILOT [VEGétation et ILOT de chaleur urbain (vegetation & urban heat island)] project, a campaign was organized by the Université Pierre et Marie Curie (UPMC) in collaboration with the Muséum National d'Histoire Naturelle in order to analyze the transport processes and the atmospheric chemistry in the urban ABL in Paris. The campaign lasted for two months (04/09/2014-06/11/2014), and provided continuous wind, aerosol and ozone profiles. The observations were recorded by utilizing the Leosphere WLS100 Doppler lidar (www.leosphere.com), the Cloud and Aerosol Micro-Lidar (CAML) by the CIMEL company (www.cimel.fr) and the modified version of the Airborne Lidar for Tropospheric Ozone (ALTO) (Ancellet and Ravetta, 1998) (Figure 20). For the current study the observations recorded by the ozone lidar were not used and therefore only the characteristics of the wind and the aerosol lidar will be presented.

(a) (b) (c)

Figure 20: Photos of the three lidars operating during the VEGILOT campaign: (a) Wind lidar, (b) aerosol lidar and (c) ozone lidar.

The primary instrument for the study was the wind lidar, that was used to retrieve the mlf-cs fields from the quasi-horizontal scans. The wind lidar was installed on the roof of the administration building (75 m high) in the Jussieu Campus, which is located in the centre of Paris. Paris city center is a rather interesting location for observations, as despite being a metropolitan city, the building heights do not exceeds 50 m, with minor exceptions, due to local regulations (Saint-Pierre et al., 2010). Therefore, the emitted signal from a 75 m height building, even for quasi-horizontal scans was not interrupted by other buildings. Figure 21 illustrates the location of the Jussieu site along with the area covered by the lidar scans. The maximum range was 5 km across all the scanning methods, with the minimum range (blind zone) reaching 100 m. The scanning area is completely urbanized, with few green areas and a part of the Seine river included. The spatial resolution was 50 m, thus it is possible to study mlf-cs, but not small turbulent structures. All the other properties of the wind lidar were unchanged for all the scanning methods

throughout the campaign with the axial resolution (measurement spacing along the beam direction) being 50 m, the accuracy on the wind observations being ± 0.1 m·s^{-1} and an accumulation time of 1 sec·beam^{-1}. Table 2 showcases the most important wind lidar properties during VEGILOT.

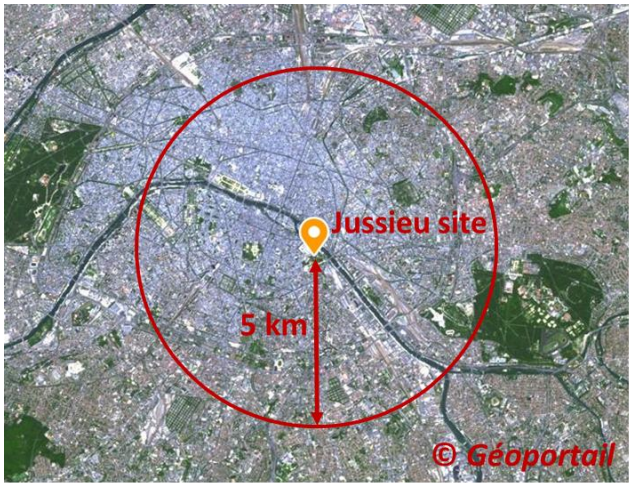


Figure 21: Jussieu site location and the maximum scanning range of a horizontal scan (red ring) (Géoportail satellite image).

Table 2: Properties of the Doppler wind lidar (Leosphere WLS100) for all the scanning methods during VEGILOT.

Altitude of lidar	75 m a.g.l.	**Radial wind speed range**	-30 to 30 m·s^{-1}
Minimum range	100 m	**Laser wavelength**	1.543 µm
Maximum range	5 km	**Radial wind accuracy**	± 0.1 m·s^{-1}
Spatial resolution	50 m	**Accumulation time**	1 sec·beam^{-1}

The ground level is about 30 to 60 m above mean sea level (a.m.s.l.) within most of the scanning area, with a few small areas exceeding 100 m a.m.s.l. (low hills) near the boundaries (Figure 22a). The near-surface scans were quasi-horizontal, but the 1° elevation angle still resulted in a 87 m height difference between the lidar position at the center and the edges of the scanning area. The beam altitude relative to the ground does not vary by more than 20 to 30 m for the first 3 km around the lidar (Figure 22b), which correspond to the area where most of the data were recorded.

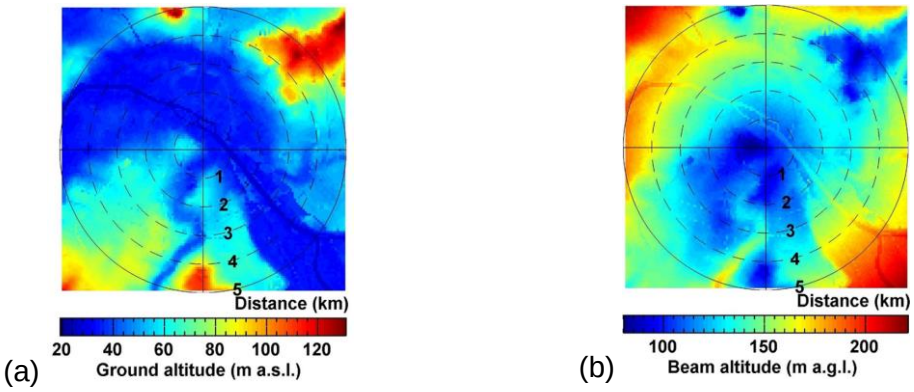


Figure 22: (a) The ground altitude (above sea level) and (b) the beam altitude maps (above ground level) for the scanning area in Paris.

In total, four scanning methods were implemented in a cycle, with various elevation and azimuth angle combinations (Table 3). The duration of the cycle, through all the methods before applying the first method again, was approximately 18 minutes. At first, the lidar performed two RHI cuts in the East-West then North-South direction. A full scan with an angular resolution of 2° lasted approximately one and half minute. The observations were recorded in order to investigate possible structures in the vertical plane. Eventually however, the observations from this method were not used in this study as the RHI orientations made most of the time an angle of about 45° from the mean wind direction, while it would have been necessary to have measurements in the longitudinal and transverse directions relative to the mean wind. The RHI method was followed by the two PPI conical scans with an elevation angle of 25° and 1°. For the full 360° with a 2° azimuth angle resolution, each scan duration was 3 minutes. The observations at 25° elevation were carried out in order to reveal information of the structures in the vertical plane, but eventually there were also not used in this study. The PPI at 1° elevation was the fundamental method for this study, as the mlf-cs fields were constructed based on these observations. After the PPI method, five LOS in fixed directions were performed, but only the vertical one was used in this study to determine the nocturnal ABL height (Section 2.4). The vertical LOS consisted of two sequences of shots lasting 30 s each. Finally, the DBS method was applied and repeated twice, which took approximately 15 seconds for each DBS. The DBS observations were used for the identification of llj cases.

Table 3: The entire measurement sequence during the VEGILOT campaign with the significant scanning characteristics for each scanning method.

	Scanning area	Purpose	Elevation & azimuth angle	Scan duration
Range Height Indicator (RHI)	Vertical arc scans (North-South & East-West)	Not used in this study	Azimuth 0° & 90°, elevation 2° to 180° with 2° resol.	2 × 1,5 min
Plan Position Indicator (PPI)	Conical scanning	Not used in this study	Elevation 25°, azimuth 0 to 360° with 2° resol.	3 min
Plan Position Indicator (PPI)	Quasi-horizontal scans near the surface	Identification of structures	Elevation 1°, azimuth 0 to 360° with 2° resol.	3 min
Line of Sight (LOS)	Vertical scans	Estimation of the ABL height (during the night)	Elevation 75° & 90°, azimuth 0°, 90°, 180° & 270°	10 × 30 sec
Doppler Beam Swinging (DBS)	Combination of LOS	Identification of low level jet cases	Elevation 75°, azimuth 0°, 90°, 180° & 270°	2 × 15 sec

Although the aerosol lidar observations were not the focus of this study, the measurements were nonetheless valuable, as the estimation of the ABL height during the day was performed based on them (see Section 2.4). This instrument records the backscattered light signal of the particles up to an effective range of about 5 km with a blind zone of 200 m and a 15 m spatial resolution. It operates at a 532 nm wavelength with a low pulse energy of 10.6 μJ but a high repetition rate of 4.7 kHz (Pelon et al., 2008). It is possible to observe aerosols and cloud profiles with this instrument. The lidar was installed in the roof of building adjacent to the one where the wind lidar was installed, but only 20 m high. The instrument operated continuously for the period 04/09/2014-06/11/2014 and provided observations with a 10-minute average. Table 4 recaps the significant properties and scanning characteristics of the aerosol lidar during the VEGILOT campaign.

Table 4: The significant scanning characteristics and properties of the aerosol lidar CAML during VEGILOT.

Altitude of lidar	20 m a. g. l.	**Scanning area**	Vertical shots
Minimum range	200 m	**Purpose**	Estimation of the ABL height (day)
Maximum effective range	~5 km	**Laser energy**	10.6 μJ
Spatial resolution	15 m	**Averaging period**	10 minute
Laser Wavelength	532 nm	**Repetition rate**	4.7 kHz

2.2 Building the training ensemble

The mlf-cs classification for this study was performed with the supervised approach. Therefore, it was necessary to recognize the different patterns and build a training ensemble to examine the performance of the algorithm. Firstly, the VAD method (Section 1.2.2.1) was applied to all the 4577 PPI scans at the 1° elevation recorded during VEGILOT. The data were fitted using Equation (2) including the offset parameter a to take into account the slight elevation of 1°. By visually observing the mlf-cs, it was noticeable that some of them contained well defined elongated patterns with alternating positive and negative values of turbulent radial wind. It was also noticeable that during the night these elongated patterns were narrow (Figure 23a) whereas during the day they were wider (Figure 24a). Other mlf-cs recorded during the day presented some large enclosed patterns of positive and negative values (Figure 25a).

2.2.1 Attributing the different patterns to structure types

Streaks

The nocturnal narrow elongated patterns resembled the streaks formation from literature. As I stated in Section 1.1.2.1, a leading factor for the formation of the streaks is the high wind shear near the surface which is revealed by the DBS observations (See Sec. 1.2.2.2). Figure 23a shows a typical scan presenting elongated narrow patterns, alternating between positive and negative values, and aligned with the mean wind direction (black arrow) which resemble streaks formation. The time-height cross-section of the horizontal wind on the same day (Figure 23b) shows a strong wind shear from

19:00 UTC until 00:00 UTC. By examining more cases, it was evident that many elongated patterns co-occurred with a high wind shear near the surface, thus indicating the formation of streaks.

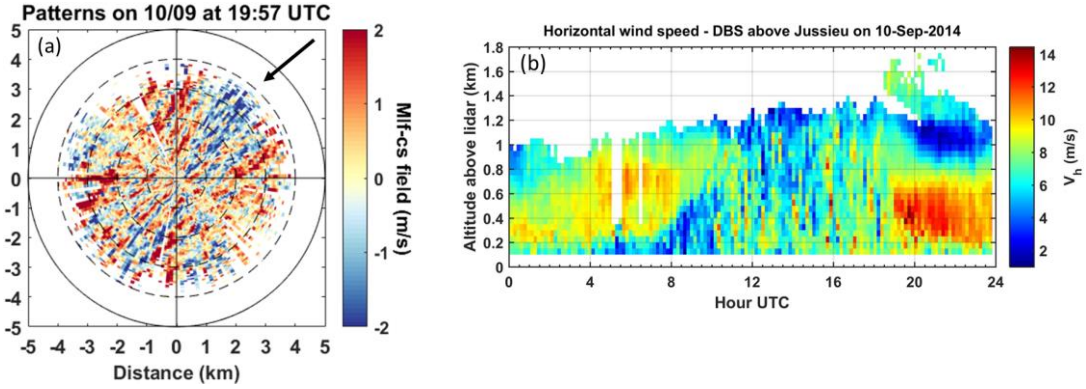


Figure 23: (a) Example of mlf-cs field for the streaks type, with the mean wind direction (black arrow) recorded on 10/09 at 19:57 UTC and (b) the time-height cross-section of the horizontal wind speed profile on 10/09.

Rolls

The wider elongated patterns were also characteristic in some mlf-cs recorded during the day, also aligned with the mean wind direction. To confirm that these wide patterns observed near the surface could be associated with rolls structures in the mixed layer, satellite images were used to detect the cloud streets cases. Figure 24a is showcasing a typical the mlf-cs field presenting wide elongated patterns of positive and negative values aligned with the wind direction. In Figure 24b, the coincident MODIS Terra satellite image shows the presence of cloud streets over the scanning area. The examination of other cloud street cases during the 2-month period under study allowed to attribute to rolls more wide elongated patterns in the mlf-cs fields. The Terra and Aqua satellites pass over Paris only around 11 and 13 UTC, but wide elongated patterns were observed at other time of day, and attributed to rolls provided they belong in a consecutive series of roll scans.

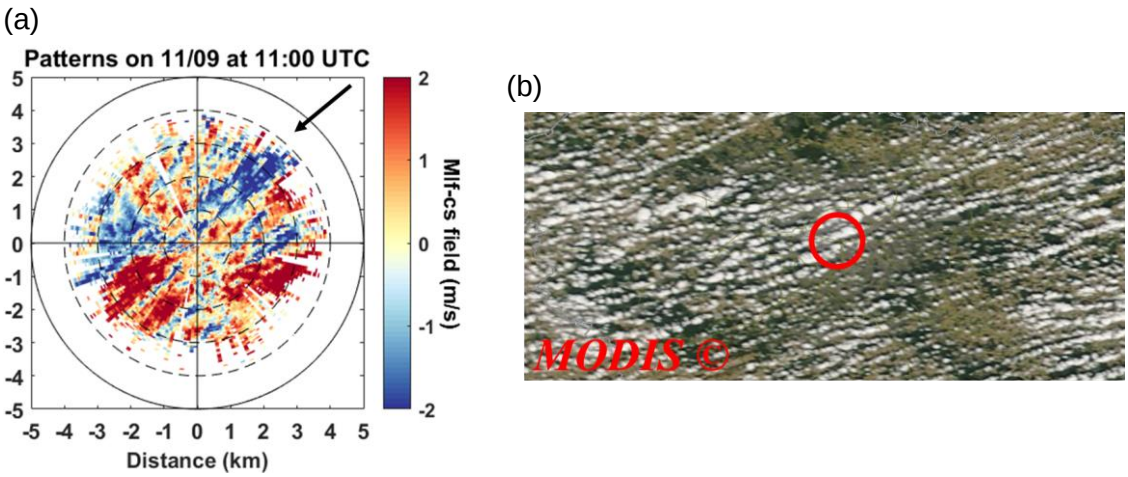


Figure 24: Example of a mlf-cs field for the rolls type, with the mean wind direction (black arrow) recorded on 11/09 at 11:00 UTC and (b) the MODIS Terra satellite image with the lidar scanning area (red ring) recorded on 11/09 at 11:10 UTC.

Thermals

Regarding the large enclosed patterns that can be observed during the day, they were likely formed during random thermals movements creating ascending and descending motions in random parts of the mlf-cs. To confirm this, the sunshine hours during these days and the hourly solar radiation values as recorder by the Paris-Montsouris station were used. The MODIS satellite images (when available) could also reveal the presence of fair weather cumuli weather. The Paris-Montsouris observations confirmed that these patterns occurred when the solar radiation values where above the monthly mean values and the sunshine hours were persistent. Moreover, the MODIS satellite images confirmed that these patterns occur under fair weather cumuli weather. Figure 25a presents an example for a mlf-cs field with large enclosed patterns of positive and negative values along with the Modis Aqua satellite image revealing fair cumuli weather, thus confirming the presence of large thermals during the day.

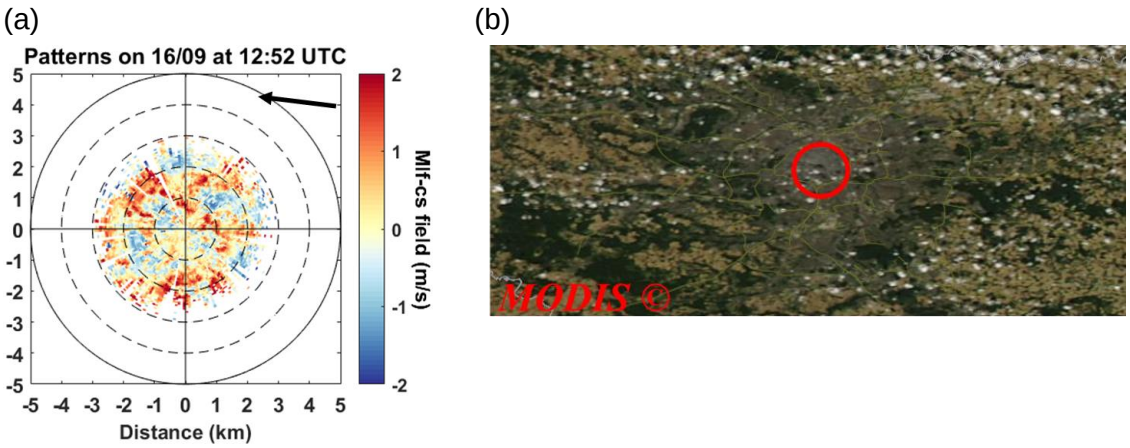


Figure 25: (a) Example of mlf-cs field for the thermals type with the mean wind direction (black arrow) recorded on 16/09 at 12:52 UTC and (b) the MODIS Aqua satellite image with the horizontal scanning area marked (red ring) recorded on 16/09 at 12:50 UTC.

"Others"

Apart from the narrow and wide elongated patterns and the large enclosed patterns, there were also patterns that did not fall in one of the above categories. These were some chaotic random patterns in the mlf-cs fields at a smaller scale than the thermal patterns, or most often cases where the VAD method had not been applied successfully due to a non-symmetric radial wind field (Figure 26a). Throughout the 2-month period under study, there were several such cases that occurred mostly during calm (0-2 $m \cdot s^{-1}$) or low wind (2-4 $m \cdot s^{-1}$) conditions, when the wind does not have a well-defined direction (Wilson et al., 1976). In a case like this, the individual rings of measurements do not follow a cosine function (Figure 26b) so the cosine fit is clearly not accurate. The VAD method was applied automatically so even this type of scan was processed, however, the patterns in the mlf-cs field resemble large slices (Figure 26c) and cannot be associated with any atmospheric structures.

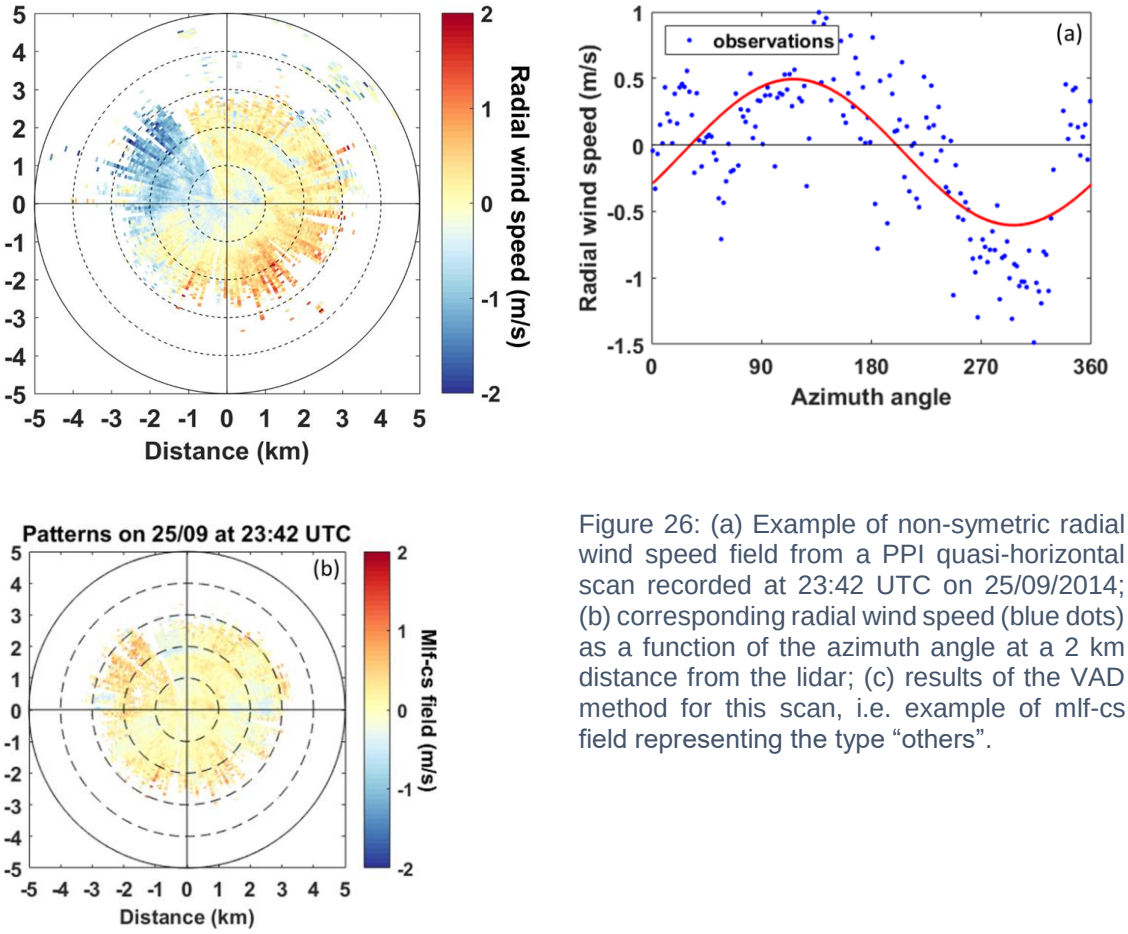


Figure 26: (a) Example of non-symetric radial wind speed field from a PPI quasi-horizontal scan recorded at 23:42 UTC on 25/09/2014; (b) corresponding radial wind speed (blue dots) as a function of the azimuth angle at a 2 km distance from the lidar; (c) results of the VAD method for this scan, i.e. example of mlf-cs field representing the type “others”.

2.2.2 Composition of the training ensemble

The repetition of the same patterns in the 2-month dataset gave birth to the idea of an automatic classification of the mlf-cs with regards to their patterns. An automatic classification requires an algorithm capable of distinguishing the different patterns. A training ensemble was built in order to test the performance of various algorithms and find the parameters capable of distinguishing the different patterns. The training ensemble consisted of 30 streak, 30 roll, 30 thermals and 60 “other” cases out of the 4577 total cases. The date and time of all the scans included in the training ensemble are displayed in Figure 27.

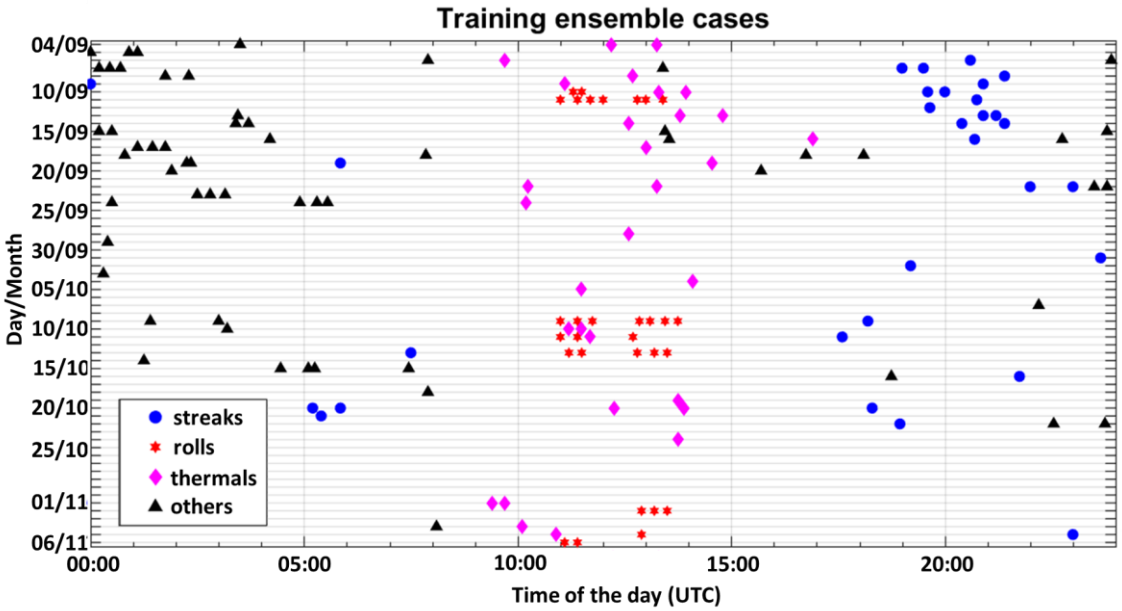


Figure 27: The date and time of each of the 150 cases of the training ensemble: streaks (blue circle), rolls (red star), thermals (purple rectangle) and others (black triangle).

The 30 streak cases were selected when narrow elongated patterns (periodicity sizes of the structures of few hundred meters) co-occurred with high wind shear (llj events) near the surface. In order to ensure that the narrow elongated patterns were mainly driven by the high wind shear and not by thermals, all 30 streaks cases were selected during the night, in particular, between 18:00 UTC and 22:00 UTC. After 22:00 UTC the wind field was usually weaker, leading to a non-symmetric radial wind field without coherent structures. Half of the streaks cases were selected from the first 15 days of the campaign with the remaining cases were distributed over the whole campaign.

Regarding the rolls, as only the presence of cloud streets can confirm the rolls formation, all 30 cases were either co-occurred with cloud streets recorded on MODIS Aqua or Terra satellite images, or they belonged to chronological series of scans presenting similar wide patterns (periodicity sizes of the structures larger than 1 km) among which a scan coincided with the observation of cloud streets. There were 8 days with cloud streets during the campaign, with 9 cases occurring on mid-September, 15 cases on mid-October and 6 cases at the beginning of September. All of theses cases were observed between 10:30 UTC and 14:00 UTC.

The 30 thermals cases were selected from 23 different days distributed over the whole campaign between 9:00 UTC and 15:00 UTC with one case around 17:00 UTC. The selection was based on the large enclosed patterns (homogeneous areas in the mlf-

cs field of approximately 1 km or above) occurring during periods with hourly solar radiation values much higher than the monthly average as displayed by the Photovoltaic Geographical Information System (Huld et al., 2012).

Finally, the type “others” was represented by 60 cases, twice more than each of the other types. The reason behind this choice is the sensitivity of the algorithms to the training ensemble. If the training ensemble is not balanced according to the expected outcome of the classification, it can perform poorly (Kubat, 2017). The whole ensemble was expected to contain a significant majority of the type “others” due to the many low wind conditions episodes during the period under study. Therefore, the representation of the type “others” with twice more cases was preferred. In order to represent the different type of “others”, for 53 out of the 60 cases, the radial wind fields were non-symmetric, and for 7 cases, the radial wind fields were symmetric but there were no defined patters. Among the 53 cases of non-symmetric radial wind fields, 46 cases occurred under calm or low wind conditions while the 7 cases without any defined pattern all occurred under moderate wind conditions (6-8 $m \cdot s^{-1}$). The cases were distributed over the two months with the majority being nocturnal and with 10 cases selected between 8:00 UTC and 16:00 UTC.

2.3 Classifying the training ensemble

After the selection of the training ensemble, it was important to identify the appropriate classifiers, i.e. the relevant parameters for the algorithms. Inspired by Srivastava et al., (2018) that were able to identify and classify elongated as well as enclosed patterns, the same parameters were selected for this classification. These are the four parameters presented in Section 1.3, namely the homogeneity, contrast, correlation and energy. For the estimation of these parameters, it is necessary to compute first the CM.

2.3.1 Pre-processing of the mlf-cs fields

The radial wind speed observations are in polar coordinates and therefore, in order to find the neighbouring pairs of pixels as explained in Section 1.3.1, it is necessary to convert the data in the cartesian coordinates. The CM can be then computed directly from the radial wind speed values, as long as they are divided into bins. The priority was to separate the positive and negative values of the radial wind speed, because they correspond to different direction of the wind fluctuations, thus different zones in the turbulent structure. To do this, the wind fluctuations values were divided into 8 bins with the first bin containing all the values higher than +0.5 $m \cdot s^{-1}$, the next six bins equally distributed between +0.5 $m \cdot s^{-1}$ and −0.5 $m \cdot s^{-1}$ and the final bin containing all values lower

than −0.5 m·s^{-1}. This bin selection provided a very good contrast between the positive and negative values, thus the structures were highlighted.

The idea was to classify the structures independently of their orientation. For example, streaks structures aligned with the northwest direction should be put in the same class as the streaks structures aligned with the south direction. In this regard, all the mlf-cs fields were rotated as if the wind had a north direction for all the cases. Figure 28 showcases an example mlf-cs with streaks patterns before and after the regriding, binning and rotation operations have been performed. It is obvious that the patterns have been enhanced and at least visually is easier to identify the patterns.

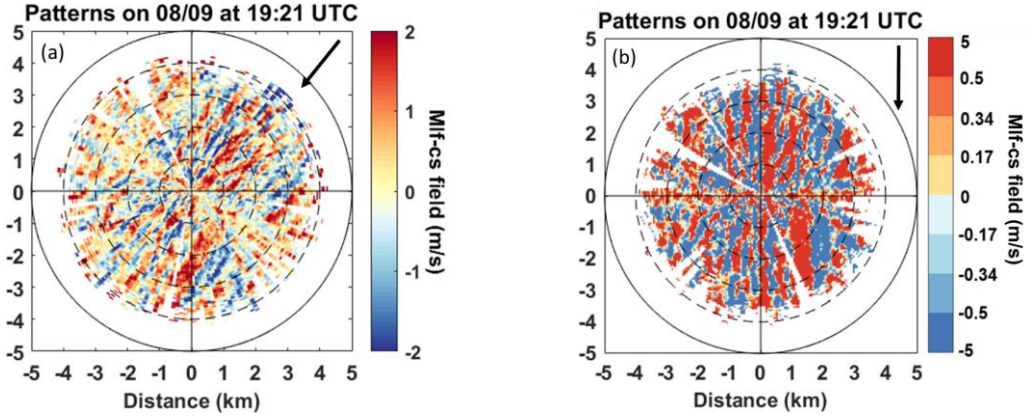


Figure 28: Streaks pattern recorded on 08/09 at 19:21 UTC: (a) observations in polar coordinates with continuous wind values and a north-east dominant wind and (b) the same after converting to Cartesian coordinates, rotating the field along the wind direction (black arrow) and separating the wind values into 8 bins.

2.3.2 Computing of the Co-occurrence Matrices

In order to show how the CM can highlight the patterns in the mlf-cs fields, this section will present two peculiar CMs computed from the mlf-cs field of Figure 28b. Similarly to the ideal case in Section 1.3.1, the first CM is computed for the adjacent “vertical” neighbours, i.e. aligned with the wind direction (Table 5). The most common pairs are the ones belonging in the 1-1 or the 8-8 bins combinations, i.e. the wind value is below −0.5 m/s in both pixels or above +0.5 m/s in both pixels. As the structures were aligned with the wind direction, which is now the north (vertical) direction, the most probable pair occurrence is for the ones that belong in the same bin, when the turbulent structure is well defined like on Figure 28, most wind values in the mlf-cs are not close to zero and fall into the bins 1 and 8.

Table 5: The CM for the adjacent vertical neighbours(50 m distance along the mean wind direction).

	1	***2***	***3***	***4***	***5***	***6***	***7***	***8***
1	3819	303	224	134	110	63	43	71
2	311	116	93	95	66	45	31	52
3	244	125	115	88	95	47	45	42
4	151	87	110	109	98	87	67	95
5	91	76	98	113	113	114	93	161
6	77	46	66	86	116	101	105	217
7	33	27	35	67	105	113	93	313
8	50	39	68	113	169	240	325	3882

On the contrary, in the CM computed for neighbour pairs located 4 grid cells away (200 m in real distance) in the horizontal direction (Table *6*), the number of pairs that belong to the 1-1 or 8-8 bins combination is equivalent to the ones of the 1-8 and 8-1 combinations. This means that in many pixel pairs, the wind is below −0.5 m/s for one and above +0.5 m/s for the other, i.e. pretty opposite values. This was expected as well, since the stripes alternate between positive and negative values in the transverse (now horizontal) direction and 200 m is the order of magnitude of a half-period on the mlf-cs field of Figure 28b.

Table 6: The CM for the horizontal neighbours four grids pixels away (200 m in real distance, perpendicularly to the mean wind direction).

	1	***2***	***3***	***4***	***5***	***6***	***7***	***8***
1	1584	261	254	220	257	230	241	1520
2	239	43	38	40	49	55	35	267
3	228	53	53	46	51	47	44	242
4	250	36	53	41	67	38	48	233
5	262	44	46	38	36	55	46	293
6	221	35	45	65	58	42	40	261
7	252	44	49	50	33	32	44	236
8	1547	251	227	272	272	262	240	1481

As the periodicities in the mlf-cs fields were not known *a priori*, the CMs were computed for all possible cell-pair orientation (angle φ from −90° to +90°, see Sec. 1.3.1) and for a neighbour spacing n from 1 (first neighbour) to 30, i.e. for real distances from 50 to 1500 m along the longitudinal/transverse relative to the mean wind.

2.3.3 Computing the texture parameters

The challenging part however, it is not to find differences in the CM for one mlf-cs field, but between the different types of patterns. For this purpose, the texture analysis parameters were computed. Each parameter was computed from each CM in the set then, based on Equations (9) to (12), the texture analysis parameters were calculated with every n giving multiple values for the different orientations φ. Therefore, it was possible to plot the curves of the texture analysis parameters for a specific n against the orientation φ. In this way we could examine possible difference in the curves for the different structures in the mlf-cs. Figure 29 showcases the curves of the homogeneity for $n = 3$ averaged for all the cases of streaks, rolls, thermals and 'others' of the training ensemble.

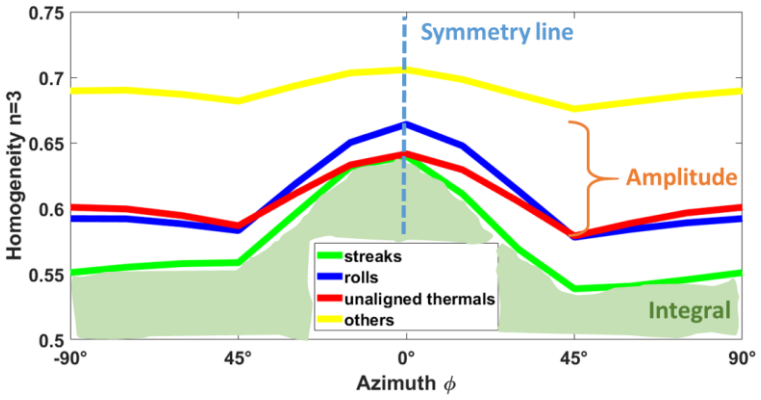


Figure 29: The homogeneity parameter for the third neighbour CMs ($n = 3$) plotted as a function of the cell pair orientation φ (±90° is horizontal, 0° is vertical) averaged for all the cases of the training ensemble for each type, streaks (green), rolls (blue), thermals (red) and others (yellow).

Three characteristics of the curves are highlighted in Figure 29, namely the amplitude, the symmetry line and the integral. The curves of the rolls and especially of the streaks cases have a prominent peak at 0° compared to the smoother curves for the thermals and the others types. The 0° orientation correspond to the vertical neighbours, which in coherent elongated structures aligned with the wind should eventually belong in the same bin, unless the structures are slightly tilted compared to the mean wind. This means that the amplitude (difference between the maximum and minimum points) of the curve should be different between the different types of structures. A perfectly elongated pattern should also be symmetric about the 0°, whereas for random structures like thermals and others that should not necessarily be the case. We also see in Figure 29 that the sum of points (integral) differs between the types and therefore this parameter was also computed. Analytically the expressions for the calculation of the amplitude, symmetry and integral were the following:

Amplitude: $Hom.Amp(n) = max_{\varphi}\big(Hom(\varphi,n)\big) - min_{\varphi}(Hom(\varphi,n))$ (13)

Symmetry: $Hom.Sym(n) = \sum_{\varphi} |Hom(\varphi,n) - Hom(-\varphi,n)|$ (14)

Integral: $Hom.Int(n) = \sum_{\varphi} Hom(\varphi,n)$ (15)

By calculating these so called “curve parameters” for all neighbour spacing n and for all 150 cases in the training ensemble, it is possible to examine whether the cases can be clustered based on the texture analysis curve parameters. In Figure 30, the scatter plot of the 150 cases of the training ensemble is displayed, based on the three first texture

analysis parameters selected by the automated classification algorithm among the 360 possible (see next section).

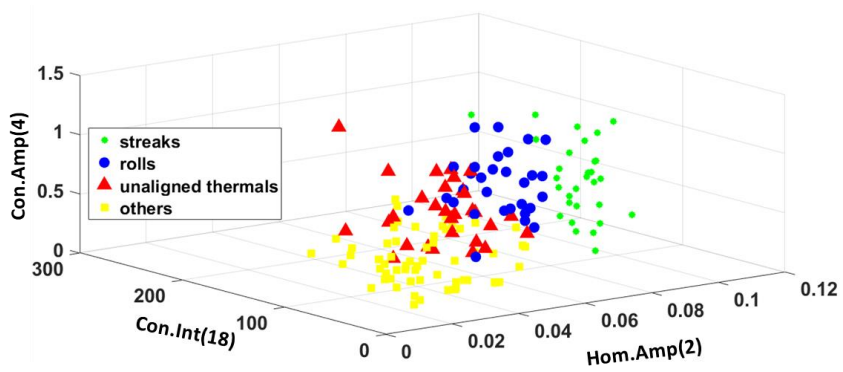


Figure 30: Scatter plot of the 150 cases in the training ensemble, based on three of the texture analysis curve parameters: amplitude of the 4th neighbour contrast curve (Eq. 13 for n=4 but the contrast instead of the homogeneity), integral of the 18th neighbour contrast curve (Eq. 15 for n=18 but for the contrast), Amplitude of the second neighbour homogeneity curve (Eq. 13 with n=2).

As we can see in Figure 30, the streak cases (green) are grouped and their cloud of points is well separated from the other types. The amplitudes of the homogeneity and contrast parameters are important for the elongated patterns, and this is confirmed by this scatter plot as well. The same, to a lesser extent, is true for the rolls cases which are also grouped for the most part, though there is some mixing with the “thermals” and “others” types. The thermal type is the most spread group, as it is mostly mixed with the types “rolls” and “others”. Finally, a large part of the “other” type is well distinguishable with only a small part being mixed with the types “thermals” mainly and “rolls” to some extent. These preliminary results indicated that it could be possible to classify the structures by utilizing an automated clustering algorithm

For the training ensemble the 3 curve parameters for the 4 texture analysis parameters were computed for 30 neighbour order n, thus bringing the number of parameters to 360. In addition to the 360 texture analysis parameters, three physical parameters were included in the classifiers. These were the average mean wind speed as estimated by the VAD method, the time in UTC hour, which is close to solar time in Paris and the root-mean-square error (RMSE) of the cosine fit. The RMSE is given by the expression:

$$RMSE = \sqrt{\frac{\sum_i^N (u_r + f(\theta))^2}{N}} \qquad (16)$$

where N is the number of non-missing observations. So the total number of classifiers was 363.

2.3.4 Evaluating the classification performance on the training ensemble

The classifiers were provided to the QDA algorithm following the greedy algorithm of stepwise forward selection (Sokolov et al., 2020) and the classification error was estimated based on the 10-fold cross validation. The principle of this validation process is that the algorithm is trained using only 90% of the training ensemble while the remaining 10 % (15 cases) are used to test the result of the classification, i.e. the outcome of the classification is compared to the expected class in order to estimate the classification error. This process is repeated 10 times to test the entire 150 cases in the training ensemble. In the process, the algorithm selects the classifiers that minimize the error. For the training ensemble of this study, the selected classifiers are presented in Figure 31. The error was slightly above 9 %.

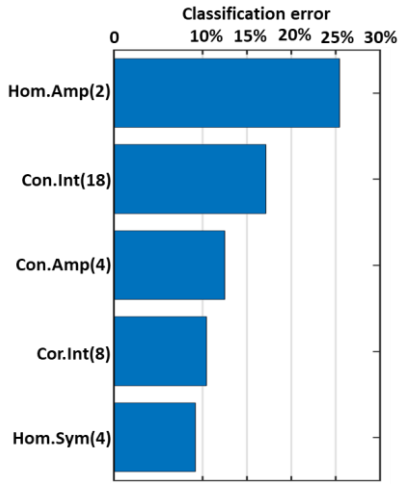


Figure 31: Classifiers that minimize the classification error of the training ensemble by using the QDA method. From top to bottom: Amplitude of the homogeneity curve for the 2nd neighbour, integral of the contrast curve for the 18th neighbour, amplitude of the contrast curve for the 4th neighbour, integral of the correlation curve of the 8th neighbour and symmetry of the homogeneity curve for the 2nd neighbour. The classification error decreases as the number of classifiers taken into account increases; it reaches a minimum of ~9% for 5 curve parameters. Adding more classifiers (not shown) makes the error increase again.

The classifiers selected by the QDA method as they minimize the error cover various distances, from short distance as the 2nd -neighbour is equivalent to 100 m, to medium range distance as the 8th neighbour is equivalent to 400 m to large distance as the 18th neighbour is equivalent to 900 m. This is particularly important for the distinction of large patterns compared to small. It is also confirmed that the amplitude of the

homogeneity and contrast curves are significant classifiers for the structures, meaning that the prominent peaks of the curves indeed differentiate the types.

The integrals of the contrast and correlation curves were also selected by the algorithm. Indeed, the “others” type is characterized by chaotic patterns while the “thermals” type is characterized by large homogeneous areas, which results in differences in the integrals of the curves for parameters such as the contrast and correlation, especially for medium (8th neighbour) and large distances (18th neighbour). It is noteworthy, that the symmetry of the homogeneity curve was selected by the algorithm. This confirms the importance to align all the mlf-cs fields in the same direction. One of the most striking results, however is that none of the three physical parameters was selected by the algorithm despite the fact that the time of day could be expected to discriminate between nocturnal (streaks) and diurnal (thermals, rolls) structures.

The analytical performance of the QDA algorithm for each type of structures is demonstrated in Figure 32 in the form of a confusion matrix.

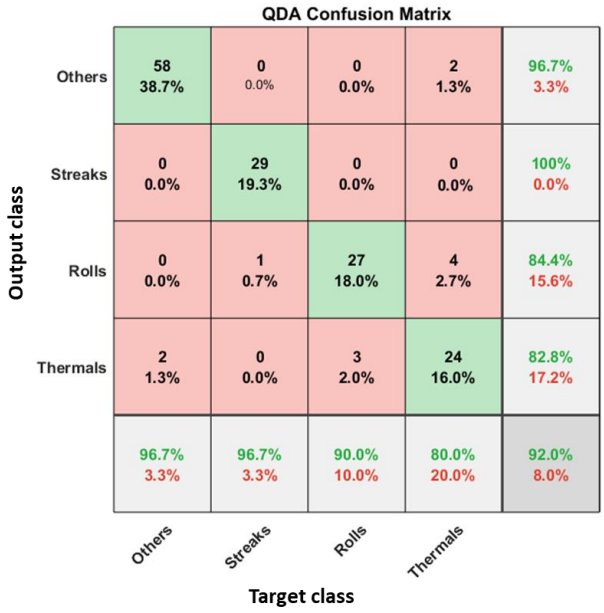


Figure 32: The confusion matrix of the training ensemble showcasing the classification error of the QDA algorithm for each type of structures. Target class = eye-made classification, output class = classification by the algorithm. For instance, the cell at the intersection of the “thermals” line and “rolls” column gives the number of roll cases classified as thermals by the algorithm.

The QDA algorithm performed the best for the classification of streaks structures. It classified accurately 29 out of the 30 streaks cases in the training ensemble, with only one misclassification as rolls. The second best performance was for the category “others” with 58 out of the 60 cases classified correctly, and only two cases misclassified as thermals.

The performance of the algorithm for the classification of rolls cases was also good with 27 cases classified correctly and 3 cases misclassified as thermals. The worst performance of the algorithm out of all the types was for the thermals. Nonetheless the algorithm still performed rather good with 24 cases classified correctly, 4 cases misclassified as rolls and 2 cases misclassified as ”others”.

As the classification of the whole ensemble of 4577 scans was carried out based on the five classifiers selected by the algorithm for the training ensemble, some misclassifications between thermals, rolls and “others” were expected to occur.

2.4 ABL height estimation

The ABL height is an important parameter with regards to the structures development. Rolls and large thermals occur during a developing or a well developed ABL. When the ABL is shallow, it is not physically possible for such large structures to occur. On the contrary streaks can form during the night (shallow ABL) under high wind shear conditions near the surface. The ABL height cannot be included as a parameter in the classification process because it cannot be estimated reliably all the time (for instance, during the afternoon transition the aerosol lidar profile often ambiguous) and because the process is more robust if relying on a single instrument, i.e. the Doppler lidar. Instead, the estimation of the ABL height can be used as a measure of validity for the classification of the whole ensemble, if the ABL height values are associated with the types of structures for the corresponding times. It is important to note that precision is not essential, an approximate value is sufficient for the purpose of this study.

For the estimation of the ABL height during two days (09/09 and 10/09) of the VEGILOT campaign, Klein et al., (2019) used the vertical LOS beams from the Doppler lidar observations during the night and the aerosol lidar observations during the day. This methodology for the estimation of the ABL height was extended to all the days of the 2-month dataset for this study. During the night, the methodology relies on the vertical wind standard deviation computed from the multiple LOS scans recorded during 30 seconds periods. On an example profile of the vertical wind standard deviation (Figure 33), it is apparent that, near the surface, there is a significant decline above a certain height. This is an indication of passing from the ABL to the free troposphere (Tucker et al., 2009). The ABL height was estimated at the maximum of the vertical wind standard deviation.

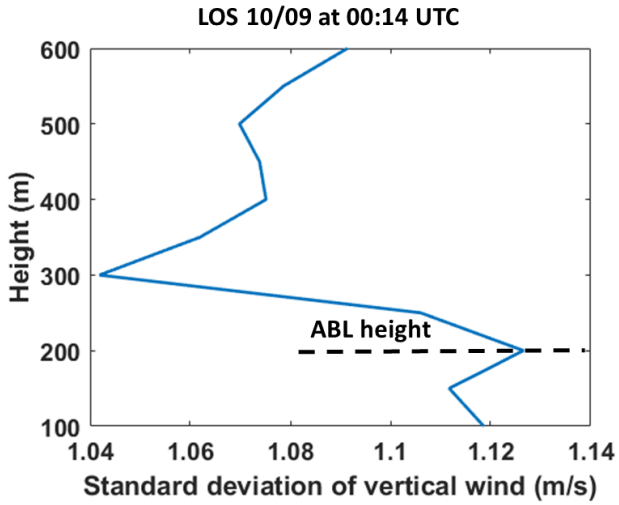


Figure 33: Vertical profile of the vertical wind standard deviation based on the LOS vertical shots recorded on 10/09 at 00:14 UTC.

This method works during the night, because aerosols in the residual layer provide a sufficient level of signal for the Doppler lidar to retrieve the wind above the ABL. During the day, there are most of the time no wind measurement above the ABL because the free troposphere is too clean. Therefore, the vertical wind variance profile is incomplete and cannot be used to estimate the ABL height.

Fortunately, there were also aerosol lidar observations. The aerosols are located within the ABL, thus there is a notable decrease in the concentrations at the transition between the ABL and the free troposphere (Stull, 1988). Flamant et al., (1997) and Menut et al., (1999) have showcased the estimation of the ABL height based on the maximum of the variance of the range corrected signal (Pr^2) recorded by the aerosol lidar. In the dataset of VEGILOT, the minimum of the first derivative of the Pr^2 standard deviation was utilized for the estimation of the ABL height, as it gave better results compared to the maximum variance of the Pr^2. It is noteworthy that the aerosol observations cannot be utilized for the estimation of the ABL height during the night due to the blind zone of the lidar, which is equivalent to 200 m. Figure 34 showcases an example of the vertical profiles of the 10-minute average and the standard deviation of the Pr^2 along with the estimated ABL height.

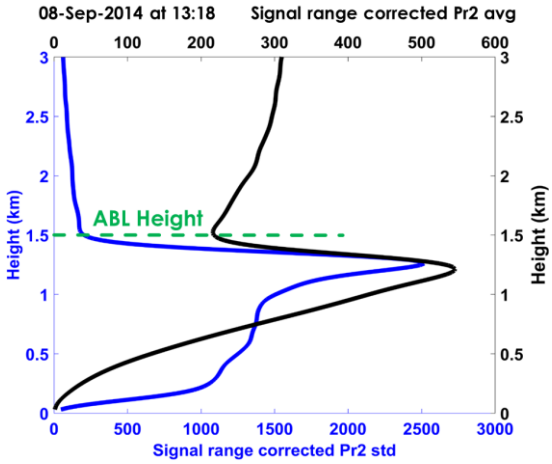


Figure 34: Vertical profiles of the average (black) and the standard deviation (blue) of the Pr^2 recorded by the aerosol lidar on 08/09 at 13:18 UTC.

An example of the results of the implemented methodology for the estimation of the ABL height is portrayed in Figure 35. The ABL height varied between 200 and 300 m during the night and grew up to 1300 m in the afternoon. The development of the ABL started a few hours after sunshine, around 8:00 UTC, and was very fast until around 11:00 UTC. Similar realistic ABL height values were observed for the others days under study and correspond to previous observations on this site (Dieudonné, 2012).

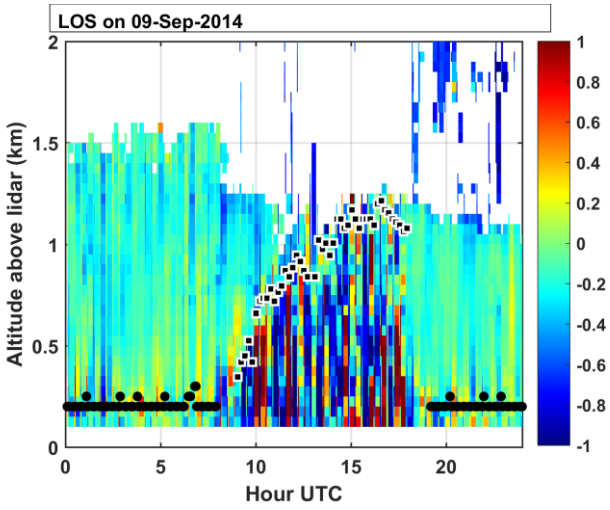


Figure 35: Time-height cross-section of the vertical wind speed observed by the Doppler lidar using the vertical LOS on 10/09 along with the ABL height as obtained from the abovementioned methodology (black dots = wind lidar, black squares with white edges = aerosol lidar).

This methodology for the estimation of the ABL height depends on clear sky conditions. Low level clouds, fog, precipitation and any other weather phenomena or events that disturbs the signal of the lidars makes it impossible to estimate the ABL height.

The ABL height was estimated accurately for the majority of the 24 hours period for 18 out of the 64 days of the campaign (the exact dates are given in Figure 36). Furthermore, the occurrence of nocturnal llj events can complicate things. According to Stull (1988), during a llj event the ABL height is estimated “under the nose of the llj”. However, the strong wind shear can generate turbulence and eventually increase the height of the ABL (Blackadar, 1957). For the current study the ABL height was only estimated via the vertical wind standard deviation values during the night, even when llj events occurred, because the precision in the estimation of the ABL height was not the priority for the purpose of this study.

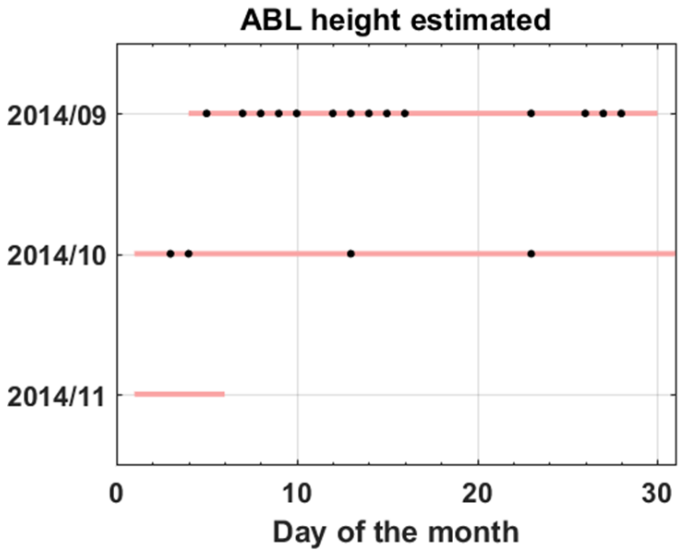


Figure 36: Days during VEGILOT that the ABL height could be estimated for the majority of the day.

2.5 Structure size retrieval

Regarding the validity of the classification, another interesting parameter is the horizontal size of the structures for the streaks and rolls cases. The patterns of these coherent structures can be observed in the mlf-cs fields, as well as the periodicities of the alternating stripes of positive and negative values. By applying the Fast Fourier Transform (FFT) analysis in the radial wind speed observations, one will obtain the spectrum of the data (Stull, 1988). Several studies have shown that the spectrum of the radial wind speed in the direction perpendicular (transverse) to the mean wind is characterized by distinct peaks corresponding to the periodicities of the structures (e. g. Calaf et al., 2013; Drobinski et al., 2004, 1998).

The two opposite beams of the quasi-horizontal PPI scan that are transverse to the mean wind direction were stacked together before computing the FFT in order to cover a

larger domain of space, thus frequency. Those two beams were recorded with a 1,5-minute difference, so it is necessary to suppose that the structures keep the same properties as they pass over the lidar, i.e. they are similar in the longitudinal direction over a distance of an order of magnitude between 200 and 700 m (90 s multiplied by the mean wind speed, which is around 2 to 8 m/s). This distance is of the same order of magnitude as the periodicities that are visible in the transverse direction, so it is reasonable to suppose that elongated structures do not vary over such a distance in the longitudinal direction.

Using a single pair of beams produces a very noisy spectrum, so the spectrums of the five pairs of beams closest to the transverse direction were computed, then averaged. Before applying the FFT, the radial wind field was converted from polar to Cartesian coordinates and interpolated on a regular grid so that all the grid points have an equal 50-m spacing in the transverse direction. Figure 37 present an example of streaks case mlf-cs field and the corresponding spectrum. There is a quite prominent peak in the spectrum a little above 400 m. This size is in agreement with the periodicity that can be observed on the mlf-cs field.

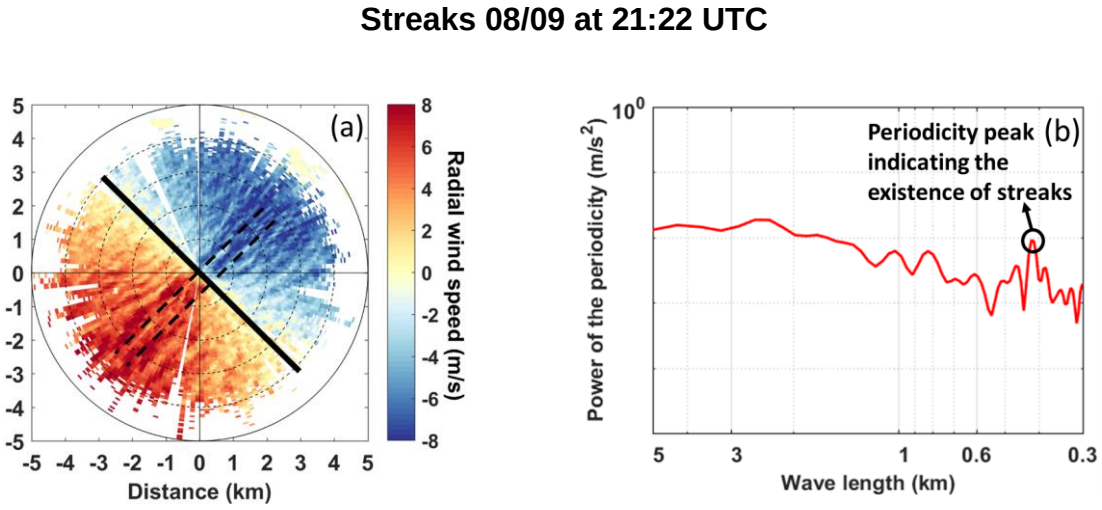


Figure 37: (a) Example of radial wind field for a streaks case recorded on 08/09 at 21:22 UTC, with the transverse direction (black line). The dashed line encloses an area covering a single period in the transverse direction. (b) Corresponding averaged spectrum for the five central lines in the transverse direction.

Figure 38 is similar to Figure 37 but for a roll case; here the peak for the periodicities is located at approximately 1300 m, again in agreement with the periodicities visible on the mlf-cs field.

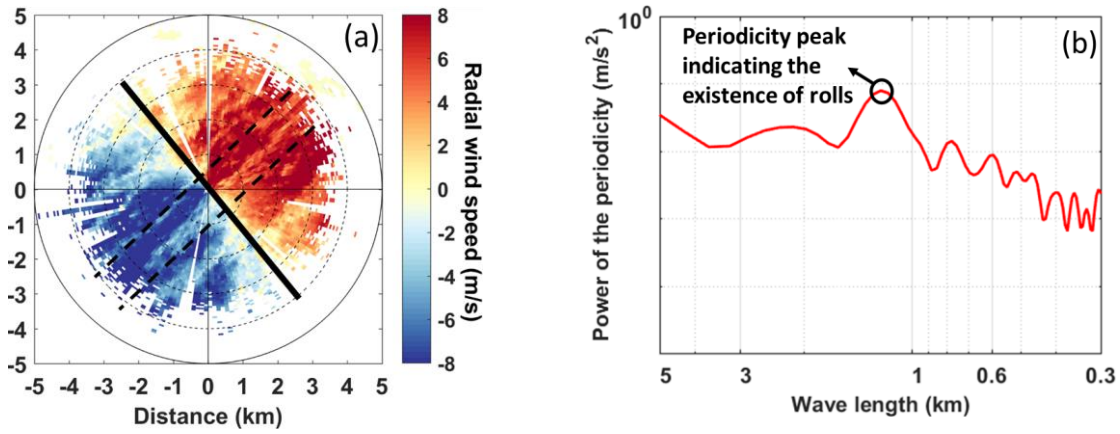


Figure 38: (a) Radial wind field and (b) the corresponding averaged spectrum for the five central lines in the transverse to the wind direction for a rolls case on 09/10 at 13:24 UTC.

The spectrums are limited to 300 m on the short wavelength side because it was not realistic to observe smaller structures when the spatial resolution of the lidar was 50 m. For the rolls cases, we observe the sizes near the surface while other studies, especially those based on airborne data, observe the rolls in the middle of the mixed layer. Therefore, it is complicated to compare the values obtained here with other studies. Despite that, the spectrums clearly showed different sizes for the rolls and the streaks. Therefore, the horizontal sizes could be a parameter to validate the classification. The horizontal sizes were estimated only for the coherent structures as the random patterns of the thermals do not produce clear peaks in the spectrums.

2.6 Implementation of the methodology

Chapter 2 presented a detailed description of the implemented methodology along with all the important means to perform this study. The performance of the methodology is displayed in the following chapter in the form of two manuscripts accompanied by introductory and discussion sections. Section 3.1.2 contains the manuscript with the performance of the algorithm in the training ensemble (published in the peer-reviewed journal of *Atmospheric Measurement Techniques*). The second manuscript (Sec. 3.2.2 – under review at the Journal of Applied Meteorology and Climatology at the time of submission) presents the results of the classification for the 2-month VEGILOT dataset and the physical properties of the classified structures.

3. Results

3.1 Automated classification of the coherent structures

3.1.1 The idea for an automated classification method

The main challenge of this study was to discover a way to exploit the information from the two-month dataset consisting of 4577 quasi-horizontal scans. The careful visual examination of the mlf-cs fields led to the identification of distinct patterns that occurred at multiple times and dates. This observation generated the idea of an automatic method for the classification of the structures based on the patterns. An idea that came to life with the combination of texture analysis parameters and a supervised machine learning algorithm. The development of the methodology have been presented thoroughly in Sections 2.2 and 2.3. A more compact version, along with the results of the classification for the 4577 cases has been published in the peer-reviewed journal *Atmospheric Measurement Techniques* and it is presented in Section 3.1.2. This article also includes the diurnal variation of the structures types found for the two-month period. It is apparent that the classified structures show a realistic behaviour with regards to the time of the day they occur.

In this article, only the algorithm set up that with the best performance for the final classification was presented. However, various different set ups were tested throughout this study. In Section 3.1.3 the alternative set ups and their results are discussed in order to give an insight to interested researchers that would like to use the same methodology. The classification depends exclusively on the texture analysis parameters of the mlf-cs fields. This is significant because the methodology can be applied to any dataset independently of the place and time of the wind lidar observations as long as the mlf-cs fields contain similar patterns.

3.1.2 The automated classification method for the detection of coherent structures

Atmos. Meas. Tech., 13, 6579–6592, 2020
https://doi.org/10.5194/amt-13-6579-2020

Atmospheric Measurement Techniques
Open Access
EGU

Detecting turbulent structures on single Doppler lidar large datasets: an automated classification method for horizontal scans

Ioannis Cheliotis[1], Elsa Dieudonné[1], Hervé Delbarre[1], Anton Sokolov[1], Egor Dmitriev[2], Patrick Augustin[1], and Marc Fourmentin[1]

[1]Laboratoire de Physico-Chimie de l'Atmosphère (LPCA), UR 4493, Université du Littoral Côte d'Opale (ULCO), Dunkirk, France
[2]Marchuk Institute of Numerical Mathematics, Russian Academy of Sciences, Moscow, Russia

Correspondence: Ioannis Cheliotis (ioannis.cheliotis@yahoo.gr)

Received: 10 March 2020 – Discussion started: 30 April 2020
Revised: 13 October 2020 – Accepted: 25 October 2020 – Published: 7 December 2020

Abstract. Medium-to-large fluctuations and coherent structures (mlf-cs's) can be observed using horizontal scans from single Doppler lidar or radar systems. Despite the ability to detect the structures visually on the images, this method would be time-consuming on large datasets, thus limiting the possibilities to perform studies of the structures properties over more than a few days. In order to overcome this problem, an automated classification method was developed, based on the observations recorded by a scanning Doppler lidar (Leosphere WLS100) installed atop a 75 m tower in Paris's city centre (France) during a 2-month campaign (September–October 2014). The mlf-cs's of the radial wind speed are estimated using the velocity–azimuth display method over 4577 quasi-horizontal scans. Three structure types were identified by visual examination of the wind fields: unaligned thermals, rolls and streaks. A learning ensemble of 150 mlf-cs patterns was classified manually relying on in situ and satellite data. The differences between the three types of structures were highlighted by enhancing the contrast of the images and computing four texture parameters (correlation, contrast, homogeneity and energy) that were provided to the supervised machine-learning algorithm, namely the quadratic discriminant analysis. The algorithm was able to classify successfully about 91 % of the cases based solely on the texture analysis parameters. The algorithm performed best for the streak structures with a classification error equivalent to 3.3 %. The trained algorithm applied to the whole scan ensemble detected structures on 54 % of the scans, among which 34 % were coherent structures (rolls and streaks).

1 Introduction

Turbulent flows are motions characterized by high unpredictability. Nevertheless, coherent structures are developed in these flows (Tur and Levich, 1992). The principal aspect that determines a coherent structure is the maintenance of the phase-averaged vorticity of the turbulent fluid mass over the spatial extent of the flow structure (Hussain, 1983). The most typical types of coherent structures are presented in the review of Young et al. (2002), who classified structures into three characteristic types: turbulent streaks, convective rolls and gravity waves. Several studies have been carried out to examine the effect of the coherent turbulent structures in the dispersion of pollutants by utilizing boundary layer simulations. The results of these studies indicate that the coherent structures can play a significant role in the pollutants' concentrations (Aouizerats et al., 2011; Soldati, 2005). Furthermore, Sandeepan et al. (2013) have demonstrated via simulations that the pollutants' concentrations can alternate from low to high during coherent-structure events. It is therefore important to be able to identify structures in the atmosphere and observe them in an efficient and consistent way. The term coherent structures in the aforementioned studies refers exclusively in the atmospheric flow, and it is the main focus in this study. This term is also encountered in studies at the laboratory scale described as hairpins or packets (Adrian, 2007; Hutchins and Marusic, 2007), but these are out of the scope of this study.

Turbulent streaks are structures aligned with the horizontal wind with alternating stripes of stronger horizontal wind associated with a subsidence and stripes of weaker horizontal

Published by Copernicus Publications on behalf of the European Geosciences Union.

wind associated with an ascendance (Khanna and Brasseur, 1998). The high wind shear between the surface layer and the lower planetary boundary layer (PBL) can lead to the formation of the turbulent streaks in the surface layer that may extend to the mixed layer. Neutral or near-neutral stratification favours the formation of streaks, though they may also form during stable and unstable conditions (Khanna and Brasseur, 1998). The physics behind their formation differs as the contribution of buoyancy varies in relation to the atmospheric conditions (Moeng and Sullivan, 1994). Formation, evolution and decay of streaks are rather short, equivalent to several tens of minutes, before they regenerate. The average streak spacing is usually hundreds of metres (Drobinski and Foster, 2003). In the mixed layer, horizontal roll vortices, also known as convective rolls, develop roughly aligned with the mean wind (LeMone, 1972). Favourable conditions for the development and maintenance of convective rolls are the spatial variations of surface-layer heat flux, the low-level wind shear and the relatively homogeneous surface characteristics (Weckwerth and Parsons, 2006). As the rolls rotate in the vertical plane, they generate ascending and descending motions. These motions under convective conditions can form clouds in rows separated by clear-sky areas known as cloud streets, which is a characteristic visual feature used to identify rolls (Lohou et al., 1998). The rolls usually extend from the surface to the capping inversion with a large variety of horizontal sizes from a few kilometres to few tens of kilometres. They are characterized by a long lifespan of hours or even days as opposed to the short lifespan of the streaks (Drobinski and Foster, 2003). Young et al. (2002) distinguish rolls in narrow mixed-layer rolls, where the ascending air masses are one thermal wide (Weckwerth et al., 1999), and wide mixed-layer rolls, where multiple thermals are grouped within each ascending area (Brümmer, 1999). As Young et al. (2002) stated, both types of rolls can be distinguished visually, with the narrow rolls having the form of a "string of pearls", whereas the wide rolls look like a "band of froth".

Remote sensors are exceptionally useful for the identification of coherent structures. Their ability to scan large areas in a short period is advantageous compared to in situ measurements (Kunkel et al., 1980). Lhermitte (1962) and Browning and Wexler (1968) were the first to implement the velocity–azimuth display (VAD) technique, also known as the plan position indicator (PPI) method, using Doppler radars. The PPI technique provides conical scans or even horizontal surface scans with the appropriate combination of elevation and azimuth angles. Kropfli and Kohn in 1978 were able to study horizontal roll structures by using a dual-Doppler radar in order to observe the wind field in the three dimensions. Several studies followed for different types of radars with more efficient configurations (Kelly, 1982; Lohou et al., 1998; Reinking et al., 1981). Weckwerth et al. (1999) were able to study the evolution of horizontal convective rolls by combining Doppler radar observations with meteorological measurements, radiosondes, flight measurements and satellite images. In recent years, various studies have been carried out by using lidars only. It has been well established that the PPI method can also be applied to Doppler lidars (Cariou et al., 2007; Vasiljević et al., 2016) with the possibility to compute the mean wind profile by using a modified version of the VAD method as has been demonstrated in the studies of Banta et al. (2002) and Chai et al. (2004). Depending on the selected scanning method of the Doppler lidar, it is possible to observe coherent structures in the atmospheric surface layer (Drobinski et al., 2004) as well as in the mixed layer (Drobinski et al., 1998). Newsom et al. (2008) and Iwai et al. (2008) introduced the dual-Doppler lidar method and revealed its benefits in the observation of coherent structures. This method was further improved by Träumner et al. (2015) using an optimized dual-Doppler technique. They were able to identify different type of structures including elongated areas resembling turbulent streaks. They combined quantitative characteristics of the coherence such as the integral scales and the anisotropy coefficients, obtained by a two-dimensional autocorrelation algorithm, with the visual observation of the scans. However, the subjective classification by observing the images is a time-consuming approach and non-systematic. Furthermore, the use of two Doppler lidars is limited to the institutes that can afford such a high cost and collaborations on short-term campaigns. A much less expensive approach, and suitable for long periods, is to detect the passage of the structures on sonic anemometer time series. For instance, Barthlott et al. (2007) analysed 10 months of data from a meteorological tower located in the surface layer 20 km south of Paris, France, and they observed coherent structures for 36 % of the cases. However, their study is limited to point measurements instead of a larger wind field that it is possible to observe via a lidar.

This study aims to identify the medium-to-large fluctuations and coherent structures (mlf-cs's) on single Doppler lidar horizontal scans and develop an automatic classification process based on the combination of texture analysis and a supervised machine-learning technique, namely the quadratic discriminant analysis (QDA), in order to handle large datasets. Texture analysis is an effective way to evaluate the distribution of the values within an image (Castellano et al., 2004). It is widely used in various scientific fields in order to classify images, covering meteorology (Alparone et al., 1990), medical studies (Holli et al., 2010) and forestry (Kayitakire et al., 2006). There is a lack of long-term studies of structures based on lidar observations, and the aforementioned automatic classification process can stimulate the interest in this research field. More particularly, it could facilitate the statistical analysis of the physical parameters of the structures, e.g. the structure size as a function of the height of the planetary boundary layer (PBL). Furthermore, it will enable us to study the transitions between structures and how these are associated with the atmospheric conditions. Finally, the impact of the structures on pollutants' concentrations could be examined for long-term studies under stable

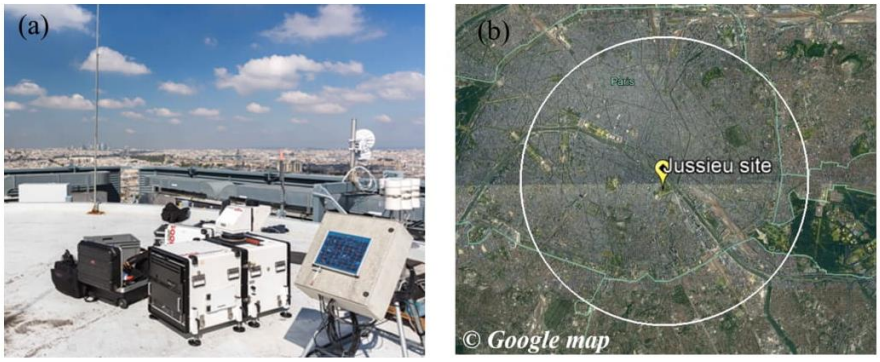


Figure 1. (a) The Doppler lidar installed on the tower roof during the VEGILOT campaign and **(b)** the measurement site in Paris with inside a circle with a 10 km diameter demonstrating the maximum range of the PPI surface scan (© Google Earth satellite image).

Table 1. Properties of the lidar used for the observation of mlf-cs's.

Doppler lidar (Leosphere WLS100)	
Altitude of lidar	75 m a.g.l.
Minimum range	100 m
Radial-wind-speed range	-30 to 30 m s^{-1}
Laser wavelength	1.543 µm
Radial wind accuracy	± 0.1 m s^{-1}
Accumulation time	1 s beam^{-1}

and unstable conditions. The classification method relies on the observations of radial wind speed recorded using a scanning Doppler lidar settled atop a 75 m high tower in the centre of Paris, during a 2-month period in late summer and early fall. Section 2 presents the experimental setup of the study. The methodology for the identification and classification of the mlf-cs's is demonstrated in Sect. 3. Subsequently, the results of the classification for the training ensemble as well as for the whole dataset are displayed in Sect. 4. Finally, the key points of the paper are summarized in Sect. 5.

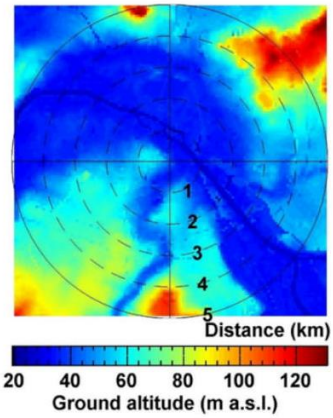


Figure 2. Ground altitude map above sea level with 75 m spatial resolution for the scanning area in Paris (credit: Institut National de l'Information Géographique et Forestière, https://www.data.gouv.fr/fr/datasets/bd-alti-r-75-m-250-m-1-000-m/, last access: 2 December 2020).

2 Experimental setup

A 2-month measurement campaign (4 September–6 November 2014) was carried out in order to study the exchange processes of ozone and aerosols in the area in the framework of the VEGILOT (VEGétation et ILOT de chaleur urbain; vegetation and urban heat island) project in the urban area of Paris (Klein et al., 2019). The Leosphere WLS100 Doppler lidar (https://www.leosphere.com, last access: 2 December 2020) with a minimum range of observations at 100 m (Fig. 1a) was installed atop a 75 m building on the Jussieu campus, located in Paris's city centre (Fig. 1b), and was used for wind measurements. Table 1 shows the significant lidar properties during the VEGILOT campaign.

The Doppler shift frequency between the emitted laser beam and the light backscattered by the aerosols is measured by heterodyne detection associated with fast Fourier transform as explained analytically by Cariou et al. (2007). A wind lidar measures the radial wind speed, i.e. the wind projection along the light beam (counted positive when going away from the lidar). Table 2 showcases the implemented scanning methods during the VEGILOT campaign. For the classification of the mlf-cs's, we focused in the current study on the almost horizontal PPI scans (1° elevation angle). During those scans, the lidar emitted beams in azimuth angles from 0 to 360° with a 2° resolution. This scenario was repeated every 18 min hence providing 4577 PPI scans during the whole campaign. The duration of each scan was 3 min, which is sufficiently fast for the observation of coher-

Table 2. Scanning methods selected during VEGILOT.

	Scanning area	Purpose	Elevation and azimuth angle	Scan duration
Plan position indicator (PPI)	Almost horizontal scans near surface	Identification of structures	Elevation 1°, azimuth 0 to 360° with 2° resolution	3 min
Doppler beam swinging (DBS)	Combination of line-of-sight beams	Identification of low-level jet cases	Elevation 75°, azimuth 0, 90, 180 and 270°	2×15 s

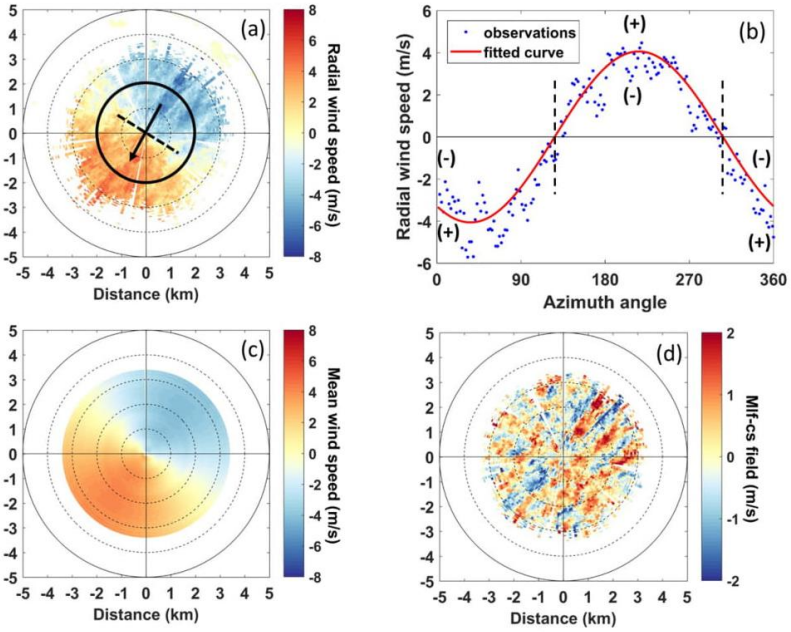


Figure 3. Observations recorded during a quasi-horizontal PPI scan on 8 September 2014 at the Jussieu site, Paris, at 09:26 UTC. **(a)** Radial wind speed along with the mean wind direction (black line) and the transverse direction perpendicular to it (black dotted line). **(b)** Radial wind speed (blue dots) as a function of the azimuth angle at a fixed 2 km distance from the lidar (black circle on **a**) along with the cosine fit function (red line). **(c)** Mean wind speed projected on the beam direction. **(d)** The mlf-cs field.

ent structures with a lifespan of several minutes. The maximum range of the scans reached 5 km (see white circle of Fig. 1b) with a spatial resolution of 50 m. It is noteworthy that the scanning area covers almost exclusively the urban area of Paris, a city famous for regulating the height of the buildings to not exceed 50 m in its centre (Saint-Pierre et al., 2010). The ground altitude enclosed by the scanning area mostly ranges between 30 and 60 m with the exception of some hills near the boundaries of the scanning range as can be seen in Fig. 2. It is fundamental for this study to assume that the wind field within the scanning area is homogeneous (see Sect. 3.1). Due to the 1° elevation, the beam was risen by about 87 m between the central point and the point at 5 km. It was also important for this study to retrieve observations regarding the vertical wind shear. For this purpose, the Doppler beam-swinging (DBS) scanning method was implemented. This method consisted of four line-of-sight beams at azimuth angles of 0, 90, 180 and 270° with an elevation angle of 75°, and it was applied twice. The duration of the four beams emission was approximately 15 s.

3 Preparation of the dataset for the classification

3.1 Turbulent radial wind fields

Assuming a homogeneous wind field for horizontal PPI scans, the radial-wind measurements u_r taken for the different beams at a given distance from the lidar should follow a

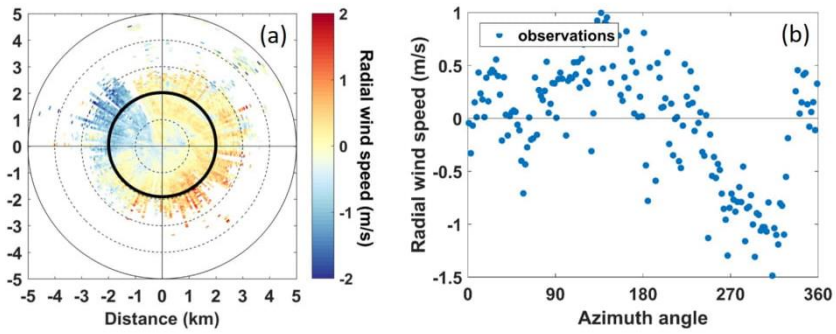


Figure 4. A case when the VAD method cannot be applied: **(a)** radial wind field on 25 September 2014 at 23:42 UTC and **(b)** radial wind speed (blue dots) as a function of the azimuth angle at a fixed 2 km distance from the lidar (black circle on **a**).

cosine function of the azimuth angle, due to the projection of the wind along the beam direction (Eberhard et al., 1989). For instance, the observations at 2 km from the lidar (black ring in Fig. 3a) are displayed in Fig. 3b and can be fitted by a cosine function in the form of Eq. (1):

$$u_r = a + b\cos(\theta - \theta_{max}), \quad (1)$$

where b is the mean wind speed, θ_{max} is the wind direction, θ is the azimuth angle of the beam and a is the offset (Browning and Wexler, 1968; Lhermitte, 1962). It is noteworthy that the value of a is much smaller than b for our data. It is possible to retrieve the mean wind from all the "rings" and subsequently calculate the mean wind projected on the beam direction which is displayed in Fig. 3c. The difference between the radial wind field u_r (Fig. 3a) and the mean wind projected on the beam direction (Fig. 3c) is the mlf-cs of the radial wind field u'_r (Fig. 3d), a parameter that indicates the existence of a turbulent atmosphere. For this study, the radial-wind-speed values for which the carrier-to-noise ratio is lower than -27 dB (CNR < -27 dB) were disregarded, since they were anomalously high, exceeding the values of the rest of the radial wind field by 2 times or more. Therefore the effective scanning range showcased in Fig. 3 is approximately 3 km. For a better visual representation of the patterns, the sign of the u'_r in the current study is positive when the radial wind speed is stronger than the mean wind speed and negative when it is weaker as is illustrated in the sign convention of Fig. 3b, and it was computed by the following expression:

$$u'_r = |u_r(\theta)| - |f(\theta)|, \quad (2)$$

where f is the fitted curve.

The Jussieu site is located in an urban area near hills; hence the surface roughness or the orography can affect the regional wind flow. Troude et al. (2002) and Lemonsu and Masson (2002) have performed numerical weather simulations in the area of Paris and have observed that during low-wind conditions (below 3 m s^{-1}) the orographic effect and the urban heat island effect could be the main drivers for the local wind speed. As a result, in some cases the radial wind field does not follow a cosine function, and therefore the VAD method cannot be applied. This is apparent especially at night when low winds (below 2 m s^{-1}) do not have a defined direction (Wilson et al., 1976). Figure 4 presents a case where the radial wind field is not homogeneous. The radial-wind-speed values e.g. at 2 km did not follow a cosine function (Fig. 4b).

The visual examination of the mlf-cs fields led to the identification of three types of remarkable mlf-cs patterns. The first type was represented by large elongated areas of positive mlf-cs's accompanied by large elongated areas of negative mlf-cs's aligned with the mean wind (Fig. 5a) during the day. In the atmosphere, these types of patterns are encountered concurrently with the existence of rolls, where strong descending motions enhance the horizontal wind speed and ascending motions reduce it. The second type of pattern was characterized by large enclosed areas of a positive mlf-cs field attached to large enclosed areas of a negative mlf-cs field (Fig. 5b) during the day. The convergence zones formed between the positive and negative mlf-cs fields during unstable conditions (e.g. high solar radiation) are able to form strong unaligned thermals. Finally, the third type of pattern consisted of narrow elongated areas alternating between positive and negative mlf-cs's aligned with the mean wind (Fig. 5c). These patterns resemble turbulent streaks as described in Sect. 1.

In order to train the classification algorithm (Sect. 4.1), it was necessary to build an ensemble of cases for which the presence of rolls, unaligned thermals or streaks was confirmed by observations other than the lidar measurements. Moderate Resolution Imaging Spectroradiometer (MODIS) true-colour images were used to detect the presence of cloud streets over Paris (Fig. 5d), which confirmed the existence of rolls as stated in Sect. 1. Close to the moment when the cloud streets were present, rolls patterns were observed at the turbulent radial fields (Fig. 5a). It is noteworthy to mention that, for the training ensemble, we selected only cases of rolls oc-

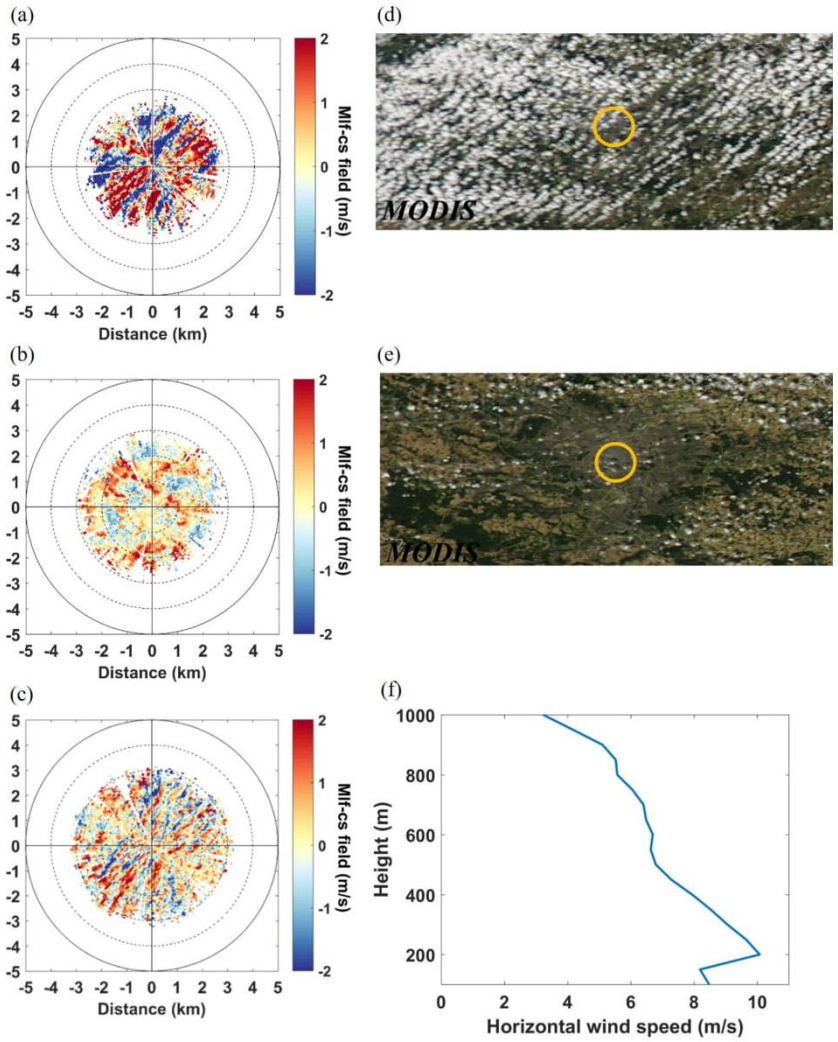


Figure 5. The upper part shows the three types of mlf-cs fields to classify: **(a)** rolls observed on 13 October 2014 at 12:52 UTC, **(b)** unaligned thermals observed on 16 September 2014 at 12:52 UTC and **(c)** streaks observed on 9 September 2014 at 20:49 UTC. The lower part shows the ancillary observations used to ascertain the structure type: **(d)** and **(e)** are true-colour images recorded by MODIS Aqua on the same day as **(a)** and **(b)** at 12:50 UTC, and **(f)** is the horizontal wind speed profile recorded by the Doppler lidar using the DBS technique on the same day as **(c)** at 20:51 UTC.

curring around the satellite overpass time to ensure the presence of cloud streets and thus the existence of rolls. However, for this classification we are interested in all the cases of rolls, with or without the formation of cloud streets. It is important to note that we observed the occurring patterns near the surface, hence near the lower part of the rolls. Regarding unaligned thermals, solar-radiation measurements from the meteorological station of Paris-Montsouris indicated the occasions when the hourly values were higher than the monthly average hourly values according to the Photovoltaic Geographical Information System (Huld et al., 2012), signifying fair-weather cumulus conditions. For approximately the same time of the day, we observe the unaligned thermals patterns. Figure 5b showcases an example of a turbulent radial wind field with unaligned thermals along with fair-weather cumuli over Paris as observed on MODIS true-colour images at approximately the same time (Fig. 5e).

Finally concerning streaks, a driving factor for their formation is the existence of a strong wind shear near the surface. The observation of the horizontal wind profiles from the DBS scans revealed when the local maxima horizontal wind speed was higher than $2\,\mathrm{m\,s^{-1}}$ compared to the local minima

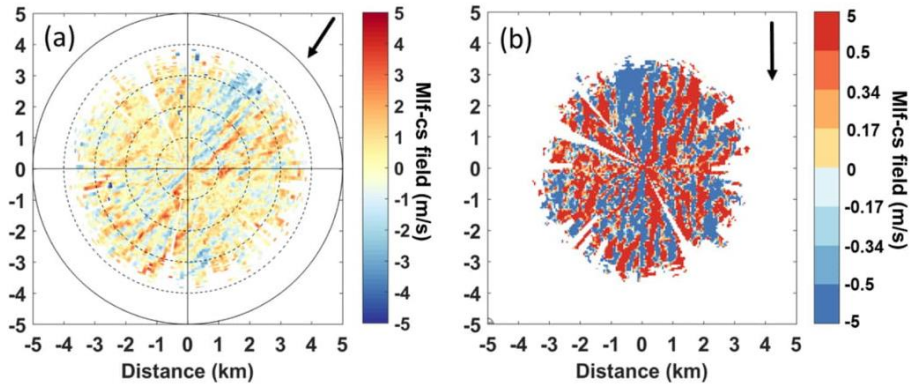


Figure 6. The mlf-cs field **(a)** before and **(b)** after image pre-processing with the arrow representing the mean wind direction on 10 September 2014 at 19:57 UTC.

Table 3. Co-occurrence matrix after the image pre-processing (Fig. 6b) for the first neighbour ($n = 1$) and for a cell pair aligned with the mean wind and oriented in the same direction (azimuth $\varphi = 0°$).

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	3065	226	164	118	113	57	35	94
2	255	67	77	58	36	26	23	48
3	181	81	59	61	44	51	35	72
4	133	58	63	91	71	50	40	92
5	98	51	59	65	67	63	58	154
6	58	36	50	53	75	72	78	169
7	46	30	38	53	60	61	55	231
8	73	45	78	104	151	201	246	3402

above it, which is defined as the threshold for nocturnal low-level jet events (Stull, 1988) (Fig. 5f). It is important to note that the location of the local maxima and minima of the horizontal wind speed were consistent during the study period, ranging from 200 to 300 and 400 to 500 m, respectively. The horizontal wind speed U_{hor} was estimated by the zonal u and meridional v winds via the expression

$$U_{\text{hor}} = \sqrt{u^2 + v^2}. \tag{3}$$

For the training ensemble, only night cases when streaks patterns (Fig. 5c) were accompanied by differences in local maxima and minima of the U_{hor} higher than $2\,\text{m}\,\text{s}^{-1}$ were selected. In total, 30 cases of each structure type were selected for the training ensemble with an extra category representing all the patterns that are not classified in the other three categories, such as chaotic patterns or cases when the VAD method cannot be applied (Fig. 4). Regarding rolls, streaks and thermals, only cases with symmetric radial wind fields were selected in order to ensure that the VAD method was applicable. The selection of symmetric radial wind fields was based on the visual examinations of the radial wind fields and the individual cosine function fits.

3.2 Computation of the co-occurrence matrices

In order to retrieve comparable texture analysis parameters from the mlf-cs field of the scans, the mlf-cs field was rotated so that the mean wind direction was aligned to the vertical (0° corresponds to a wind blowing from the north). Then, the coordinates were converted from polar to Cartesian. It was also important to adjust the contrast of the image so that the difference between the areas of positive and negative turbulent wind speed became more prominent. For this purpose, the contrast of the images was increased by mapping the turbulent-wind-speed values into eight levels. One bin included all the negative values below $-0.5\,\text{m}\,\text{s}^{-1}$; six bins were equally distributed between -0.5; and $+0.5\,\text{m}\,\text{s}^{-1}$ and one bin included all the positive values above $+0.5\,\text{m}\,\text{s}^{-1}$ (Fig. 6b).

For the automated classification of patterns, we need to map them to a space of corresponding numerical parameters. Each reconstructed mlf-cs field is represented by a matrix (cells corresponds to pixels) from which 8×8 co-occurrence matrices (CMs) can be constructed (Haralick et al., 1973). The rows and columns of the CM represent the wind levels from 1 to 8, whereas the cells contain the frequency of the combination of two neighbour pixels in the image. More specifically, the element at line i and column j contains the number of pixels with value i which are neighboured by pixels with value j. The first neighbour can be searched at different direction (e.g. left to right, top to bottom or diagonally) defining the cell pair orientation. In the same way a second, third, etc. neighbour can be selected. Thus, the CMs can be calculated for any cell pair orientations and neighbour order. CM were computed for various distances, i.e. neighbour orders n from 1 to 30 (distance from 50 m to 1.5 km), and all possible cell pair orientations, i.e. azimuth angles φ from $-90°$ (transverse direction from the mean wind in the anticlockwise direction) to $+90°$ (transverse direction in the clockwise direction). Table 3 shows the cell values of the CM

Table 4. Co-occurrence matrix after the image pre-processing (Fig. 6b) for the third neighbour ($n = 3$) and for the transverse direction in the clockwise direction (azimuth $\varphi = +90°$).

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	1497	231	203	182	165	168	170	1149
2	185	19	25	43	27	27	25	200
3	183	29	26	29	33	31	21	207
4	195	32	37	39	29	31	28	185
5	203	29	38	31	36	31	26	208
6	201	26	25	25	26	39	29	198
7	175	27	23	26	32	21	37	212
8	1063	179	187	196	243	206	217	1719

built from the image of Fig. 6b for the first neighbour ($n = 1$) and for a cell pair aligned with the mean wind and oriented in the same direction (azimuth $\varphi = 0°$). It is apparent that the vast majority of the occurrences are concentrated in the cells [1,1] and [8,8], as the structures are elongated and aligned with the mean wind direction.

On the other hand, Table 4 shows the CM of Fig. 6b for the third neighbour ($n = 3$) and for a cell pair oriented perpendicularly to the mean wind (transverse direction) with a clockwise rotation (azimuth angle $\varphi = +90°$). In this case, the occurrences have been distributed to the cells [1,1] and [8,8], as well as to the cells [1,8] and [8,1]. As we can see in Fig. 6b, the structures alternate between positive and negative values in the direction transverse to the mean wind, thus creating this difference in the CM compared to Table 3.

3.3 Texture analysis parameters for the classification of the turbulent structures

It is possible to compute several texture analysis parameters from each CM. Srivastava et al. (2018) were able to distinguish different synthetic patterns by using four texture analysis parameters: correlation, contrast, homogeneity and energy. Correlation indicates the existence of linear structures in the image, with high values associated with a large amount of linear structure in the image. Contrast reveals the local variations in an image, where a large amount of variations leads to high values. Homogeneity is self-explanatory, and the high values represent a homogeneous image. Finally, energy measures the uniformity of an image with the highest values corresponding to constant or periodic forms (Haralick et al., 1973; Yang et al., 2012). In the study of Srivastava et al. (2018), the striped patterns resemble the elongated patterns of streaks and rolls that we observe in the turbulent radial wind field. Therefore, the same texture analysis parameters were selected for calculation in our dataset. More particularly, the following parameters were computed by Eqs. (4)–(7).

$$\text{Homogeneity: } \mathrm{Hom}(\varphi,n)=\sum_{i,j}\frac{p(i,j)}{1+|i-j|}, \tag{4}$$

$$\text{Contrast: } \mathrm{Con}(\varphi,n)=\sum_{i,j}p(i,j)|i-j|^2, \tag{5}$$

$$\text{Correlation: } \mathrm{Cor}(\varphi,n)=\sum_{i,j}\frac{(i-\mu_i)(j-\mu_j)p(i,j)}{\sigma_i\sigma_j}, \tag{6}$$

$$\text{Energy: } \mathrm{En}(\varphi,n)=\sum_{i,j}p(i,j)^2, \tag{7}$$

where $p(i,j)=\frac{\mathrm{CM}(i,j)}{\sum_{i,j}\mathrm{CM}(i,j)}$ for the i,j position in the CM, marginal expectations $\mu_i=\sum_i\sum_j i\cdot p(i,j)$, $\mu_j=\sum_i\sum_j j\cdot p(i,j)$, and the marginal SDs $\sigma_i=\sqrt{\sum_i\sum_j(i-\mu_i)^2\cdot p(i,j)}$ and $\sigma_j=\sqrt{\sum_i\sum_j(j-\mu_j)^2\cdot p(i,j)}$.

At a given neighbour order n, it is then possible to study the dependence of the texture parameters to the azimuth angle φ (see an example of such a dependence in Fig. 7). The streaks and rolls have a more prominent peak in the longitudinal direction ($\varphi = 0°$) compared to the unaligned thermals and patterns of "others". As streaks and rolls are aligned with the mean wind (azimuth $\varphi = 0°$), those peaks result from the elongated shapes of these patterns.

Three parameters of the curve in Fig. 7 were selected in order to distinguish the different types of structures. For instance, for the homogeneity curves, these parameters are defined as follows in Eqs. (8)–(10).

$$\text{Amplitude: } \mathrm{Hom.Amp}(n)=\max_{\varphi}(\mathrm{Hom}(\varphi,n))-\min_{\varphi}(\mathrm{Hom}(\varphi,n)) \tag{8}$$

$$\text{Integral: } \mathrm{Hom.Int}(n)=\sum_{\varphi}\mathrm{Hom}(\varphi,n) \tag{9}$$

$$\text{Symmetry: } \mathrm{Hom.Sym}(n)=\sum_{\varphi}|\mathrm{Hom}(\varphi,n)-\mathrm{Hom}(-\varphi,n)| \tag{10}$$

These three curve parameters were calculated for the 4 texture analysis parameters and for each of the 30 neighbour orders, which gives 360 parameters. In addition to these parameters, the time in UTC (close to solar time in Paris), the average mean wind speed and the root-mean-square error (RMSE) of the cosine fit (Fig. 3b) were included in the classification parameters. The total number of classification parameters associated with each scan was therefore 363.

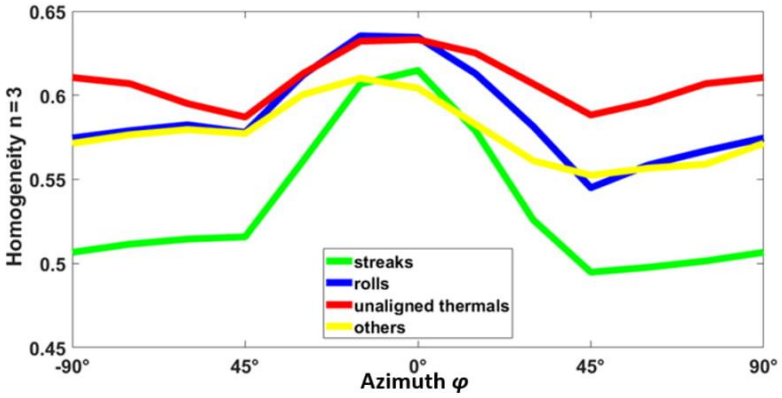


Figure 7. Third-neighbour homogeneity as a function of azimuth for one selected scan of each type.

4 Classification using supervised machine learning

4.1 Algorithm training and classification error

In order to classify the mlf-cs's according to the aforementioned texture analysis parameters, the supervised machine-learning methodology was applied (Bonamente, 2017; James et al., 2000; Kubat, 2017). The QDA algorithm was used, which minimizes the total probability of misclassification, assuming that features of each class have a multidimensional Gaussian distribution. QDA or normal Bayesian classification (Hastie et al., 2009) is the parametric approach implying that probability density functions (PDFs) belong to the family of normal distributions. It is a classical algorithm of supervised machine learning, based on the principle of maximum likelihood. The general idea is to estimate the PDF for each class and then select the most probable class (Kubat, 2017).

The greedy algorithm of stepwise forward selection was used in the article, which is the standard and frequently used method of reduction of the feature space. As indicated in Sokolov et al. (2020), it can be formulated as follows. The features are divided into two groups: accepted in the classification model and remaining, for which an estimate of the possibility of acceptance into the model is checked. Features from the set of "remains" are consecutively added to the model, and corresponding estimations of the classification error are calculated. From the received set of errors, the minimum is chosen and compared with the error of the previous model. If a significant reduction of the error occurred, then the corresponding feature is accepted into the model; if this is not true, then the process stops. The QDA was trained (Hastie et al., 2009; Sokolov et al., 2020) with the 150-case ensemble described in Sect. 3.1: 30 cases of streaks, 30 cases of rolls, 30 cases of unaligned thermals and 60 cases of others. The category of others was represented by twice as many cases, since it is expected to be the dominant category in the classification, as it includes the chaotic mlf-cs fields and the cases where the mlf-cs field was not computed successfully by the VAD method. The algorithm can be sensitive to an unbalanced training ensemble. Therefore, the selection of a training ensemble based on the expected results was preferred (Kubat, 2017).

The total omission error (see Sokolov et al., 2020) of the classification based on the QDA technique could be estimated for the training ensemble by means of 10-fold cross validation. This error is referred further as the classification error. In this method, the algorithm is trained using 90 % of the training ensemble (135 cases); it is then applied to the remaining 10 % (15 cases), and the resulting (output) classes are compared to the expected (target) classes. The process is repeated 10 times, each time extracting a different 10 % sample for testing, until the entire training ensemble has been tested.

As the number of dimensions of the feature space (363) was significantly higher than the number of patterns of the training ensemble (150), the application of all the features leads to the curse of the dimensionality problem, when the classification works well only for the training data and fails for the test set. In order to deal with this problem, we reduced the feature space by selecting the most informative components using the stepwise forward-selection algorithm (Sokolov et al., 2020). The resulting sequence of these components and the decrease of the 10-fold cross-validation classification error are presented in Fig. 8. The classification error reached a minimum of about 9.2 % when five parameters were used; taking more into account increased the classification error.

Analytically, these parameters are the amplitude of the 2nd-neighbour homogeneity curve, the integral of the 18th-neighbour contrast curve, the amplitude of the 4th-neighbour contrast curve, the integral of the 8th-neighbour correlation curve and the symmetry of the 2nd-neighbour homogeneity curve. These results show that the prominent peaks are a distinctive characteristic for the elongated patterns, as the amplitude of the homogeneity and contrast curves are two of the

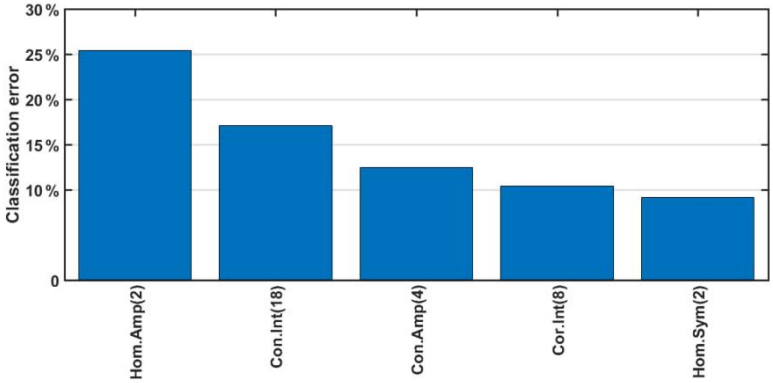


Figure 8. Parameters selected to minimize the classification error of the training ensemble by the QDA method. From left to right: amplitude of the homogeneity for the 2nd neighbour, integral of the contrast for the 18th neighbour, amplitude of the contrast for the 4th neighbour, integral of the correlation of the 8th neighbour and symmetry of the homogeneity for the 2nd neighbour.

significant parameters. Furthermore, the integral or more precisely the sum of the points of the curves for the contrast and for the correlation curves are significant parameters as well. This is important especially for the distinction between the categories thermals and others, as their amplitude may not differ substantially, since the patterns are not towards a specific direction, yet a chaotic area will have higher values of contrast and lower values of correlation compared to an enclosed homogeneous area. Finally, the symmetry of the homogeneity curve as a classifier reveals the urgency to align the turbulent radial wind fields to the mean wind direction and thus align the structures such as streaks and rolls with the mean wind direction in order to be distinguishable from the random positions of the enclosed structures of the thermals or the chaotic structures of others. It is also crucial to note that the parameters cover various distances, from the 2nd neighbour, which in grid points is 100 m, to the 18th neighbour, which is 900 m. This is necessary for our classification, since streaks and rolls are both elongated patterns, but their transverse horizontal sizes differ. Furthermore, it demonstrates the ability of the algorithm to distinguish structures with different sizes. It is noteworthy that the curve parameters play a more significant role in the classification of the structures in comparison to time, mean wind field and cosine fit RMSE.

The detailed results of the cross validation of the QDA classification for the algorithm with five predictors are displayed in Table 5. The algorithm allowed for classifying correctly about 91 % of the training ensemble. The algorithm performs the most precise classification for the streaks with a classification error of only 3.3 %, as one case was misclassified as rolls. Regarding the category of others, the results are equivalently accurate with a classification error of 3.3 %, as two cases were misclassified as thermals. Moreover, the performance of the algorithm for rolls was good with a classification error of 10 % with three cases being misclassified as thermals. Thermals were the most troublesome type for classification by the algorithm; the algorithm classified correctly 24 cases. Four cases were misclassified as rolls, and two cases were misclassified as others, showing a classification error of 20 %.

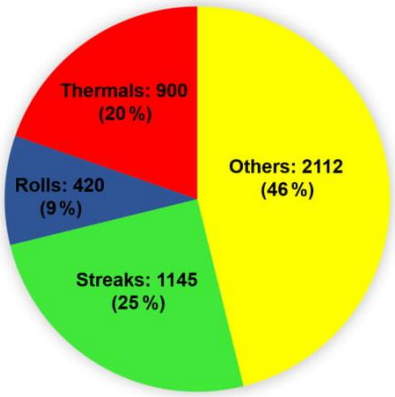


Figure 9. Classification of the whole ensemble using the QDA method according to the parameters of Fig. 8.

4.2 Results of the trained algorithm over the 2-month dataset

The whole dataset, consisting of 4577 scans, was classified according to the five parameters showcased in Fig. 8. The results are displayed in Fig. 9.

The algorithm classifies 54 % of the 2-month dataset as containing mlf-cs's and 34 % in particular as coherent structures (streaks and rolls). The most frequent cases of mlf-cs's were streaks (25 %), and the least frequent cases were rolls (9 %). It is important to note that, in our classification, we considered only thermals and rolls during daytime. Figure 10 illustrates the number of occurrences for each type of struc-

Table 5. Confusion matrix calculated for the training dataset. The "target class" corresponds to the visual classification, while the "output class" corresponds to the class attributed by the algorithm. Therefore, the cells in the "roll" column, for instance, give the number of roll cases that were classified properly (roll line) or improperly (other lines) in the different categories.

	Target class				
Output class	Others	Streaks	Rolls	Thermals	
Others	58	0	0	2	**96.7 %**
	38.7 %	0.0 %	0.0 %	1.3 %	**3.3 %**
Streaks	0	29	0	0	**100.0 %**
	0.0 %	19.3 %	0.0 %	0.0 %	**0.0 %**
Rolls	0	1	27	4	**84.4 %**
	0.0 %	0.7 %	18.0 %	2.7 %	**15.6 %**
Thermals	2	0	3	24	**82.8 %**
	1.3 %	0.0 %	2.0 %	16.0 %	**17.2 %**
	96.7 %	**96.7 %**	**90.0 %**	**80.0 %**	**92.0 %**
	3.3 %	**3.3 %**	**10.0 %**	**20.0 %**	**8.0 %**

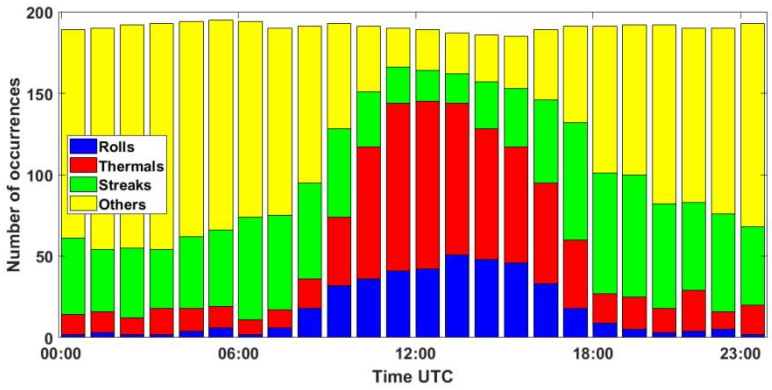


Figure 10. Histogram of the number of occurrences of the different types of structures with respect to time in UTC.

ture at a particular time of the day during the 2 months of the campaign. It is evident that despite time not being one of the selected classifiers, the number of occurrences of the structures shows a distribution that can be associated with the atmospheric conditions. More particularly, rolls and thermals were mainly classified during the day. This result is noteworthy, as these structures are linked to a well-developed atmospheric boundary layer during the day. On the contrary, there were scarcely any roll cases observed at night, and a few unaligned thermals were classified at night. This stems from the training process, where some cases of thermal were improperly classified as others and the reverse. Regarding the cases of others, these were mostly observed during the night. This was expected, since the cases of low winds with no defined direction – when the VAD method cannot be applied – occur mainly during the night. We also see that streaks were observed more frequently during the night, when mechanical turbulence becomes dominant. This was also expected as the nocturnal low-level jets are a main driving force for the formation of streaks, and we observed the occurrence of the local maxima of the horizontal wind speed near the surface higher than $2\,\mathrm{m\,s^{-1}}$ compared to the local minima over Paris for 20 out of the 62 nights during the VEGILOT campaign.

5 Conclusions

The current study showcases that it is possible to identify and classify mlf-cs's such as streaks, rolls and unaligned thermals with horizontal scans from a single Doppler lidar by combining texture analysis parameters and the QDA supervised machine-learning technique. By applying the VAD method to the radial-wind observations, it is possible to identify mlf-cs's that can be distinguished as narrow elongated (streaks), wide elongated (rolls), large enclosed (thermals) and chaotic (others) patterns. The diversity of the patterns was also depicted in the curves of the texture analysis parameters with

the elongated patterns (streaks and rolls) showing a prominent peak compared to more chaotic or enclosed patterns (unaligned thermals).

A training ensemble of 150 cases was selected by combining visual examination of the patterns and studying characteristic physical properties corresponding to streaks, rolls and unaligned thermals. Subsequently, the QDA algorithm with stepwise forward selection of the features was applied to the training ensemble, and its performance was estimated using the cross-validation technique. The results showed a successful classification for 91 % of the training ensemble using five texture analysis parameters as predictors. More particularly, these parameters were the amplitude of the 2nd-neighbour homogeneity curve and the amplitude of the 4th-neighbour contrast curve which were associated with the prominent peaks of the elongated patterns (streaks and rolls). Furthermore, the integral of the 18th-neighbour contrast curve and the integral of the 8th-neighbour correlation curve which could distinguish, for example, chaotic patterns (others) with high contrast and lower values of correlation between neighbour points compared to an enclosed homogeneous type (thermals). Finally, the symmetry of the 2nd-neighbour homogeneity curve revealed the importance to align the mlf-cs fields to the mean wind direction. Another striking outcome of the QDA classification was the variety of the classifiers in terms of distance between the grid points. The 2nd neighbour translates in a distance between two grid points equivalent to 100 m, and for the 18th neighbour this is 900 m. This is essential for the classification between patterns with different sizes such as streaks and rolls. The algorithm performed best for the category of streaks with a classification error of only 3.3 %. Time, mean wind speed and the cosine fit RMSE of the VAD method were not selected by the algorithm for the classification.

The whole ensemble of the 4577 scans was classified by the trained QDA algorithm using the five selected texture analysis parameters. The results showed that 54 % of cases were classified as mlf-cs's, among which 34 % were coherent structures (streaks and rolls). The streaks were mostly observed during night, whereas the thermals and rolls were almost exclusively observed during the day, with only a few cases classified between sunset and sunrise. The classified ensemble can be used for statistical studies of the mlf-cs physical parameters, such as structure size, as a function of weather conditions (PBL height, temperature, wind speed, radiation, etc.). Moreover, the development of the structures can be analysed and comprehended.

Data availability. All lidar data used in the study are property of the Laboratoire de Physico-Chimie de l'Atmosphere (LPCA), Dunkirk, France, and are not publicly available.

Supplement. The supplement related to this article is available online at: https://doi.org/10.5194/amt-13-6579-2020-supplement.

Author contributions. IC, EDE, HD and AS conceptualized this study and developed the methodology. HD, PA and MF installed and monitored the instrument on the field. IC processed the data and analysed the results for all parts of the study, with the help of HD, AS and EDM for Sect. 4. IC wrote the original draft of the paper, with contributions from HD, EDE and AS. All authors participated in the review and editing of the paper and agreed to this version.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. The authors thank François Ravetta, Jacques Pelon, Gilles Plattner and Amelie Klein of the LATMOS, Sorbonne University, Paris, for organizing and carrying out the VEGILOT campaign.

We acknowledge the use of imagery from the NASA Worldview application (https://worldview.earthdata.nasa.gov/, last access: 2 December 2020), part of the NASA Earth Observing System Data and Information System (EOSDIS).

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).

Financial support. This work is a contribution to the CPER (Contrat de Plan Etat-Région) research project IRenE (Innovation et Recherche en Environnement) and Climibio. The work is supported by the French Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation, the region Hauts-de-France and the European Regional Development Fund. The work is also 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; contract no. ANR-11-LABX-0005-01) and by the regional council of Nord-Pas-de-Calais and the European Regional Development Fund. This study was funded by the RFBR (Russian Foundation for Basic Research; project no. 20-07-00370) and Moscow Center for Fundamental and Applied Mathematics (Agreement 075-15-2019-1624 with the Ministry of Education and Science of the Russian Federation; MESRF).

Review statement. This paper was edited by Marcos Portabella and reviewed by two anonymous referees.

References

Adrian, R. J.: Hairpin vortex organization in wall turbulence, Phys. Fluids, 19, 41301, https://doi.org/10.1063/1.2717527, 2007.

Alparone, L., Benelli, G., and Vagniluca, A.: Texture-based analysis techniques for the classification of radar images, IET Digital Library, IEE Proc. F, 137, 276–282, https://doi.org/10.1049/ip-f-2.1990.0041, 1990.

Aouizerats, B., Tulet, P., Pigeon, G., Masson, V., and Gomes, L.: High resolution modelling of aerosol dispersion regimes during the CAPITOUL field experiment: from regional to local scale interactions, Atmos. Chem. Phys., 11, 7547–7560, https://doi.org/10.5194/acp-11-7547-2011, 2011.

Banta, R. M., Newsom, R. K., Lundquist, J. K., Pichugina, Y. L., Coulter, R. L., and Mahrt, L.: Nocturnal low-level jet characteristics over Kansas during cases-99, Bound.-Lay. Meteorol., 105, 221–252, https://doi.org/10.1023/A:1019992330866, 2002.

Barthlott, C., Drobinski, P., Fesquet, C., Dubos, T., and Pietras, C.: Long-term study of coherent structures in the atmospheric surface layer, Bound.-Lay. Meteorol., 125, 1–24, https://doi.org/10.1007/s10546-007-9190-9, 2007.

Bonamente, M.: Functions of random variables and error propagation, in: Statistics and Analysis of Scientific Data, Grad. Texts Phys., Springer, New York, USA, 55–83, https://doi.org/10.1007/978-1-4939-6572-4, 2017.

Browning, K. A. and Wexler, R.: The determination of kinematic properties of a wind field using Doppler radar, J. Appl. Meteorol., 7, 105–113, https://doi.org/10.1175/1520-0450(1968)007<0105:tdokpo>2.0.co;2, 1968.

Brümmer, B.: Roll and cell convection in wintertime Arctic cold-air outbreaks, J. Atmos. Sci., 56, 2613–2636, https://doi.org/10.1175/1520-0469(1999)056<2613:RACCIW>2.0.CO;2, 1999.

Cariou, J. P., Parmentier, R., Valla, M., Sauvage, L., Antoniou, I., and Courtney, M.: An innovative and autonomous 1.5 µm Coherent lidar for PBL wind profiling, in: Proceedings of the 14th Coherent Laser Radar Conference, Snowmass, Colorado, USA, 8–13 July 2007, 35–38, 2007.

Castellano, G., Bonilha, L., Li, L. M., and Cendes, F.: Texture analysis of medical images, Clin. Radiol., 59, 1061–1069, https://doi.org/10.1016/j.crad.2004.07.008, 12, 2004.

Chai, T., Lin, C.-L., and Newsom, R. K.: Retrieval of microscale flow structures from high-resolution Doppler lidar data using an adjoint model, J. Atmos. Sci., 13, 1500–1520, https://doi.org/10.1175/1520-0469(2004)061<1500:ROMFSF>2.0.CO;2, 2004.

Drobinski, P. and Foster, R. C.: On the origin of near-surface streaks in the neutrally-stratified planetary boundary layer, Bound.-Lay. Meteorol., 108, 247–256, https://doi.org/10.1023/A:1024100125735, 2003.

Drobinski, P., Brown, R. A., Flamant, P. H., and Pelon, J.: Evidence of organized large eddies by ground-based Doppler lidar, sonic anemometer and sodar, Bound.-Lay. Meteorol., 88, 343–361, https://doi.org/10.1023/A:1001167212584, 1998.

Drobinski, P., Carlotti, P., Newsom, R. K., Banta, R. M., Foster, R. C., and Redelsperger, J.-L.: The structure of the near-neutral atmospheric surface layer, J. Atmos. Sci., 61, 699–714, https://doi.org/10.1175/1520-0469(2004)061<0699:TSOTNA>2.0.CO;2, 2004.

Eberhard, W. L., Cupp, R. E., and Healy, K. R.: Doppler lidar measurement of profiles of turbulence and momentum flux, J. Atmos. Ocean. Tech., 6, 809–819, https://doi.org/10.1175/1520-0426(1989)006<0809:dlmopo>2.0.co;2, 1989.

Haralick, R. M., Dinstein, I., and Shanmugam, K.: Textural features for image classification, IEEE T. Syst. Man. Cyb., 6, 610–621, https://doi.org/10.1109/TSMC.1973.4309314, 1973.

Hastie, T., Tibshirani, R., and Friedman, J.: The elements of statistical learning: Data mining, inference, and prediction, Springer Series in Statistics, Springer, New York, USA, 2009.

Holli, K., Lääperi, A. L., Harrison, L., Luukkaala, T., Toivonen, T., Ryymin, P., Dastidar, P., Soimakallio, S., and Eskola, H.: Characterization of breast cancer types by texture analysis of magnetic resonance images, Acad. Radiol., 17, 135–141, https://doi.org/10.1016/j.acra.2009.08.012, 2010.

Huld, T., Müller, R., and Gambardella, A.: A new solar radiation database for estimating PV performance in Europe and Africa, Sol. Energy, 86, 1803–1815, https://doi.org/10.1016/j.solener.2012.03.006, 2012.

Hussain, A. K. M. F.: Coherent structures – Reality and myth, Phys. Fluids, 26, 2816–2850, https://doi.org/10.1063/1.864048, 1983.

Hutchins, N. and Marusic, I.: Evidence of very long meandering features in the logarithmic region of turbulent boundary layers, J. Fluid Mech., 579, 1–28, https://doi.org/10.1017/S0022112006003946, 2007.

Iwai, H., Ishii, S., Tsunematsu, N., Mizutani, K., Murayama, Y., Itabe, T., Yamada, I., Matayoshi, N., Matsushima, D., Weiming, S., Yamazaki, T., and Iwasaki, T.: Dual-Doppler lidar observation of horizontal convective rolls and near-surface streaks, Geophys. Res. Lett., 35, L14808, https://doi.org/10.1029/2008GL034571, 2008.

James, G., Witten, D., Hastie, T., and Tibshirani, R.: An introduction to statistical learning, Springer Texts in Statistics, Springer, New York, USA, https://doi.org/10.1007/978-1-4614-7138-7, 2000.

Kayitakire, F., Hamel, C., and Defourny, P.: Retrieving forest structure variables based on image texture analysis and IKONOS-2 imagery, Remote Sens. Environ., 102, 390–401, https://doi.org/10.1016/j.rse.2006.02.022, 2006.

Kelly, R. D.: A single Doppler radar study of horizontal-roll convection in a lake-effect snow storm (Lake-Michigan), J. Atmos. Sci., 39, 1521–1531, https://doi.org/10.1175/1520-0469(1982)039<1521:ASDRSO>2.0.CO;2, 1982.

Khanna, S. and Brasseur, J. G.: Three-dimensional buoyancy and shear-induced local structure of the atmospheric boundary layer, J. Atmos. Sci., 55, 710–743, https://doi.org/10.1175/1520-0469(1998)055<0710:TDBASI>2.0.CO;2, 1998.

Klein, A., Ravetta, F., Thomas, J. L., Ancellet, G., Augustin, P., Wilson, R., Dieudonné, E., Fourmentin, M., Delbarre, H., and Pelon, J.: 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, https://doi.org/10.1016/j.atmosenv.2018.10.009, 2019.

Kropfli, R. A. and Kohn, N. M.: Persistent horizontal rolls in the urban mixed layer as revealed by dual-Doppler radar, J. Appl. Meteorol., 17, 669–676, https://doi.org/10.1175/1520-0450(1978)017<0669:phritu>2.0.co;2, 1978.

Kubat, M.: An introduction to machine learning, Springer International Publishing, Springer, New York, USA, https://doi.org/10.1007/978-3-319-63913-0, 2017.

Kunkel, K. E., Eloranta, E. W., and Weinman, J. A.: Remote determination of winds, turbulence spectra and energy dissipation rates in the boundary layer from lidar measurements, J. Atmos. Sci., 37, 978–985, https://doi.org/10.1175/1520-0469(1980)037<0978:rdowts>2.0.co;2, 1980.

LeMone, M.: The structure and dynamics of the horizontal roll vortices in the planetary boundary layer, J. Atmos. Sci., 30, 1077–1091, https://doi.org/10.1175/1520-0469(1973)030<1077:tsadoh>2.0.co;2, 1972.

Lemonsu, A. and Masson, V.: Simulation of a summer urban breeze over Paris, Bound.-Lay. Meteorol., 104, 463–490, https://doi.org/10.1023/A:1016509614936, 2002.

Lhermitte, R. M.: Note on wind variability with Doppler radar, J. Atmos. Sci., 19, 343–346, https://doi.org/10.1175/1520-0469(1962)019%3C0343:NOWVWD%3E2.0.CO;2, 1962.

Lohou, F., Druilhet, A., and Campistron, B.: Spatial and temporal characteristics of horizontal rolls and cells in the atmospheric boundary layer based on radar and in situ observations, Bound.-Lay. Meteorol., 89, 407–444, https://doi.org/10.1023/A:1001791408470, 1998.

Moeng, C.-H. and Sullivan, P. P.: A comparison of shear and buoyancy-driven planetary boundary layer flows, J. Atmos. Sci., 51, 999–1022, https://doi.org/10.1175/1520-0469(1994)051<0999:acosab>2.0.co;2, 1994.

Newsom, R., Calhoun, R., Ligon, D., and Allwine, J.: Linearly organized turbulence structures observed over a suburban area by Dual-Doppler lidar, Bound.-Lay. Meteorol., 127, 111–130, https://doi.org/10.1007/s10546-007-9243-0, 2008.

Reinking, R. F., Doviak, R. J., and Gilmer, R. O.: Clear-air roll vortices and turbulent motions as detected with an airborne gust probe and dual-Doppler radar, J. Appl. Meteorol., 20, 678–685, https://doi.org/10.1175/1520-0450(1981)020<0678:CARVAT>2.0.CO;2, 1981.

Saint-Pierre, C., Becue, V., Diab, Y., and Teller, J.: Case study of mixed-use high-rise location at the Greater Paris scale, WIT Trans. Ecol. Envir., 129, 251–262, https://doi.org/10.2495/SC100221, 2010.

Sandeepan, B. S., Rakesh, P. T., and Venkatesan, R.: Observation and simulation of boundary layer coherent roll structures and their effect on pollution dispersion, Atmos. Res., 120, 181–191, https://doi.org/10.1016/j.atmosres.2012.08.016, 2013.

Sokolov, A., Dmitriev, E., Gengembre, C., and Delbarre, H.: Automated classification of regional meteorological events in a coastal area using in-situ measurements, J. Atmos. Ocean. Tech., 37, 723–739, https://doi.org/10.1175/JTECH-D-19-0120.1, 2020.

Soldati, A.: Particles turbulence interactions in boundary layers, ZAMM J. Appl. Math. Mech., 85, 683–699, https://doi.org/10.1002/zamm.200410213, 2005.

Srivastava, D., Rajitha, B., Agarwal, S., and Singh, S.: Pattern-based image retrieval using GLCM, Neural Comput. Appl., 32, 1–14, https://doi.org/10.1007/s00521-018-3611-1, 2018.

Stull, R. B.: An introduction to boundary layer meteorology, Kluwer Academic Publishers, Springer, Dordrecht, Germany, https://doi.org/10.1007/978-94-009-3027-8, 1988.

Träumner, K., Damian, T., Stawiarski, C., and Wieser, A.: Turbulent structures and coherence in the atmospheric surface layer, Bound.-Lay. Meteorol., 154, 1–25, https://doi.org/10.1007/s10546-014-9967-6, 2015.

Troude, F., Dupont, E., Carissimo, B., and Flossmann, A. I.: Relative influence of urban and orographic effects for low wind conditions in the Paris area, Bound.-Lay. Meteorol., 103, 493–505, https://doi.org/10.1023/A:1014903627803, 2002.

Tur, A. V. and Levich, E.: The origin of organized motion in turbulence, Fluid Dyn. Res., 10, 75–90, https://doi.org/10.1016/0169-5983(92)90009-L, 1992.

Vasiljević, N., Lea, G., Courtney, M., Cariou, J.-P., Mann, J., and Mikkelsen, T.: Long-range WindScanner system, Remote Sens.-Basel, 8, 896, https://doi.org/10.3390/rs8110896, 2016.

Weckwerth, T. M. and Parsons, D. B.: A review of convection initiation and motivation for IHOP_2002, Mon. Weather Rev., 134, 5–22, https://doi.org/10.1175/MWR3067.1, 2006.

Weckwerth, T. M., Horst, T. W., and Wilson, J. W.: An observational study of the evolution of horizontal convective rolls, Mon. Weather Rev., 127, 2160–2179, https://doi.org/10.1175/1520-0493(1999)127<2160:AOSOTE>2.0.CO;2, 1999.

Wilson, R. B., Start, G. E., Dickson, C. R., and Ricks, N. R.: Diffusion under low windspeed conditions near Oak Ridge, Tennessee, NOAA Technical Memorandum ERL ARL-61, 83, 1976.

Yang, X., Tridandapani, S., Beitler, J. J., Yu, D. S., Yoshida, E. J., Curran, W. J., and Liu, T.: Ultrasound GLCM texture analysis of radiation-induced parotid-gland injury in head-and-neck cancer radiotherapy: An *in vivo* study of late toxicity, Med. Phys., 39, 5732–5739, https://doi.org/10.1118/1.4747526, 2012.

Young, G. S., Kristovich, D. A. R., Hjelmfelt, M. R., and Foster, R. C.: Rolls, streets, waves, and more: A review of quasi-two-dimensional structures in the atmospheric boundary layer, B. Am. Meteorol. Soc., 83, 997–1002, https://doi.org/10.1175/1520-0477(2002)083<0997:RSWAMA>2.3.CO;2, 2002.

3.1.3 Additional discussion regarding the classification methodology

The results of the abovementioned paper were rather satisfying as the classification error was approximately 9 % and the diurnal variation of the structures followed the expected behaviour according to the atmospheric conditions. This classification was estimated by converting the mlf-cs field into an 8-bins field with the values divided as explained in Section 2.3.1: one bin with values greater than +0.5 $m \cdot s^{-1}$, six equal bins with values between +0.5 and −0.5 $m \cdot s^{-1}$ and one bin with values lower than −0.5 $m \cdot s^{-1}$. Other binning configurations were tested in order to obtain the lowest classification error possible.

The bins should be selected in a way to enhance the differences, which consist mainly in distinguishing the positive and negative values in the mlf-cs fields, because they correspond to opposite wind directions. One may primarily think to use only 2 bins, containing respectively all the positive and all the negative values. This was indeed one of the first binning options that was tested. For this case, the KNN (3 neighbours) algorithm gave the lowest classification error, but it was approximately of 18 % with 13 texture analysis parameters as classifiers. This unsatisfying result indicated that, apparently, the size of the co-occurrence matrix is particularly important. The 2 bins will create 2 by 2 co-occurrence matrices, which appear to be too small to properly compute the texture analysis parameters. In Equations (9) to (11), it is evident that the difference in bin number is important for the estimation of the texture analysis parameters and thus, more bins should be selected for this type of data. It can be noted that, for the 2-bin configuration, the cosine fit error parameter was selected by the algorithm as 5th classifier.

Consequently, a 4-bin configuration was tested, with one bin containing all the values greater than +0.5 $m \cdot s^{-1}$, one bin with the values between +0.5 and 0 $m \cdot s^{-1}$, one bin with the values between 0 and −0.5 $m \cdot s^{-1}$ and finally one bin with the values lower than −0.5 $m \cdot s^{-1}$. This binning selection slightly lowered the classification error at 16 % for the KNN (1 neighbour) algorithm with 14 texture analysis parameters as classifiers. One of the classifiers selected by the algorithm in this case was time (in 6th position). The selection of 8 bins allowed to decrease the classification error significantly without any physical parameter being selected by the algorithm. There were no tests with more than 8 bins because we concluded that more bins near 0 $m \cdot s^{-1}$ will not be useful for a further significant decrease of the classification error.

It is noteworthy that one more binning configuration was tested and produced a classification error below 7 %. It was the default option of the graycomatrix function in MATLAB for which the 8 bins were divided as follows: one bin with all the negative values, 6 equally distributed bins between 0 and 1 $m \cdot s^{-1}$ and one bin with the values greater than

1 m·s^{-1}. For this binning configuration, the classification error reached 6.7 % for the KNN (1 neighbour) algorithm, with 13 texture analysis parameters as classifiers. However, even if this binning configuration provided a smaller classification error than the one that was eventually retained, the difference between the errors was not large and the algorithm needed 8 more texture analysis parameters. This is not a good sign regarding the stability of the classification, as the classification error oscillated around the minimum values. Furthermore, the KNN algorithm with one neighbour, as explained in Section 1.3.2, has the disadvantage to classify one case according to only one neighbour. For a relatively small training ensemble of 150 cases, this could introduce systematic errors. As a result, the binning options with the classification error of approximately 9 % was preferred. The classification error with the classifiers that minimize it for the three other binning options test are presented in Figure 39.

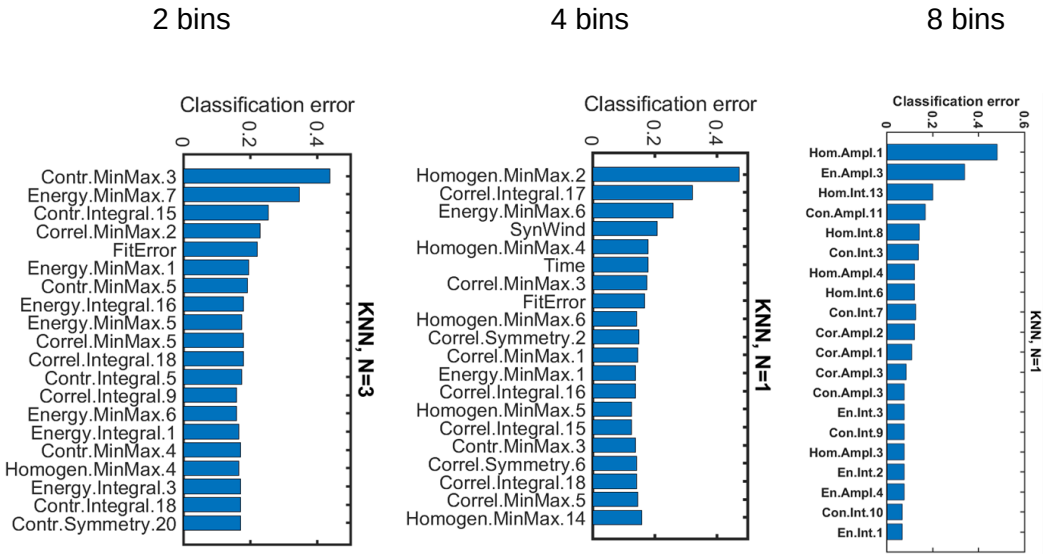


Figure 39: The classification error with the classifiers that minimize it for the other tested binning options and for the KNN algorithm using only the first neightbor. From left to right: 2 bins (one for positive values and one for negative), 4 bins (one for values greater than 0.5 m·s-1, one between 0.5 and 0 m·s-1, one between 0 and -0.5 m·s-1 and one lower than -0.5 m·s-1) and 8 bins (one for values greater than 0.5 m·s-1, six equally distributed bins between 0.5 and -0.5 m·s-1 and one lower than -0.5 m·s-1).

It is also important to note that only four out of the fourteen Haralick parameters were used for this classification, the same ones that had been used for the classification of patterns similar to those that exist in the mlf-cs fields. Other Haralick parameters could be more suitable for the classification of the mlf-cs fields, but they were not tested in this study, as the number of parameters was already very high (363). The development of an algorithm that would both identifies the best binning configuration and the appropriate Haralick parameters in order to minimize the classification error could be a separate

subject of study. Last but not least, increasing the size of the training ensemble could be another way to improve the classification error.

3.2 Physical properties of coherent structures over Paris

3.2.1 Analyzing the results of the automated classification for a large dataset

The classification of the whole ensemble of 4577 cases was interesting by itself, but the actual goal was to associate the classified structures with physical parameters. In Section 3.1.2, the diurnal variation of the classified structures for the whole ensemble shows a good agreement with the time of occurrence that can be expected from theoretical knowledge. Rolls and thermals cases were mostly classified during the day and “others” and streaks were mostly found mostly during the night. This was a good indication that the classification was working as intended. In that regard more physical parameters were associated with the structures in order to validate the classification. Section 3.2.2 presents the article, under review at the Journal of Applied Meteorology and Climatology at the time of submission, regarding the physical properties of the structures for the whole dataset recorded during the VEGILOT campaign.

Apart from revealing the physical parameters associated to the structures and showcasing the performance of the classification methodology, the study of Section 3.2.2 demonstrates how easy and accessible the data become after the classification. It is possible to examine all the data simultaneously, as it is the case for the distribution of the mean wind speed values for each type of the structures. Additionally, a sub-ensemble of random cases can be also examined, for instance when an additional parameter (the ABL height for this study) is not available for all times/dates. Furthermore, it facilitates the depiction of peculiar case studies, as the classification results can be plotted easily against any parameter. In this article, a case study is depicted with several nllj events observed in concurrence with streaks cases, followed by rolls cases during the day.

3.2.2 The validation of the automated classification based on the physical properties of the coherent structures

Submitted to the Journal of Applied Meteorology and Climatology

Properties of turbulent coherent structures over Paris: a study based on an automated classification method for Doppler lidar observations

Ioannis Cheliotis[1], Elsa Dieudonné[1], Hervé Delbarre[1], Anton Sokolov[1], Egor Dmitriev[2], Patrick Augustin[1], Marc Fourmentin[1], François Ravetta[3], Jacques Pelon[3]

[1] Laboratoire de Physico-Chimie de l'Atmosphère (LPCA), UR 4493, Université du Littoral Côte d'Opale (ULCO), Dunkirk, France

[2] Marchuk Institute of Numerical Mathematics Russian Academy of Sciences, Moscow, Russia

[3] Laboratory Atmosphere, Backgrounds, Space Observations (LATMOS) / Pierre Simon Laplace Institute (IPSL), Sorbonne University / CNRS, Paris, France

Abstract The previous studies that were carried out in order to examine the coherent structures in the atmosphere, using Doppler wind lidar observations, focused exclusively on specific cases for short periods of time. In particular, the structures were visually detected and the lidar images were classified accordingly, making this process time-consuming. We have instead developed an automated classification based on texture analysis parameters and the quadratic discriminant analysis algorithm for medium-to-large fluctuations and coherent structures as can be observed on the quasi-horizontal scans recorded by a single Doppler wind lidar. The algorithm classified a training ensemble of 150 cases into four types of patterns, namely streaks (narrow stripes), rolls (wide stripes), thermals (enclosed areas) and "others" (impossible to classify), with a good accuracy (91%). We then applied the trained algorithm to the full dataset of 4577 lidar scans recorded in Paris, atop a 75-m tower for a 2-month period (September-October 2014). In order to assess the quality of the classification, we now examine the physical properties of the classified cases. The results show a realistic classification of the data: roll and thermal cases were mostly classified during the day, when the atmospheric boundary layer was well developed, whereas the majority of streak cases were found during the night and often associated with nocturnal low-level jets (nllj) events. Roll and streak cases were mostly observed under moderate or high wind conditions. The detailed analysis of a four-day period reveals that the occurrence of streaks coincided exactly with the occurrence of nllj events. The streaks were replaced by rolls in the morning, when the thermal instabilities increased. The analysis of the space spectra in the direction transverse to the mean wind during these four days revealed a streak spacing of 200 to 400 m, in accordance to previous studies, and rolls sizes, as observed in the lower level of the mixed layer, of approximately 1 km.

Keywords Coherent structures, Atmospheric boundary layer, Automated classification, Wind lidar

1. Introduction

Coherent structures in the atmospheric boundary layer (ABL) have been the subject for a significant number of studies throughout the years. Their role in the atmospheric processes is substantial as they are key factors in the transport of heat and moisture (e. g. Brient et al., 2019), the dispersion of pollutants (e.g. Han et al., 2019) and the dust transport (e.g. Zhang et al., 2018) throughout the ABL. For this purpose, various methods have been developed in order to observe and better understand those structures. More specifically, some studies have utilized aircraft measurements (e. g. Brümmer et al., 1985; Martin and Bakan, 1991), some meteorological towers observations (e. g. Barthlott et al., 2007; Smedman, 1991), others remote sensing methods, such as radar (e. g. Eymard and Weill, 1988; Rabin et al., 1982) and lidar (e. g. Lin et al., 2008; Lothon et al., 2006) and others numerical simulations (e. g. Bartello et al., 1994; Jiang et al., 2017) or combination of the aforementioned methods (e. g. Bernard-Trottolo et al., 2004; Drobinski et al., 1998; Lohou et al., 1998). The remote sensing methods have the advantage to provide the visualization of the coherent structures (Drobinski et al., 2004). Despite this asset of the lidars, the studies so far have been limited to few cases covering only short periods of time. This is caused mainly because the analysis of the lidar data requires also the subjective observation of the researcher (e. g. Träumner et al., 2015).

We have developed an automated method with the aim to achieve consistent medium-to-large fluctuations and coherent structures (mlf-cs) classifications for single Doppler wind lidar observations (Cheliotis et al., 2020). We focus on the streaks and rolls as presented by Young et al., (2002). The streaks occur in the surface layer with the possibility to extend in the mixed layer. The main characteristics of the streaks is their formation which resembles stripes where stronger horizontal winds alternate with weaker ones and their direction is aligned with the mean wind (Khanna and Brasseur, 1998). The average space between the stronger horizontal winds is equivalent to hundreds of meters and their lifespan is no more than several tens of minutes (Drobinski and Foster, 2003). One of the main driving factors for the formation of the streaks is the high wind shear near the surface. The buoyancy contributes also to the streaks formation with the physics varying with regards to the atmospheric conditions (Moeng and Sullivan, 1994). They usually form under neutral or near-neutral stratification but it is possible to form under stable or unstable conditions as well (Khanna and Brasseur, 1998). The streaks formation can be visible during low clouds or fog events (e. g. MacDonald et al., 2020; Nakanishi, 2000).

On the other hand, rolls occur in the mixed layer, extending from the surface to the capping inversion and they are also aligned with the mean wind (LeMone, 1972). The main factors for the formation and maintenance of rolls are the low-level wind shear, the spatial variations of surface-layer heat flux and the homogeneity of the surface (Weckwerth and Parsons, 2006).The formation of the rolls resemble also stripes, but they are wider compared to streaks. It is possible to identify the rolls in the atmosphere during convective conditions. The rotation of the rolls in the vertical plane generates ascending and descending motions which lead to the formation of cloud rows and clear sky areas. This phenomenon is known as "cloud streets" (Lohou et al., 1998). However, the convective conditions is not a requirement for the formation of rolls, as they can form during low relative humidity conditions, in a phenomenon known as "dry streets" (Kuettner, 1971). The horizontal sizes of the rolls vary from a few kilometres to a few tens of kilometres and their lifespan ranges from hours to days (Drobinski and Foster, 2003). Regarding their width, Young et al. (2002) separate the rolls in two categories: narrow and wide. The ascending air masses inside the narrow rolls are one thermal wide (Weckwerth et al., 1999), whereas in the wide rolls each ascending area contains multiple thermals (Brümmer, 1999).

The remote sensing methods can provide the mlf-cs fields, by applying the Velocity Azimuth Display (VAD) method in the radial wind observations (Browning and Wexler, 1968; Drobinski et al., 2004; Lhermitte, 1962). In a previous article (Cheliotis et al., 2020), we presented our automated mlf-cs classification methodology based on texture analysis parameters (Haralick et al., 1973) and a supervised machine learning method, namely the quadratic discriminant analysis (Hastie et al., 2009) and we assessed its' performance on a training ensemble of 150 cases. The training ensemble consisted of four mlf-cs types: streaks (narrow elongated patterns associated with high wind shear near the surface), rolls (wide elongated patterns

associated with cloud streets), thermals (large enclosed patterns associated with fair weather conditions) and "others" (any other pattern that does not belong in one of the other types). Our algorithm revealed five texture analysis parameters that minimized the classification error at approximately 9% for the training ensemble. In the current study, these five texture analysis parameters were utilized for the classification, in one of the four mlf-cs types, of the 4577 lidar scans recorded in the urban area of Paris for the period 04 September 2014 – 06 November 2014. The aim of the study is to assess the quality of the classification in terms of physics/meteorology for a large dataset. In this regard, we examined the physical parameters of the classified structures for the whole ensemble. These parameters are the time of the day, the ABL height and the horizontal wind speed during the occurrence of the structures, and the sizes of the structures. Furthermore, we investigated a four-day period when streaks, rolls and thermals cases were all detected in order to retrieve information regarding the transition between the types of structures and their life cycle. Up to this point, large datasets mlf-cs classifications have been performed only visually (e.g. Yagi et al., 2015). This is the first time an automated classification has been used for a large dataset.

The main body of the manuscript is composed of three sections. In Section 2, we present the methodology of our study which includes the experimental set-up, the mlf-cs fields of our classification, the method to estimate the ABL height and the horizontal size of the structures, plus the relevant weather conditions in the study area during the study period. Section 3 contains the results, separated into two subsections, one regarding the classification of the whole ensemble and the other focusing on a peculiar four-day period. Finally, Section 4 summarizes all the main points of the study.

2. Methodology

2.1 Experimental set-up

The current study relies on radial wind observations derived from a two-months campaign (04/09-06/11/2014) performed in Paris in the framework of the project VEGILOT (VEGétation et ILOT de chaleur urbain, i.e. vegetation & urban heat island) (Klein et al., 2019). The aim of the VEGILOT campaign was to study the atmospheric chemistry and transport processes inside the urban boundary layer in Paris. For the radial wind measurements, the Leosphere WLS100 scanning Doppler lidar (www.leosphere.com) was set up atop a 75 m building in the Jussieu Campus, located in the centre of Paris city as showcased in Figure 1. The Cloud and Aerosol Micro-Lidar (CAML), manufactured by the CIMEL company (www.cimel.fr), was installed on the roof of an adjacent building (20 m height) in order to provide aerosol measurements. The maximum range for the Doppler lidar was 5 km, but the effective range was smaller for the vast majority of the data as backscatter signals with carrier-to-noise ratio (CNR) lower than -27 dB was filtered out, along with outliers in wind speed values. The red circle in Figure 1 represents the maximum horizontal scanning range for the Doppler lidar. It is important to note that despite Paris being a metropolitan city, there is a regulation with respect to the building heights in the center and therefore these do not exceed 50 m (Saint-Pierre et al., 2010). Specifically, for the scanning area, the ground altitude ranges between 30 to 60 m with some exceptions near the boundaries where there are some hills. Hence the scanned urban area is rather homogeneous. The significant lidar properties for the Doppler and the aerosol lidar during the VEGILOT campaign are displayed in Table 1.

Figure 1: The Jussieu measurement site with the scanning maximum range (5 km) of the PPI surface scan (Géoportail satellite image https://www.geoportail.gouv.fr/, last access date: 09/12/2020).

Table 1: Properties of the Doppler and the aerosol lidar utilized during VEGILOT

	Doppler lidar (Leosphere WLS100)	**Aerosol lidar (CAML)**
Altitude of lidar:	75 m a.g.l.	20 m a. g. l.
Minimum range:	100 m	200 m
Maximum range:	5 km	20 km
Spatial resolution:	50 m	15 m
Radial wind speed range:	-30 to 30 m/s	-
Laser wavelength:	1.543 µm	532 nm
Radial wind accuracy:	± 0.1 m/s	-
Accumulation time:	1 sec/beam	10 minutes

The way the Doppler operates has been described extensively by Cariou et al., (2007) It was taking measurements continuously during the two-month campaign, following a sequence of scanning methods that included quasi-horizontal scans (Plan Position Indicator or PPI), vertical beams (line of sight or LOS) and vertical profiling using the Doppler Beam Swinging (DBS). The full measurement cycle was repeated every 18 minutes. The LOS beams consisted of 10 vertical observations during 30 seconds, which average was used for the estimation of the ABL height during the night (Section 2.3). Concerning the PPI method, quasi-horizontal scans at a 1° elevation angle were carried out for azimuth angles ranging from 0° to 360° with a 2° resolution. This 1° elevation created a height difference between the central point and the boundaries of the scans equivalent to 87 m. The duration of a PPI scan was 3 minutes, which is adequate for the observation of the mlf-cs with an equivalent or longer lifespan. The DBS method was used to identify the meteorological phenomena relevant to interpret the classification result, such as low level jets. The DBS method was a combination of LOS beams in the North, East, South and West directions with a 75° elevation and a total duration of 15 seconds. This method was applied two times consecutively in each cycle.

Regarding the aerosol lidar, its' operation has been demonstrated by Pelon et al., (2008). For the current study, the aerosol lidar observations were used exclusively for the daytime estimation of the ABL height during the two-month campaign (Section 2.3). The aerosol lidar was emitting beams vertically and data were averaged on 10-minute periods. Table 2 summarizes the different scanning methods selected during the VEGILOT campaign.

125 Table 2: Scanning methods and their characteristics during VEGILOT

Lidar	Scanning area	Purpose	Elevation & azimuth angle	Scan duration
Doppler-LOS	Vertical profile	Estimation of the ABL height (night)	Elevation 90°, azimuth 0°	10 x 30 sec
Doppler-PPI	Almost horizontal scans near surface	Identification of structures	Elevation 1°, azimuth 0 to 360° with 2° resol.	3 min
Doppler-DBS	Combination of LOS	Identification of low level jet cases	Elevation 75°, azimuth 0°, 90°, 180° & 270°	2 x 15 sec
Aerosol	Vertical profile	Estimation of the ABL height (day)	Elevation 90°, azimuth 0°	10 min average

2.2 The mlf-cs fields

As we stated in Section 1, we have developed an automated methodology in order to detect and classify mlf-cs near the surface, based on single Doppler lidar observations by applying the VAD method. A detailed description of this process is presented in a peer-reviewed paper (Cheliotis et al., 2020). A characteristic example for each of the identified mlf-cs fields 130 that we aimed to classify is displayed in Figure 2.

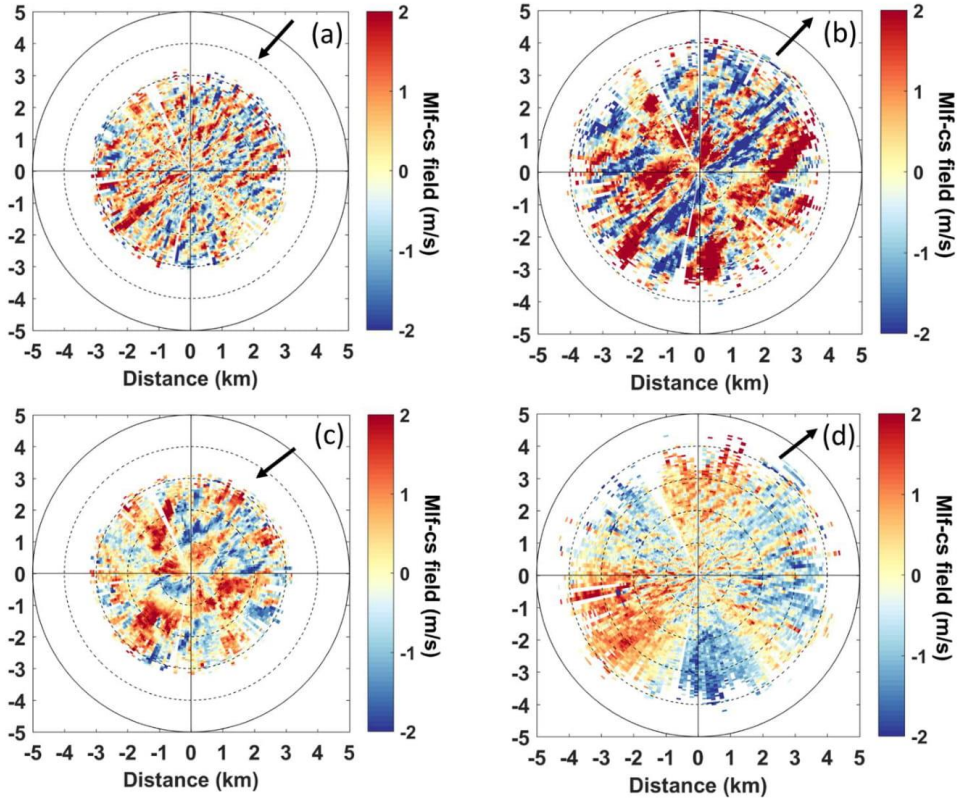


Figure 2: The mlf-cs types of our classification along with the horizontal wind direction (black arrow). (a) Streaks observed on 08/09 at 21:22. (b) Rolls observed on 09/10 at 13:24. (c) Thermals observed on 04/09 at 13:13. (d) "Others" observed on 10/10 at 03:08.

Each field showcases some distinctive characteristics. In Figure 2a, representing the streaks case, it is possible to see some narrow stripes aligned with the horizontal wind direction. Similarly, in Figure 2b, there are some wide stripes aligned with the horizontal wind direction, representing the rolls case. Figure 2c showcases a more chaotic pattern where it was still possible to identify some organized enclosed areas of positive and negative values corresponding to the thermals case. Finally, Figure 2d portrays a type that it would not be possible to classify into one of the aforementioned types and as a result it was identified as a separate type named "others". The algorithm selected five classifiers based on the texture analysis parameters of the 150 cases training ensemble. Regarding the training ensemble, the algorithm performed the best for the streaks type where 29 out of the 30 cases were classified accurately. The performance for the "others" type was also impressive as 58 out of the 60 cases were classified correctly. The rolls type did not lag far behind, with 27 out of 30 cases classified correctly and only 3 cases misclassified as thermals. Last but not least with respect to the performance, was the category thermals. However, it did not perform poorly either with 24 out of 30 cases classified correctly, 4 cases misclassified as rolls and two misclassified as "others". Apparently, these chaotic enclosed patterns of the thermals were more challenging to distinguish as the "others" type was also chaotic and the rolls type consisted of large patterns as well.

2.3 Estimation of the ABL height

The Doppler lidar vertical LOS beams and the aerosol lidar observations were used for the estimation of the ABL height during the night and during the day respectively. Klein et al. (2019) utilized the same VEGILOT observations in order to estimate the ABL height on 09/09-10/09/2014 for their study related to the surface ozone variability in Paris. The basis for the estimation of the ABL height was the identification of the maximum variance of the vertical wind speed, as the vertical wind variance is significantly larger in the transition zone between the boundary layer and the residual layer or the free troposphere (Tucker et al., 2009). However, this technique is limited to the cases when the wind profile extends above the ABL, hence it is applicable during the night and early morning, when the aerosols' concentrations in the residual layer provide a sufficient CNR for wind retrieval. During the day, the aerosols' concentrations are low in the free troposphere, thus the CNR is significantly lower than the -27 dB threshold. Therefore, Klein et al. estimated the ABL height during the day using the maximum of the variance of the range corrected signal of the aerosol lidar, a principle well described by Flamant et al., (1997) and Menut et al., (1999). This method takes advantage of the abundance of aerosol in the boundary layer in contrast to the free troposphere (Stull, 1988). For the current study we applied the same methods as Klein et al. (2019) with slight modifications, for the two-month dataset of the VEGILOT campaign. More precisely, the ABL height was estimated using the wind variance method between 0000 UTC and 0800 UTC and between 1800 UTC and 2359 UTC, while from 0800 UTC to 1800 UTC. We considered the ABL height to be the level where the minimum derivative of the standard deviation of the range corrected signal occurred because it gave better results than the maximum of the standard deviation. Figure 3 illustrates the estimated ABL height along with the vertical wind speed LOS observations on 09/09/2014. The ABL height values are in agreement with previous observations in the area (Dieudonné, 2012).

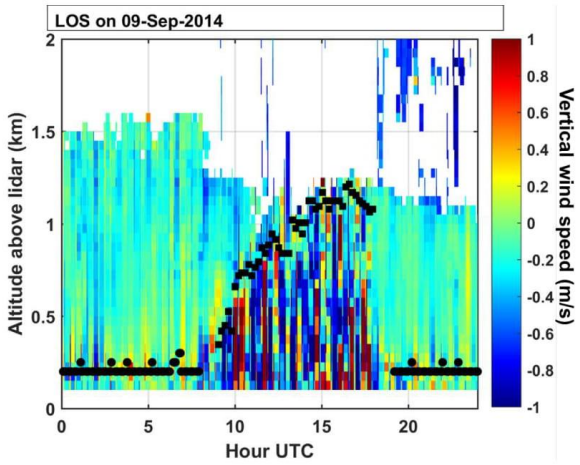


Figure 3: Time-height cross-section of the vertical wind speed observations obtained from LOS beams. The black dots represent the values of the ABL height as estimated by the Doppler wind lidar and the black squares estimated by the aerosol lidar methods on 09/09/2014.

It is important to remark that low clouds, such as stratus and stratocumulus, could inhibit the lidar signal. In such cases, the abovementioned methodology for the estimation of the ABL height was ineffective. Hence the estimation of the ABL height was performed solely for low clouds free cases. It was possible to estimate the ABL height for the entire 24-hour day (or close) for 18 out of the 64 days of the study (noted in Figure 5b). The nocturnal stable boundary layer (SBL) height values vary between 200 and 300 m during this period. Regarding the ABL height during daytime the values vary between 1.2 and 1.8 km.

2.4 Estimation of the horizontal sizes for the coherent structures

In Figure 2a-b, characteristic cases of coherent structures are displayed. The periodicity of the alternating positive and negative stripes is evident. It is possible to estimate the sizes of these periodicities by the means of the spectral analysis. The spectrum is computed by applying the Fast Fourier Transform (FFT) analysis for physical space data (Stull, 1988). Flamant et al. (1997) indicated the existence of turbulence in the spectrum peaks of the wind speed. Several studies were carried out since then, which take advantage of the spectrum peaks in order to estimate the horizontal sizes of coherent structures e. g. Calaf et al., 2013; Drobinski et al., 2004, 1998. The spectrum is calculated for the radial wind speed along the transverse to the mean wind direction. More precisely the radial wind of the opposite pair of beams closest to the transverse direction were adjoined before computing the spectrum. In order to reduce the noise, the spectrum was computed for several pairs of beams, then averaged. We observed that averaging five pairs of beams reduced the noise without blurring the peaks appearing on a single spectrum. On the other hand, an average on only three pairs of beams did not reduce the noise enough, while using seven pairs of beams smoothened the peaks significantly. As the five pairs of beams span over an 8°-wide cone, the radial wind speed values were converted from polar to Cartesian coordinates to maintain a consistent size of the grid throughout the scanning range. Figure 4a,c showcase the radial wind speed observations from the PPI scans at the same time as Figure 2a,b. Alongside the corresponding averaged spectrums (Figure 4b,d) in the transverse direction, depicted by the black lines, are displayed. Furthermore, dashed curves highlighted an adjacent pair of a stronger (positive mlf-cs) and a weaker (negative mlf-cs) radial wind speed stripe which is the examined periodicity. The effective scanning range for these cases is between 3 and 4 km, thus we can only observe periodicities within this range. The signal of the spectrum below 300 m it is very noisy and therefore excluded from the study. The peaks representing the structures are also highlighted.

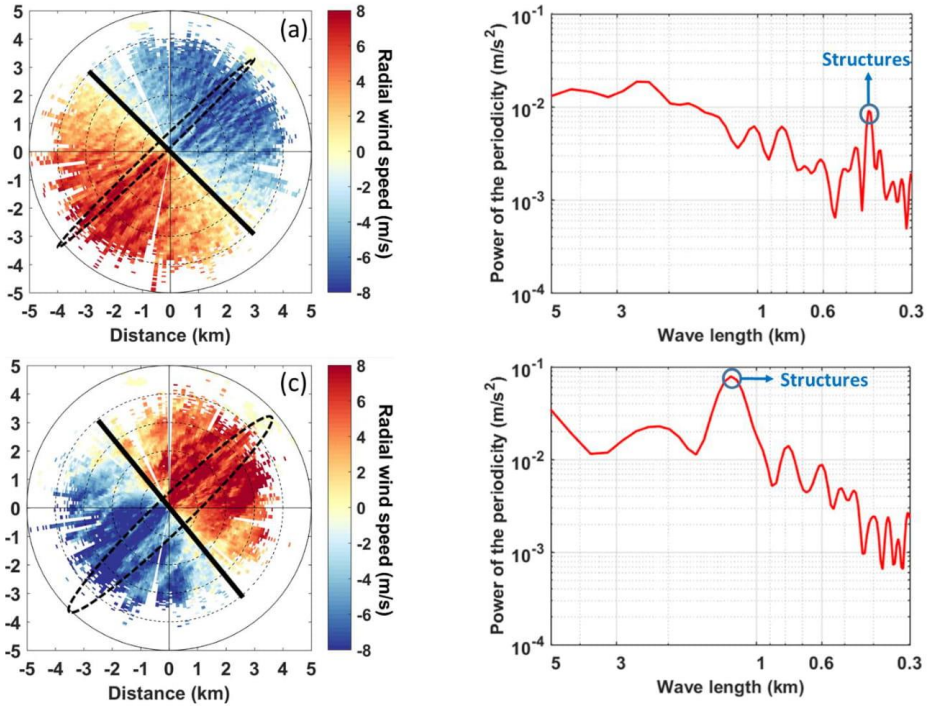


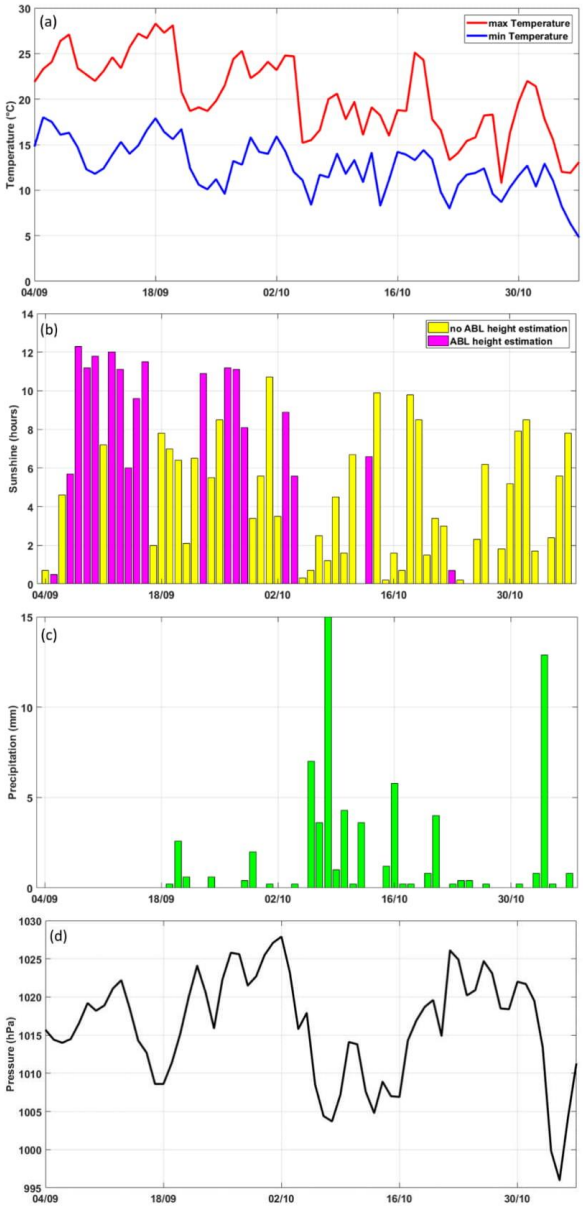
Figure 4: Radial wind speed PPI observations with the corresponding spectra (average over five pairs of beams) in the transverse direction (black line). (a-b) Streaks case observed on 08/09 at 21:22. (c-d) Rolls case observed on 09/10 at 13:24. The dashed curves contain a stronger and a weaker radial wind speed stripe (one periodicity).

Regarding the streaks case (Figure 4a-b), a prominent peak of periodicity is detected slightly above 400 m. From Figure 4a, it is visible that the dashed curve, equivalent to a stronger and a weaker radial wind speed stripes, contains an area of an approximate size of 400 m. This value is comparable with the approximate 300 m horizontal spacing of the streaks (distance between the streaks, weaker radial wind area in Figure 4a-b) observed in other Doppler lidar studies (e.g. Drobinski et al., 2004; Weckwerth et al., 1997) as well as large eddy simulation studies (e.g. Deardorff, 1972; Moeng and Sullivan, 1994). For the rolls case (Figure 4c-d), the most prominent peak is detected at approximately 1.3 km. This size is comparable to the highlighted area in Figure 4b by the dashed curve. So, if we assume that there is half-space between the rolls as well, then the size of the individual rolls should be smaller than 1 km. According to previous studies, the width of the rolls may vary (Young et al., 2002). The horizontal size of narrow mixed-layer rolls could be equivalent to one thermal wide of hundreds of meters (Weckwerth et al., 1999). Although this is a comparable size to the abovementioned case, it is tricky to compare our results to studies focusing on mixed-layer rolls since the observations of the current study were retrieved near the surface. It is also important to mention the occurrence of secondary peaks in the spectrums. Therefore, it is of the utmost importance to handpick the peaks of the spectrums by detecting equivalent size patterns in the mlf-cs fields.

2.5 Weather conditions during the period under study

For a better understanding of the classification results, it was fundamental to examine some relevant meteorological parameters during the period under study (04/09/2014-06/11/2014). The nearest meteorological station with available results for that period was the Météo France station of Paris-Montsouris, located in park, at the southern edge of inner Paris, 2.9 km south-southwest from the Jussieu site. Regarding the thermals pattern, they are expected to occur during fair cumuli weather, hence the sunshine duration and the atmospheric pressure are appropriate parameters to indicate their development.

Furthermore, the detection of precipitation events reveals the cases when the estimation of the ABL height via the lidar observations were not applicable. Figure 5 illustrates the minimum and maximum temperatures, the daily values of sunshine in hours, the daily accumulated precipitation and the daily average atmospheric pressure for the period under study. The days were the estimation of the ABL height was feasible are pointed out with the magenta coloured bars in Figure 5b.



220 Figure 5: (a) Minimum and maximum temperatures, (b) daily sunshine duration (ABL height estimation only for the magenta coloured bars), (c) daily accumulated precipitation, (d) daily average pressure for the period 04/09/2014-06/11/2014. Data are from the Météo France station Paris Montsouris.

It is apparent from Figure 5b-d that during the period under study there were many hours when fair cumuli weather conditions prevailed. So the thermals type was expected to occur frequently during our observations. It is also evident from Figure 5c that there was precipitation recorded for 30 out of the 64 days of VEGILOT. The ABL height was estimated mostly during days with abundant sunshine hours, except 05/09 and 23/10 where the total sunshine hours did not exceed two hours. During these days the aerosol lidar signal was not interrupted by clouds and therefore it was still possible to estimate the ABL height.

In Section 1, we mentioned the weather phenomena related to coherent structures. More particularly, the high wind shear near the surface is a driving factor the development of streaks, and the cloud streets are an evidence for the formation of rolls. Regarding the wind shear, the horizontal wind profile was built from the DBS soundings.

According to Stull, (1988) a nocturnal low-level jet (nllj) is characterized by a low-altitude maximum of the horizontal wind speed (jet core), with an increment of at least 2 $m \cdot s^{-1}$ compared to the wind minimum located above it. Figure 6 demonstrates a DBS profile of the horizontal wind speed with a jet core at 300 m altitude and for which it is evident that the difference in wind speed is more than 2 $m \cdot s^{-1}$.

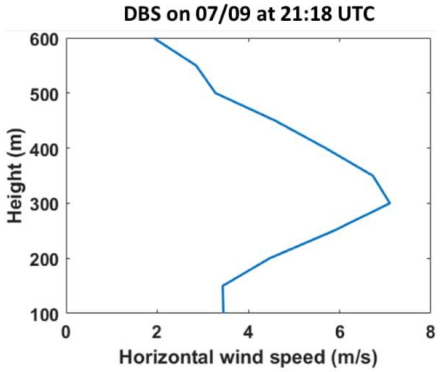


Figure 6: Vertical profile of the horizontal wind speed from the DBS scan on 07/09 at 2118 UTC.

We identified nllj events for 20 out of the 63 nights based on this description. As a result, we expected to observe several cases of streaks during the nights. The days when the nllj events were observed are displayed in Figure 7a. Regarding the cloud streets, MODIS satellite images were used for the identification of such cloud formations. Cloud streets formations were observed for 8 out of the 64 days of VEGILOT. This is a confirmation for the formation of rolls during these days. However, it is important to emphasize that the observations we analysed where near the surface, thus we observed the structures occurring near the surface during rolls formation. Figure 7b displays the days when cloud streets were observed.

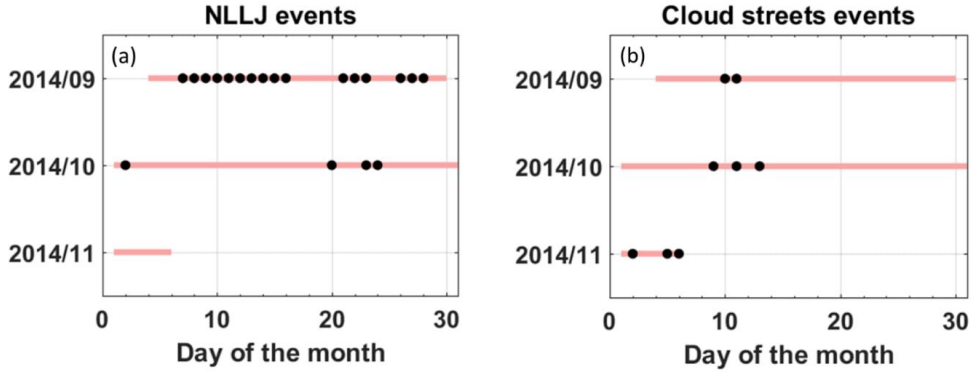


Figure 7: Days of VEGILOT when (a) nllj and (b) cloud streets events occurred.

3. Results

3.1 Classification of the two-month dataset

In Section 2.2 we briefly summarized the main points of our classification and indicated the classification results for the training ensemble made by the QDA algorithm. The classification error was minimized at approximately 9% for five texture analysis parameters (Cheliotis et al., 2020). We thereupon utilized the QDA algorithm using these five texture analysis parameters in order to classify the whole ensemble consisted of 4577 PPI scans. Table 3 displays the outcome of the classification. More specifically, the number of occurrences for each type and the corresponding percentage over the whole ensemble are presented. Apart from the total number of occurrences, we included the individual number of occurrences for the following time periods 0000-0800 UTC (late night/early morning), 0800-1800 UTC (daytime) and 1800-2359 UTC (evening/early night) in order to examine whether the results are realistic.

Table 3: Classification of the structure types for the whole PPI scan ensemble, with a distinction between daytime (0800-1800), evening/early night (1800-2359) and late night (0000-0800).

Classified cases / Type	0000-0800 UTC	0800-1800 UTC	1800-2359 UTC	Total
Streaks	376 (8.2%)	394 (8.6%)	375 (8.2%)	1145 (25%)
Rolls	27 (0.6%)	365 (8%)	28 (0.6%)	420 (9.0%)
Thermals	98 (2%)	695 (15%)	107 (3%)	900 (20%)
"Others"	1036 (22.6%)	438 (9.6%)	638 (13.9%)	2112 (46.0%)
				4577 (100.0%)

As expected the category "others" was dominant with almost half of the ensemble (46%) classified as such. The second most frequent type was the streaks, with 25%. In Section 2.5, we discussed the frequent occurrences of nllj during VEGILOT (approximately 1/3 nights), hence the frequent occurrence of streaks was also expected. The thermals type accounted for 20% of the total cases. An expected outcome as well, since the duration of sunshine over Paris was long during the 2-month period under study, as portrayed in Figure 5d-f. Finally, 9% of the cases were classified as rolls. The cloud streets cases were not frequent during VEGILOT, so the rolls were realistically classified as the least frequent type. However, the most striking result regarding rolls is the scarcity of the classification during night. This is a remarkable result as the five texture analysis parameters were the only classifiers used by the algorithm and not the time of the day. The same outcome ensues for the thermals type to a lesser extent. The number of cases classified as thermals during day-time outweighs the one classified during night time. However, the number of "night-time thermals" is not negligible. As we explained in Section 2.2, this was due to the random nature of the patterns in thermals that in some cases resembles the "others" category.

Another interesting outcome from Table 3 is that the vast majority of the "others" cases were classified during night time. As we stated in Section 1, the estimation of the mlf-cs fields via the VAD method requires a homogeneous mean wind field resulting in a symmetrical radial wind field. A non-symmetric radial wind field will lead to a bad case of a mlf-cs field that is included in the "others" type. The most common cause for a non-symmetric radial wind field is the varying wind direction during calm or low wind conditions (Wilson et al., 1976). It was therefore interesting to examine the horizontal wind speed distribution for the patterns type. Figure 8 demonstrates the horizontal wind speed distribution through histograms for the number of occurrences for each wind speed group. It confirms that the vast majority of cases, more than 1500 (75 %), characterized by calm (<2 $m \cdot s^{-1}$) or low wind conditions (2-4$m \cdot s^{-1}$) were classified as "others". Furthermore, there are no cases with calm wind conditions classified as streaks or rolls. The cases classified as streaks and rolls were mainly found during moderate or high wind conditions (4-8 $m \cdot s^{-1}$) with approximately 700 (61 %) and 330 (79 %) cases respectively.

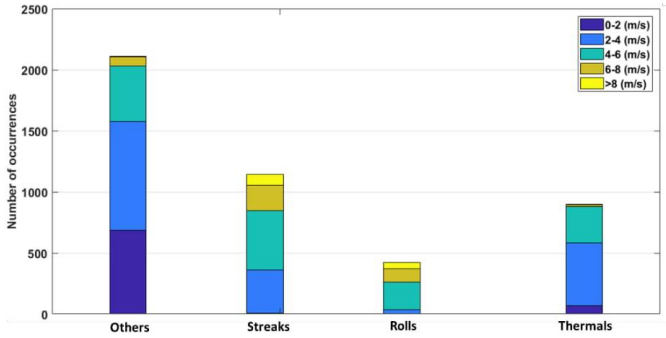


Figure 8: Histogram with the horizontal wind speed distributions for each pattern type.

Table 3 presents the classification of the types for specific time periods in order to examine whether the classification was realistic. This idea was extended by associating the classified cases with the estimated ABL height for the 18 days when the estimation was possible for the majority of the observations. The time difference between the PPI scans for the estimation of the mlf-cs field and the LOS vertical scans was approximately 10 minutes. Regarding the PPI and the aerosol lidar scans, the time difference was few minutes. These time differences are relatively short and therefore the ABL height was expected to remain steady. In Figure 9, the distribution of the ABL height in relation to the classified pattern is portrayed via histograms. The ABL height was divided into three subranges to represent the SBL (0-350 m), the developing ABL (350-1000 m) and the well-developed ABL (1000-2000 m).

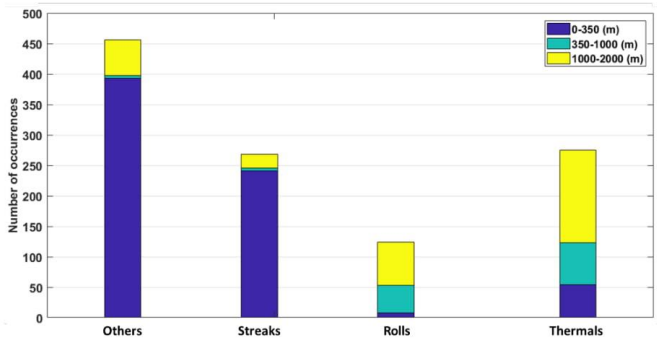


Figure 9: Histogram with the ABL height distributions for each patterns type.

It is apparent from Figure 9 that there were only 8 cases out of the 124 (6 %) classified as rolls concurrently with the existence of the SBL. Almost all of the classified rolls cases correspond to the developing or the well-developed ABL. Moreover, the classified “others” cases during the SBL conditions dominated the number of occurrences, close to 400 cases (86 %), as expected and analysed from Table 3. During 13 out of the 18 days for which we estimated the ABL height, we observed a nllj. Therefore, the substantial number of streaks occurrences, more than 240 cases (90 %), during the SBL conditions was expected as well. Finally, despite the majority of the thermals cases being classified during the developing or the well-developed SBL, there was also a significant amount during the SBL conditions exceeding 50 cases (20 %). These misclassifications were associated with the randomness of the patterns as we stated before and not with the atmospheric physics. The possible addition of more/other texture analysis parameters could improve the classifications of the thermals.

3.2 Mlf-cs during a peculiar four-day period

We examined further the results of the classification by focusing on a period when the three mlf-cs cases (streaks, rolls and thermals) have been detected (07/09-10/09). These days were characterized by moderate wind conditions and only few low wind cases. This was important, since as we have previously stated, the mlf-cs fields were constructed by utilizing the VAD method that is not applicable under low wind conditions. Figure 10 presents the wind rose for the period 07/09-10/09 computed from the PPI scans using the VAD method. The wind speed values ranged between 4 and 8 m/s for more than 70% of the cases. The wind direction was north-easterly during most of the period.

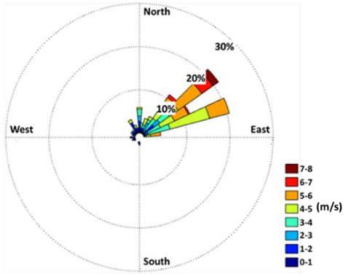


Figure 10: Wind rose for the period 07/09-10/09 for the horizontal wind extracted from the PPI observations by utilizing the VAD method.

During this period there were anti-cyclonic conditions over Paris. Additionally, as we can see in Figure 5c, there was no precipitation during this four-day period, but instead plenty of sunshine hours (Figure 5b). Except on the 07/09 when the sunshine duration was approximately 6 hours, for the 08/09, 09/09 and 10/09, the sunshine duration exceeded 12 hours. This allowed a reliable estimation of the ABL height for these days. In Figure 11(a), the ABL height is displayed along with the time-height cross sections of the horizontal wind speed from the DBS soundings. In order to examine the classification, each type is represented with different colour and symbol. The first element to notice in Figure 11(a) was the accurate representation of the ABL height development throughout the four-day period. The ABL height ranged between 200 and 300 m during the night and 1.2 to 1.8 km during the day. Furthermore, the classification of the cases was realistic with respect to the time and the ABL height for almost the entire four-day period. Only four cases were misclassified in that regard with three cases of thermals found on 07/09 between 2230 UTC and 2330 UTC and one case of thermals around 0300 UTC on 09/09. Regarding rolls, no cases were found on 07/09 which is the day with the least sunshine hours. Most of the rolls cases were found on 10/09. Since there was no weather system affecting the region at this date, these structures developed mainly due to the thermal instabilities. The classification of the streaks cases was particularly interesting. It is obvious in all four days that the streaks are associated with a high wind shear near the surface. The streaks appear for the first time around 1900 UTC simultaneously with the wind shear presence near the surface. The missing ABL height values on 08/09 between 0000 UTC and 0800 UTC and between 1800 and 2000 UTC were due to the absence of LOS vertical wind speed observations and therefore the structure type is plotted at the 0 m altitude. Other missing points were due to the lack of PPI observations.

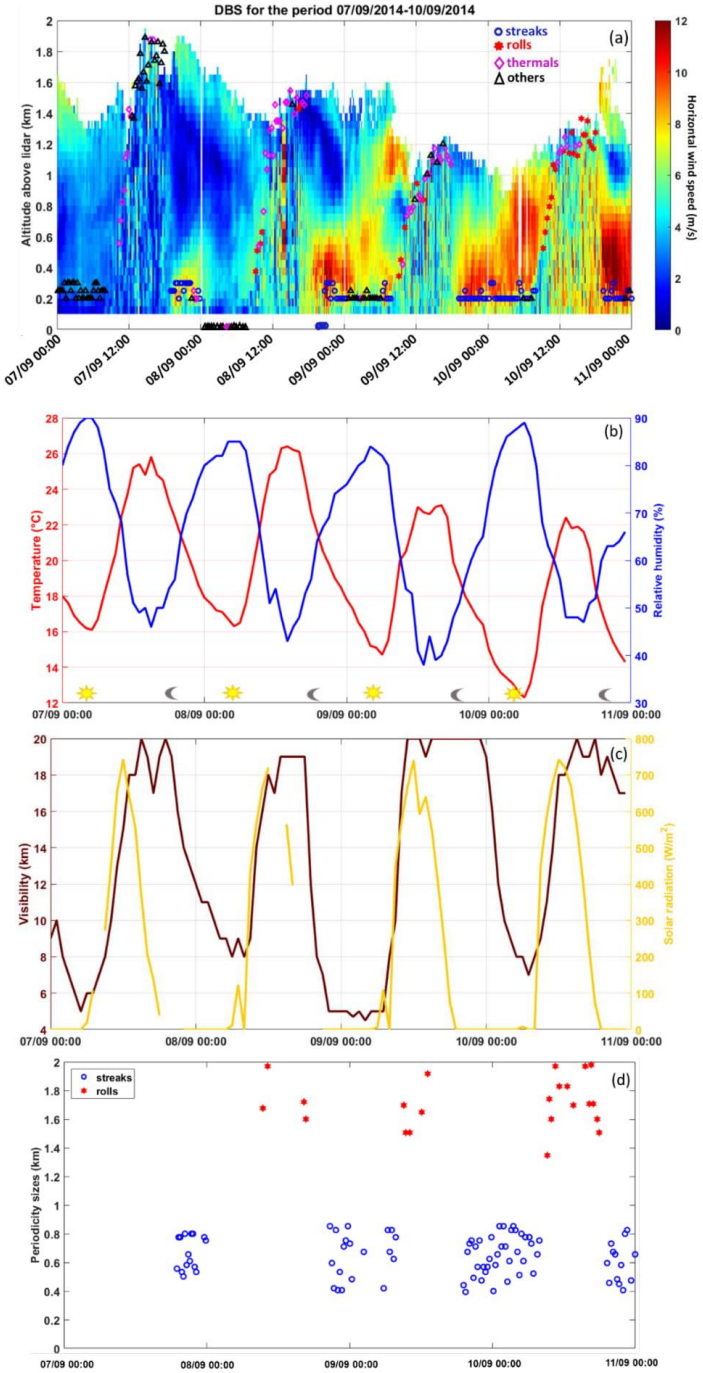


Figure 11: For the period 07/09/2014-10/09/2014: (a) The horizontal wind speed derived from the DBS observations along with the ABL height points representing the different ty of patterns with the points plotted at zero altitude correspond to missing ABL height data. (b-c) Hourly measurements of the temperature and relative humidity and visibility and solar radiation respectively observed at the Paris-Montsouris station. (d) Periodicity sizes for the streaks (blue rings) and rolls cases (red stars).

Additionally, there was a shift in the types of patterns in the early morning on 09/09 and 10/09 from streaks to rolls, as soon as the wind shear dissipated and the ABL height began to develop. We examined the meteorological parameters for the period 07/09/2014-10/09/2014, in order to identify the possible changes during the shift in the types of patterns. The temperature, relative humidity, visibility and solar radiation observed at the Paris-Montsouris station are displayed in Figure 11b,c. An interesting aspect to encounter was the reduced visibility of 5 and 7 km during the early morning on 09/09 and 10/09 respectively (Figure 11c). The Paris-Montsouris station recorded fog events only at 0200 and 0400 UTC. The reduced visibility signifies possible convections near the surface with the time of the occurrence indicating radiation fog events (Roach et al., 1976). In Section 1 we referred to studies which observed the streaks formed during fog events. As the surface heat increased in the morning the smaller structures were replaced by larger ones. The temperature increase during the three-hour period from 0600 UTC to 0900 UTC was at the magnitude of 6 °C which resulted in a drop of relative humidity of 40 % for both days.

In order to reveal possible periodicities for these structures and their corresponding sizes we computed the spectrums of the radial wind speed in the transverse direction as explained in Section 2.4. Figure 11d contains the values of the highest peaks of the spectra representing the sizes of the periodicities for the streaks and rolls cases. The sizes of the periodicities differ between the streaks and rolls cases by a magnitude of at least 500 m. In Section 2.4, we have already stated that the size of the periodicity referred to the combined red (positive) and blue (negative) stripes (Figure 2a,b). For the streaks cases, the size ranged from 400 to 800 m. So the width of the stripes would be approximately 200 to 400 m respectively. These sizes, as showcased in Section 2.4, are in agreement with other studies where the spacing was estimated to be around 300 m (Drobinski et al., 2004; Weckwerth et al., 1997b). Since the lifespan of the streaks is several minutes and the quasi horizontal PPI scans in our study had an approximate 18 minutes' time difference, most of the streaks we observed dissolved and were replaced by new ones before the next scan occurred, thus we did not observe the evolution of the same structures. Regarding the rolls cases, the sizes ranged from 1.3 km to 2 km therefore the size of the structures would probably be close to 600 m to 1 km. Since we observed the structures only near the surface, it is complicated to compare these results to other studies focusing on the rolls structures. The low values of relative humidity indicated that these were dry rolls cases. Concerning the rolls, it is also important to note that, for 3 cases out of the 31 it was not possible to estimate the size of the periodicities from the spectrum, which also led to a misclassification of these cases by the algorithm.

4. Conclusions

In the current study we examined the performance of an automated classification method which we developed in order to classify mlf-cs based on Doppler lidar observations recorded during two months over Paris. We intended to identify and classify streaks (narrow stripes), rolls (wide stripes), thermals (enclosed areas) and "others" (impossible to classify). In our previous work, we had presented the algorithm itself and assessed its' performance on the training ensemble: a classification error of approximately 9 % for a training ensemble of 150 cases with five texture analysis parameters as classifiers. We then applied these five parameters to the full dataset of 4577 quasi-horizontal scans and in this paper, we analyse the results of the classification from the physical/meteorological point of view.

The results of the classification showed that the most frequent type was the category "others" with 46 %, then the streaks type with 25 %, the third frequent type was the thermals with 20 % and finally the rolls with 9 %. By examining the occurrence of each type for three different time periods of the day, in particular 0000 to 0800 UTC, 0800 to 1800 UTC and 1800 to 2359 UTC it was possible to examine the distribution of the types during the day. The results showed a realistic distribution as rolls and thermals cases were mainly classified during the day. Streaks cases were mostly classified during the night which is an expected result as nllj events occurred during 20 out of the 63 nights of the study period. Regarding the category "others" the results were also expected, as the majority of the cases occurred during the night, when the low winds

result in asymmetric radial wind fields and the VAD method cannot be applied. This observation was confirmed by the distribution of the classified cases with regards to the horizontal wind speed: for 75 % of the "others" occurred when the horizontal wind speed was lower than 4 $m \cdot s^{-1}$. On the contrary, for streaks and rolls which are driven by wind shear, the majority of the cases were characterized by moderate or high winds (4–10 $m \cdot s^{-1}$) for approximately 67 % and 89 % of the cases cases respectively. The accurate representation of the types as a function of the time of day was further supported by the distribution of the estimated ABL height. For the 18 days, during which it was possible to estimate the ABL height for the majority of the 24 hours, the vast majority of the rolls cases occurred in a developing or developed ABL, whereas the vast majority of streaks and "others" cases were found in the SBL or the nocturnal boundary layer. The thermals had the worst performance but still 80% of the cases were found in the developed or developing ABL.

The investigation for a four-day period, when the three types of structures were classified, showed that the streaks occur at the same time as nllj and dissolve during the morning when the thermal instabilities prevail. During the two days with more than 12 sunshine hours, which resulted in strong thermal instabilities, rolls cases were classified. The spectrum of the radial wind speed in the direction transverse to the horizontal wind vector was estimated in order to detect periodicities during the two days when the streaks cases were followed by rolls. The spectra exhibited prominent peaks for almost all but three rolls cases. The periodicities of the streaks and rolls were clearly separated, with at least a 500 m difference. Regarding the streaks, the sizes of periodicities ranged between 400 and 800 m; and for rolls, they were between 1.3 and 2.0 km, indicating realistic sizes of the structures that would be around 200 and 400 m for the streak spacing and more than 1 km for the rolls.

Overall the performance of the algorithm with regards to the classification of the structures based solely on texture analysis parameters is in agreement with the physical parameters. The results were as expected, especially for the coherent structures rolls and streaks. However, there is room for improvement, especially for the category thermals.

Data availability

All lidar data used in the study are property of the Laboratoire de Physico-Chimie de l'Atmosphere (LPCA), Dunkirk, France and are not publicly available.

Acknowledgements

The authors thank François Ravetta, Jacques Pelon, Gilles Plattner and Amelie Klein of the LATMOS, Sorbonne University, Paris, for organizing and carrying out the VEGILOT campaign.

We acknowledge the use of imagery from the NASA Worldview application (https://worldview.earthdata.nasa.gov/, last access: 2 December 2020), part of the NASA Earth Observing System Data and Information System (EOSDIS).

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).

This work is a contribution to the CPER (Contrat de Plan Etat-Région) research project IRenE (Innovation et Recherche en Environnement) and Climibio. The work is supported by the French Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation, the region Hauts-de-France and the European Regional Development Fund. The work is also 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; contract no. ANR-11-LABX-0005-01) and by the regional council of Nord-Pas-de-Calais and the European Regional Development Fund.

This study was funded by the RFBR (Russian Foundation for Basic Research; project no. 20-07-00370) and Moscow Center for Fundamental and Applied Mathematics (Agreement 075-15-2019-1624 with the Ministry of Education and Science of the Russian Federation; MESRF).

References

Bartello, P, Métais, O, Lesieur, M (1994) Coherent structures in rotating three-dimensional turbulence. J Fluid Mech 273: 1–29, https://doi.org/10.1017/S0022112094001837

Barthlott, C, Drobinski, P, Fesquet, C, Dubos, T, Pietras, C (2007) Long-term study of coherent structures in the atmospheric surface layer. Boundary-Layer Meteorol 125: 1–24, https://doi.org/10.1007/s10546-007-9190-9

Bernard-Trottolo, S, Campistron, B, Druilhet, A, Lohou, F, Saïd, F (2004) TRAC98: Detection of coherent structures in a convective boundary layer using airborne measurements. Boundary-Layer Meteorol 111: 181–224, https://doi.org/10.1023/B:BOUN.0000016465.50697.63

Brient, F, Couvreux, F, Villefranque, N, Rio, C, Honnert, R (2019) Object-oriented identification of coherent structures in large eddy simulations: importance of downdrafts in stratocumulus. Geophys Res Lett 46: 2854–2864, https://doi.org/10.1029/2018GL081499

Browning, KA, Wexler, R (1968) The determination of kinematic properties of a wind field using Doppler radar. J Appl Meteorol 7: 105–113, https://doi.org/10.1175/1520-0450(1968)007<0105:tdokpo>2.0.co;2

Brümmer, B (1999) Roll and cell convection in wintertime Arctic cold-air outbreaks. J Atmos Sci 56: 2613–2636, https://doi.org/10.1175/1520-0469(1999)056<2613:RACCIW>2.0.CO;2

Brümmer, B, Bakan, S, Hinzpeter, H (1985) Kontur: Observations of cloud streets and open cellular structures. Dyn Atmos Ocean 9: 281–296, https://doi.org/10.1016/0377-0265(85)90024-7

Calaf, M, Hultmark, M, Oldroyd, HJ, Simeonov, V, Parlange, MB (2013) Coherent structures and the k-1 spectral behaviour. Phys Fluids 25: 125107, https://doi.org/10.1063/1.4834436

Cariou, JP, Parmentier, R, Valla, M, Sauvage, L, Antoniou, I, Courtney, M (2007) An innovative and autonomous 1.5 µm Coherent lidar for PBL wind profiling. In: Proceedings of 14th Coherent Laser Radar Conference.

Cheliotis, I, Dieudonné, E, Delbarre, H, Sokolov, A, Dmitriev, E, Augustin, P, Fourmentin, M (2020) Detecting turbulent structures on single Doppler lidar large datasets: an automated classification method for horizontal scans. Atmos Meas Tech 16: 6579–6592, https://doi.org/10.5194/amt-2020-82

Deardorff, JW (1972) Numerical investigation of neutral and unstable planetary boundary layers. J Atmos Sci 29: 91–115, https://doi.org/10.1175/1520-0469(1972)029<0091:nionau>2.0.co;2

Dieudonné, E (2012) Multi-instrumental analysis of the influence of boundary layer depth variability on the vertical distribution of nitrogen oxides in Paris region. Ph.D. thesis, Pierre & Marie Curie University (Paris, France). In French. Available online: https://www.theses.fr/2012PA066017

Drobinski, P, Brown, RA, Flamant, PH, Pelon, J (1998) Evidence of organized large eddies by ground-based Doppler lidar, sonic anemometer and sodar. Boundary-Layer Meteorol 88: 343–361, https://doi.org/10.1023/A:1001167212584

Drobinski, P, Carlotti, P, Newsom, RK, Banta, RM, Foster, RC, Redelsperger, J-L (2004) The structure of the near-neutral atmospheric surface layer. J Atmos Sci 61: 699–714, https://doi.org/10.1175/1520-0469(2004)061<0699:TSOTNA>2.0.CO;2

Drobinski, P, Foster, RC (2003) On the origin of near-surface streaks in the neutrally-stratified planetary boundary layer. Boundary-Layer Meteorol 108: 247–256, https://doi.org/10.1023/A:1024100125735

Eymard, L, Weill, A (1988) Dual Doppler radar investigation of the tropical convective boundary layer. J Atmos Sci 45: 853–864, https://doi.org/10.1175/1520-0469(1988)045<0853:DDRIOT>2.0.CO;2

Flamant, C, Pelon, J, Flamant, PH, Durand, P (1997) Lidar determination of the entrainment zone thickness at the top of the unstable marine atmospheric boundary layer. Boundary-Layer Meteorol 83: 247–284, https://doi.org/10.1023/A:1000258318944

Han, BS, Baik, JJ, Kwak, KH (2019) A preliminary study of turbulent coherent structures and ozone air quality in Seoul using

the WRF-CMAQ model at a 50 m grid spacing. Atmos Environ 218: 117012, https://doi.org/10.1016/j.atmosenv.2019.117012

Haralick, RM, Dinstein, I, Shanmugam, K (1973) Textural Features for Image Classification. IEEE T Syst Man Cyb 6: 610–621, https://doi.org/10.1109/TSMC.1973.4309314

Hastie, T, Tibshirani, R, Friedman, J (2009) The elements of statistical learning: Data mining, inference, and prediction, The elements of statistical learning: Data mining, inference, and prediction, Springer Series in Statistics, Springer, New York, USA.

Jiang, P, Wen, Z, Sha, W, Chen, G (2017) Interaction between turbulent flow and sea breeze front over urban-like coast in large-eddy simulation. J Geophys Res Atmos 122: 5298–5315, https://doi.org/10.1002/2016JD026247

Khanna, S, Brasseur, JG (1998) Three-dimensional buoyancy and shear-induced local structure of the atmospheric boundary layer. J Atmos Sci 55: 710–743, https://doi.org/10.1175/1520-0469(1998)055<0710:TDBASI>2.0.CO;2

Klein, A, Ravetta, F, Thomas, JL, Ancellet, G, Augustin, P, Wilson, R, Dieudonné, E, Fourmentin, M, Delbarre, H, Pelon, J, (2019) 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, https://doi.org/10.1016/j.atmosenv.2018.10.009

Kubat, M (2017) An introduction to machine learning, Springer International Publishing, Springer, New York, USA, https://doi.org/10.1007/978-3-319-63913-0

Kuettner, JP (1971) Cloud bands in the earth's atmosphere: observations and theory. Tellus 23: 404–426, https://doi.org/10.3402/tellusa.v23i4-5.10519

LeMone, M, (1972) The structure and dynamics of the horizontal roll vortices in the planetary boundary layer. J Atmos Sci: 30, 1077–1091, https://doi.org/10.1175/1520-0469(1973)030<1077:tsadoh>2.0.co;2

Lhermitte, RM (1962) Note on wind variability with Doppler radar. J Atmos Sci 19: 343–346, https://doi.org/10.1175/1520-0469(1962)019%3C0343:NOWVWD%3E2.0.CO;2, 1962.

Lin, CL, Xia, Q, Calhoun, R (2008) Retrieval of urban boundary layer structures from Doppler lidar data. Part II: Proper orthogonal decomposition. J Atmos Sci 65: 21–42, https://doi.org/10.1175/2007JAS2329.1

Lohou, F, Druilhet, A, Campistron, B, 1998. Spatial and temporal characteristics of horizontal rolls and cells in the atmospheric boundary layer based on radar and in situ observations. Boundary-Layer Meteorol 89: 407–444, https://doi.org/10.1023/A:1001791408470

Lothon, M, Lenschow, DH, Mayor, SD (2006) Coherence and scale of vertical velocity in the convective boundary layer from a Doppler lidar. Boundary-Layer Meteorol 121: 521–536, https://doi.org/10.1007/s10546-006-9077-1

MacDonald, M, Kurowski, MJ, Teixeira, J (2020) Direct numerical simulation of the moist stably stratified surface layer: Turbulence and fog formation. Boundary-Layer Meteorol (175): 343–368, https://doi.org/10.1007/s10546-020-00511-2

Martin, T, Bakan, S (1991) Airplane investigation of a case of convective cloud bands over the North Sea. Boundary-Layer Meteorol 56: 359–380: https://doi.org/10.1007/BF00119212

Menut, L, Flamant, C, Pelon, J, Flamant, PH (1999) Urban boundary-layer height determination from lidar measurements over the Paris area. Appl Opt 38: 945-954, https://doi.org/10.1364/ao.38.000945

Moeng, CH, Sullivan, PP (1994) A comparison of shear and buoyancy-driven planetary boundary layer flows. J Atmos Sci 51: 999–1022, https://doi.org/10.1175/1520-0469(1994)051<0999:acosab>2.0.co;2

Nakanishi, M (2000) Large-eddy simulation of radiation fog. Boundary-Layer Meteorol 94: 461–493, https://doi.org/10.1023/A:1002490423389

Pelon, J, Mallet, M, Mariscal, A, Goloub, P, Tanré, D, Bou Karam, D, Flamant, C, Haywood, J, Pospichal, B, Victori, S (2008) Microlidar observations of biomass burning aerosol over Djougou (Benin) during African monsoon multidisciplinary analysis special observation period 0: dust and biomass-burning experiment. J. Geophys. Res. 113, D00C18, https://doi.org/10.1029/2008jd009976

Rabin, RM, Doviak, RJ, Sundara-Rajan, A (1982) Doppler radar observations of momentum flux in a cloudless convective layer with rolls. J Atmos Sci (39): 851–863, https://doi.org/10.1175/1520-0469(1982)039<0851:DROOMF>2.0.CO;2

Roach, WT, Brown, R, Caughey, SJ, Garland, JA, Readings, CJ (1976) The physics of radiation fog: I - a field study. Q J R Meteorol Soc 102: 313–333, https://doi.org/10.1002/qj.49710243204

Saint-Pierre, C, Becue, V, Diab, Y, Teller, J (2010) Case study of mixed-use high-rise location at the Greater Paris scale. WIT Trans Ecol Environ 129: 251–262, https://doi.org/10.2495/SC100221

Smedman, AS (1991) Occurrence of roll circulations in a shallow boundary layer. Boundary-Layer Meteorol 57: 343–358, https://doi.org/10.1007/BF00120053

Stull, RB (1988) An introduction to boundary layer meteorology. Kluwer Academic Publishers, Springer, Dordrecht, Germany, https://doi.org/10.1007/978-94-009-3027-8

Träumner, K, Damian, T, Stawiarski, C, Wieser, A (2015) Turbulent structures and coherence in the atmospheric surface layer. Boundary-Layer Meteorol 154: 1-25 https://doi.org/10.1007/s10546-014-9967-6

Tucker, SC, Brewer, WA, Banta, RM, Senff, CJ, Sandberg, SP, Law, DC, Weickmann, AM, Hardesty, RM (2009) Doppler lidar estimation of mixing height using turbulence, shear, and aerosol profiles. J Atmos Ocean Technol 26: 673-688, https://doi.org/10.1175/2008JTECHA1157.1

Weckwerth, TM, Grund, CJ, Mayor, SD (1997a) Linearly-Organized Coherent Structures in the Surface Layer, In: 12th Symposium on Boundary Layers and Turbulence. American Meteorological Society, Vancouver, BC, Canada, pp. 22–23.

Weckwerth, TM, Horst, TW, Wilson, JW (1999) An observational study of the evolution of horizontal convective rolls. Mon Weather Rev 127: 2160–2179, https://doi.org/10.1175/1520-0493(1999)127<2160:AOSOTE>2.0.CO;2

Weckwerth, TM, Parsons, DB (2006) A review of convection initiation and motivation for IHOP_2002. Mon Weather Rev 134: 5-22, https://doi.org/10.1175/MWR3067.1

Wilson, RB, Start, GE, Dickson, CR, Ricks, NR (1976) Diffusion under low windspeed conditions near Oak Ridge, Tennessee, NOAA Technical Memorandum ERL ARL-61: 83.

Yagi, A, Inagaki, A, Kanda, M, Fujiwara, C, Fujiyoshi, Y (2015) Analysis of spacing of streaky structures within surface layer above real urban, In: 9th International Conference on Urban Climate Jointly with 12th Symposium on the Urban Environment. Toulouse, France.

Young, GS, Kristovich, DAR, Hjelmfelt, MR, Foster, RC (2002) Rolls, streets, waves, and more: A review of quasi-two-dimensional structures in the atmospheric boundary layer. Bull Am Meteorol Soc 83: 997-1002, https://doi.org/10.1175/1520-0477(2002)083<0997:RSWAMA>2.3.CO;2

Zhang, J, Liu, P, Zhang, F, Song, Q (2018) CloudNet: Ground-based cloud classification with deep convolutional neural network. Geophys Res Lett 45: 8665–8672, https://doi.org/10.1029/2018GL077787

3.2.3 Additional discussion regarding the physical properties of the coherent structures

The association of the available physical properties with the structures revealed a realistic classification. The time, the mean wind speed, the ABL height (when available), the horizontal sizes of the coherent structures (when available) and the occurrence of particular weather events (nllj) were associated to the appropriate structures as expected. It is important to mention that for the coherent structures, when the texture analysis parameters were plotted against the neighbor order in the transverse direction, they showed a peak for a distance comparable to the size retrieved by the spectrum. Figure 40 showcases an example of the amplitude of the homogeneity curve plotted against the neighbor order n for the pair orientation φ=0°, as defined in Section 1.3.1, for a case of streaks and for a case of rolls. For the example of streaks (Figure 40a) there is a peak at the 3rd neighbor order, meaning at 150 m distance. For the example of rolls (Figure 40b), the peak appears for the 14th neighbor order, equivalent to 700 m. These peaks are indications that the horizontal sizes could be retrieved from such figures, but more testing is necessary in order to find which texture analysis parameter is more suitable for an accurate estimation of the horizontal sizes.

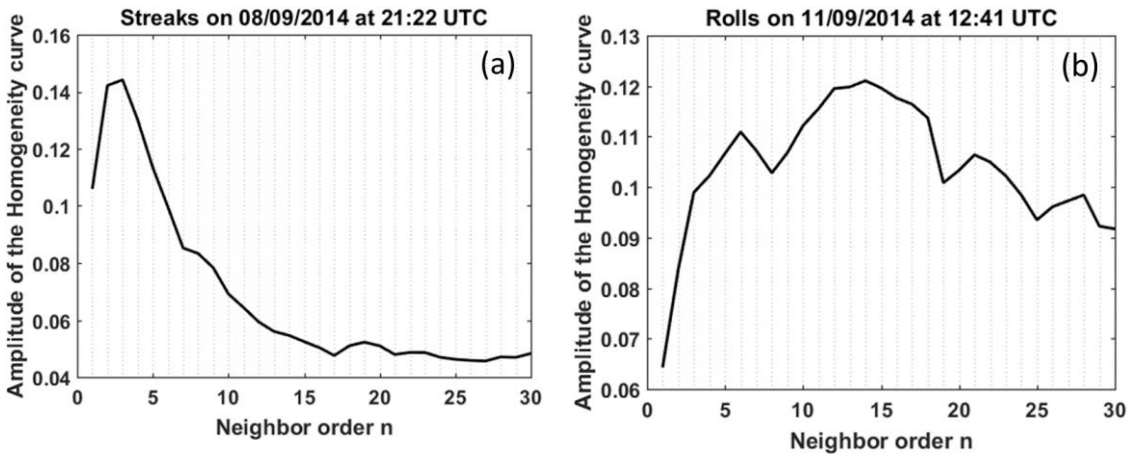


Figure 40: The amplitude of the homogeneity curve plotted as a function of the neighbor order in the transverse direction (cell pair orientation φ=0°) for a case of streaks (a) and for a case of rolls (b).

It is also important to note that the estimation of the nocturnal ABL height was limited due to the addition of the height of the building where the lidar was installed and the minimum range of the lidar. These can cause some overestimation of the height of shallow ABL. However, for the purpose of this study an approximate value for the height was sufficient as the aim was to reveal the distributions of the cases with respect to the ABL evolution. It is particularly interesting that during the two-month period under study, coherent structures formations seemed to be frequent over Paris (1145 streaks and

420 rolls cases out of the 4577 scans). An observation also made by Barthlott et al. (2007) for a 10-month period study using sonic anemometer measurements 20 km southwest from Paris.

As stated in this manuscript, regarding the streaks cases, the wind shear near the surface plays a significant role in their formation. Under a llj, the shear can be defined as the horizontal wind gradient between the near-surface and jet nose, but this definition cannot be applied all the time and used in order to examine the association with the streaks cases. Therefore, it was decided to define the near-surface wind shear as the horizontal wind speed (Eq. (8) difference between the lidar level (~70 m a.g.l.) and 400 m above it, computed using the DBS observations. The 400 m level was selected as it was the altitude where the maximum horizontal wind speed of the llj was approximately located during the period 07/09 – 11/09. Figure 41 shows the distribution of the streaks cases according to the aforementioned wind gradient. The results show more than about 40% of the streak cases occurred under low wind differences (≤ 3 $m.s^{-1}$). The number of occurrences then declined for wind speed differences higher than 3 $m \cdot s^{-1}$, with very few cases corresponding to wind speed difference above 10 $m \cdot s^{-1}$. Further analysis is necessary in order to extract conclusions regarding the association of the wind shear and the streaks cases. Nevertheless, these preliminary results are an indication that a net majority of the classified cases were related to a significant wind shear near the surface as about 72% of the streaks cases were associated with wind speed differences between the lidar level and 400 m above it higher than 2 $m \cdot s^{-1}$.

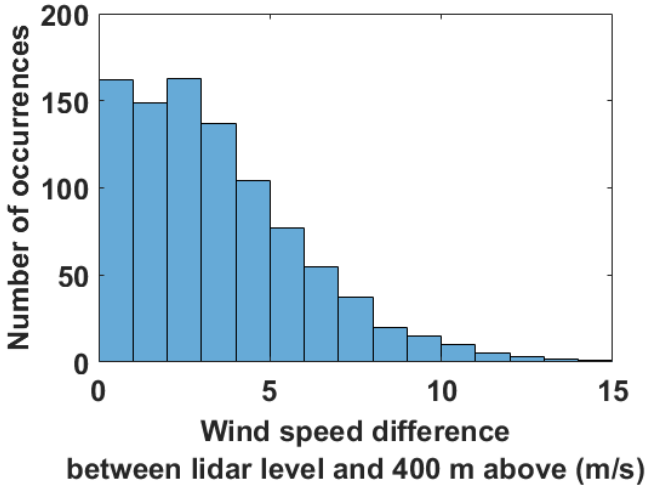


Figure 41: Distribution of streaks cases based on the horizontal wind speed difference between the lidar level and 400 m above it as recorded by the DBS scans.

A two-month or a longer dataset classification provides the ability to perform statistical studies. In the future, monthly, seasonal or annual variations could be estimated

and analyzed in order to reveal possible modifications of the structures throughout the year. This type of study has not been done yet based on lidar measurements, since it would have required the subjective identification of the structures by the researcher, which is a cumbersome process. A consistent automated classification like the one presented in this study can overcome this issue and allow the statistical estimation of structure parameters. For this purpose, it would be crucial to combine the lidar measurements with other meteorological instruments, such as a sonic anemometer in order to validate the wind lidar observations and estimate the atmospheric stability and vertical momentum fluxes, or a water vapour lidar for the estimation of the latent heat fluxes. Unfortunately, the nearest meteorological instruments (the Paris-Montsouris station) are located too close to the ground to allow a proper comparison of the wind observations with the results from the VAD method. There was a significant difference in wind direction between the two for several cases, that would be interesting to investigate. However, the closest meteorological tower offering observations more in altitude (including turbulent fluxes) is located on the Saclay plateau, ~20 km southwest from the Jussieu site (SIRTA site, https://sirta.ipsl.fr/) in a suburban area.

Additional cloud images, taken at different time of the day, would also have been exceptionally useful. Indeed, the Terra and Aqua satellites that were used in this study follow sun-synchronous orbits, so they pass over Paris region around the same time every day (~11 UTC for Terra and ~13 UTC for Aqua). Having satellite images throughout the day would not only have allowed to enlarge the training ensemble, by selecting more rolls cases based on possible cloud streets formations, but it would also have served for the verification of the cases classified as rolls. However, time lacked during this project to obtain and exploit images from a geostationary satellite (for instance from the SEVIRI instrument aboard MetOp).

Also, the scanning sequence in the VEGILOT campaign alternated between the different scanning techniques, among which only a part was finally used for this study. An optimized measurement cycle could therefore be implemented in future campaigns to achieve a higher repetition frequency of the quasi-horizontal scans. Moreover, these types of measurements are sufficient for the estimation of the mean wind speed and the mlf-cs, but not for turbulent parameters such as momentum flux, turbulence intensity or TKE. For the estimation of turbulent parameters, it is vital to record continuous observations using one single scanning technique as, by definition, the average in time of the radial wind speed is mandatory. In a future campaign this should be taken into account. A study combining the structures with turbulent parameters would be exceptionally interesting for the scientific community as it could reveal subcategories of the structures.

It would be also interesting to use LES models to simulate the days when multiple cases of rolls were found in order to examine whether the simulations reveal similar patterns and if the model predict correctly the time when the thermals/rolls transitions occur. These results would be compelling, as the studies so far have shown a connection between rolls structures and pollutants' concentrations. Another interesting aspect to examine is whether the models can simulate the formation and dissipation of coherent structures during nllj events and the effect on pollutants' concentrations in this case.

Finally, it is important to note that the range and the resolution of the scans play an important role in the sizes and shapes of the observable patterns. In addition, the meteorological conditions and the topography have an impact in the patterns as well. Even minor differences in the sizes and shapes of the patterns can make the texture analysis parameters used for the classification of this study irrelevant. In that regard, for the replication of this method for a new dataset, a new training ensemble should be built in order to retrieve the appropriate texture analysis parameters for the classification of the ensemble.

General Conclusions

The current manuscript presented three years of work in the analysis of the two-month of mlf-cs observations recorded by a single Doppler wind lidar in Paris. The most astonishing outcome of this study is the possibility to identify and classify the mlf-cs based on the patterns of their fields, retrieved from the radial wind observations by implementing the VAD method. In this manner, an innovative classification methodology was developed by combining texture analysis and the supervised machine learning technique. The identified mlf-cs for this study were: streaks represented by narrow elongated patterns, rolls represented by wide elongated pattern, thermals represented by large enclosed areas and “others” represented by any other pattern that cannot be classified in one of the other three previous categories. The differences between these patterns were evident when observing the texture analysis parameters as a function of the cell pair orientation. The so called curve parameters (amplitude, integral and symmetry of the curves) differentiated the pattern types in a noticeable degree.

A training ensemble of 150 cases (30 streaks, 30 rolls, 30 thermals, 60 “others”) was built based on the visual examination of the mlf-cs patterns and ancillary information (satellite images, weather parameters). When comparing the texture analysis curves parameters, it was apparent that cases belonging to the same type formed clusters of points. Consequently, the texture analysis curves parameters were provided to the QDA algorithm with the stepwise forward selection of the process. The 10-fold cross-validation on the algorithm revealed a classification error of approximately 9% with five texture analysis curves parameters as classifiers. For the individual type of structures, the classification errors were 3.3 % for streaks and “others”, 10 % for rolls and 20 % for thermals. The misclassified thermals cases were classified as rolls and “others” types by the algorithm. The inclusion of three curve amplitude parameters in the classifiers shows the necessity to rotate all the mlf-cs in the same wind direction so that the peak in the texture analysis curve occurs for the same angle. Furthermore, the classifiers cover various neighbour distances from short (2^{nd} neighbour – 100 m) to medium (8^{th} neighbour – 400m) and to large (18^{th} neighbour – 900 m) which is fundamental for the distinction between structures with different sizes such as streaks and rolls. Along with the texture analysis curves parameters, three physical parameters (time of day, mean wind speed, RMSE from the VAD cosine fit) were also provided to the algorithm but they were not selected as relevant classifiers.

The five classifiers which gave the classification error of 9 % were utilized for the classification of the whole 4577 cases ensemble. The algorithm classified 46 % of the cases as “others”, 25 % as streaks, 9 % as rolls (coherent structures sum up to 34 %) and

20 % as thermals. Despite the classifiers being only texture analysis parameters, the diurnal variations of the structures during the two-month period agreed with the theoretical knowledge. In particular, the vast majority of rolls (87%) and thermals (77 %) cases were classified during the day between 08:00 UTC and 18:00 UTC as expected because these structures develop after strong instabilities. Oppositely, most of the cases of streaks (66 %) and “others” (79 %) were classified during the night, from 18:00 to 08:00 UTC. This was expected for streaks as nllj events occurred during 20 out of the 63 nights of the campaign, so that favorable conditions for the formation of streaks structures occurred frequently. Regarding the “others” cases, this type is mainly composed of asymmetric radial wind field for which the VAD method does not work, and this type of field is mainly associated with calm and low wind conditions that occur primarily during the night, which was confirmed by analyzing the mean wind conditions. The cases classified as “others” associated with mean wind values lower than 4 $m \cdot s^{-1}$ made up 75 % of the total “others” cases. Cases classified as streaks and rolls, on the other hand, were mostly associated with moderate mean wind values between 4 and 10 $m \cdot s^{-1}$, as it is the case in to other studies.

The automatic classification methodology made the analysis of random cases within the two-month dataset fast, easy and smooth. More particularly, the 18 days during which it was possible to estimate the ABL height for the majority of the 24 hours were selected in order to examine the relationship between ABL height and the structures types. The 94 % of the cases classified as rolls and 80% of the cases classified as thermals occurred in a developing or a well-developed ABL. On the contrary, cases classified as streaks and “others” were associated with SBL for 90 % and 86 % respectively.

A 4-day case study with daily nllj events showed that streaks structures appeared concurrently with the occurrence of the nllj. During the morning, as the nllj disappeared and the thermal instabilities prevailed, the streaks were replaced by thermals or rolls structures. The difference in horizontal sizes were confirmed by the estimation of the periodicity sizes in the direction transverse to the mean wind, via the spectrum graphs of the radial wind speed. The periodicity sizes (including one positive and one negative stripe in the mlf-cs) ranged between 400 and 800 m for the cases classified as streaks cases and between 1.3 and 2 km for the cases classified as rolls.

The automated classification of the mlf-cs provides a consistent way to identify and classify the structures. It can be implemented in any other dataset independently of the date, duration and place as long as the patterns of the mlf-cs fields are similar. The results indicated a good agreement between the classified structures and the physical parameters. There is still room for improvement by optimizing the selection of the texture

analysis parameters and/or the supervised machine learning algorithm. Nevertheless, it creates the possibilities for various new exciting studies.

So far the wind lidar observations have not been used for a statistical analysis of the physical properties of the structures due to the limitations of visually identifying structures in large datasets. The automated classification can assist such studies. Furthermore, it can be combined with LES models in order to examine the impact of the structures on the pollutants' concentrations. Finally, as the lidar is capable to estimate turbulent parameters, with the appropriate scanning configuration, it could be possible to alternate between horizontal scans, thus mlf-cs fields, and turbulent parameters. In this way, the possibility to observe subcategories of the structures could be examined.

References

Achtert, P., Brooks, I.M., Brooks, B.J., Moat, B.I., Prytherch, J., Persson, P.O.G., Tjernström, M., 2015. Measurement of wind profiles by motion-stabilised ship-borne Doppler lidar. Atmos. Meas. Tech. 8, 4993–5007. https://doi.org/10.5194/amt-8-4993-2015

Adrian, R.J., 2007. Hairpin vortex organization in wall turbulence. Phys. Fluids 19, 41301. https://doi.org/10.1063/1.2717527

Alparone, L., Benelli, G., Vagniluca, A., 1990. Texture-based analysis techniques for the classification of radar images. IEE proceedings. Part F. Commun. radar signal Process. 137, 276–282. https://doi.org/10.1049/ip-f-2.1990.0041

Ancellet, G., Ravetta, F., 1998. Compact airborne lidar for tropospheric ozone: description and field measurements. Appl. Opt. 37, 5509-5521. https://doi.org/10.1364/ao.37.005509

Anurose, T.J., Subrahamanyam, D.B., Sunilkumar, S. V., 2018. Two years observations on the diurnal evolution of coastal atmospheric boundary layer features over Thiruvananthapuram (8.5◦ N, 76.9◦ E), India. Theor. Appl. Climatol. 131, 77–90. https://doi.org/10.1007/s00704-016-1955-y

Ao, C.O., Waliser, D.E., Chan, S.K., Li, J.-L., Tian, B., Xie, F., Mannucci, A.J., 2012. Planetary boundary layer heights from GPS radio occultation refractivity and humidity profiles. J. Geophys. Res. Atmos. 117. https://doi.org/10.1029/2012JD017598

Aouizerats, B., Tulet, P., Pigeon, G., Masson, V., Gomes, L., 2011. Atmospheric Chemistry and Physics High resolution modelling of aerosol dispersion regimes during the CAPITOUL field experiment: from regional to local scale interactions. Atmos. Chem. Phys 11, 7547–7560. https://doi.org/10.5194/acp-11-7547-2011

Arya, P.S., 2001. Introduction to Micrometeorology. Elsevier, Amsterdam, the Netherlands.

Azimi-Sadjadi, M.R., Zekavat, S.A., n.d. Cloud classification using support vector machines, in: IGARSS 2000. IEEE 2000 International Geoscience and Remote Sensing Symposium. Taking the Pulse of the Planet: The Role of Remote Sensing in Managing the Environment. Proceedings (Cat. No.00CH37120). IEEE, pp. 669–671. https://doi.org/10.1109/IGARSS.2000.861666

Balaji, V., Clark, T.L., 1988. Scale selection in locally forced convective fields and the initiation of deep cumulus. J. Atmos. Sci. 45, 3188–3211. https://doi.org/10.1175/1520-0469(1988)045<3188:SSILFC>2.0.CO;2

Bankert, R.L., 1994. Cloud classification of AVHRR imagery in maritime regions using a probabilistic neural network. J. Appl. Meteorol. 33, 909–918. https://doi.org/10.1175/1520-0450(1994)033<0909:CCOAII>2.0.CO;2

Banta, R.M., Newsom, R.K., Lundquist, J.K., Pichugina, Y.L., Coulter, R.L., Mahrt, L., 2002. Nocturnal low-level jet characteristics over Kansas during cases-99. Boundary-Layer Meteorol. 105, 221–252. https://doi.org/10.1023/A:1019992330866

Banta, R.M., Pichugina, Y.L., Brewer, W.A., 2006. Turbulent velocity-variance profiles in

the stable boundary layer generated by a nocturnal low-level jet. J. Atmos. Sci. 63, 2700–2719. https://doi.org/10.1175/JAS3776.1

Bartello, P., Métais, O., Lesieur, M., 1994. Coherent structures in rotating three-dimensional turbulence. J. Fluid Mech. 273, 1–29. https://doi.org/10.1017/S0022112094001837

Barthlott, C., Drobinski, P., Fesquet, C., Dubos, T., Pietras, C., 2007. Long-term study of coherent structures in the atmospheric surface layer. Boundary-Layer Meteorol. 125, 1–24. https://doi.org/10.1007/s10546-007-9190-9

Bernard-Trottolo, S., Campistron, B., Druilhet, A., Lohou, F., Saïd, F., 2004. TRAC98: Detection of coherent structures in a convective boundary layer using airborne measurements. Boundary-Layer Meteorol. 111, 181–224. https://doi.org/10.1023/B:BOUN.0000016465.50697.63

Bianco, L., Djalalova, I. V., King, C.W., Wilczak, J.M., 2011. Diurnal evolution and annual variability of boundary-layer height and its correlation to other meteorological variables in California's Central Valley. Boundary-Layer Meteorol. 140, 491–511. https://doi.org/10.1007/s10546-011-9622-4

Blackadar, A.K., 1957. Boundary Layer Wind Maxima and Their Significance for the Growth of Nocturnal Inversions.

Bolin, B., Persson, C., 1975. Regional dispersion and deposition of atmospheric pollutants with particular application to sulfur pollution over Western Europe. Tellus 27, 281–310. https://doi.org/10.1111/j.2153-3490.1975.tb01679.x

Bonamente, M., 2017. Statistics and analysis of scientific data. Grad. Texts Phys. 55–83. https://doi.org/10.1007/978-1-4939-6572-4

Bonin, T.A., Choukulkar, A., Brewer, W.A., Sandberg, S.P., Weickmann, A.M., Pichugina, Y.L., Banta, R.M., Oncley, S.P., Wolfe, D.E., 2017. Evaluation of turbulence measurement techniques from a single Doppler lidar. Atmos. Meas. Tech. 10, 3021–3039. https://doi.org/10.5194/amt-10-3021-2017

Browning, K.A., Wexler, R., 1968. The determination of kinematic properties of a wind field using Doppler radar. J. Appl. Meteorol. 7, 105–113. https://doi.org/10.1175/1520-0450(1968)007<0105:tdokpo>2.0.co;2

Brümmer, B., 1999. Roll and cell convection in wintertime Arctic cold-air outbreaks. J. Atmos. Sci. 56, 2613–2636. https://doi.org/10.1175/1520-0469(1999)056<2613:RACCIW>2.0.CO;2

Brümmer, B., Bakan, S., Hinzpeter, H., 1985. Kontur: Observations of cloud streets and open cellular structures. Dyn. Atmos. Ocean. 9, 281–296. https://doi.org/10.1016/0377-0265(85)90024-7

Caicedo, V., Rappenglück, B., Lefer, B., Morris, G., Toledo, D., Delgado, R., 2017. Comparison of aerosol lidar retrieval methods for boundary layer height detection using ceilometer aerosol backscatter data. Atmos. Meas. Tech. 10, 1609–1622. https://doi.org/10.5194/amt-10-1609-2017

Calaf, M., Hultmark, M., Oldroyd, H.J., Simeonov, V., Parlange, M.B., 2013. Coherent structures and the k-1 spectral behaviour. Phys. Fluids 25, 125107. https://doi.org/10.1063/1.4834436

Calhoun, R., Heap, R., Princevac, M., Newsom, R., Fernando, H., Ligon, D., 2006. Virtual towers using coherent doppler lidar during the Joint Urban 2003 dispersion experiment. J. Appl. Meteorol. Climatol. 45, 1116–1126. https://doi.org/10.1175/JAM2391.1

Cariou, J.P., Parmentier, R., Valla, M., Sauvage, L., Antoniou, I., Courtney, M., 2007. An innovative and autonomous 1.5 µm Coherent lidar for PBL wind profiling, in: Proceedings of 14th Coherent Laser Radar Conference.

Castellano, G., Bonilha, L., Li, L.M., Cendes, F., 2004. Texture analysis of medical images. Clin. Radiol. 59, 1061-1069. https://doi.org/10.1016/j.crad.2004.07.008

Chai, T., Lin, C.-L., Newsom, R.K., 2004. Retrieval of microscale flow structures from high-resolution Doppler lidar data using an adjoint model, J. Atm. Sci. 6, 1500-1520. https://doi.org/10.1175/1520-0469(2004)061<1500:ROMFSF>2.0.CO;2

Chandra, S., Kumar, M., Dwivedi, A.K., 2014. Characterization of the atmospheric boundary layer from radiosonde observations along eastern end of monsoon trough of india. J. Earth Syst. Sci. 123, 1233–1240. https://doi.org/10.1007/s12040-014-0458-4

Cheliotis, I., Dieudonné, E., Delbarre, H., Sokolov, A., Dmitriev, E., Augustin, P., Fourmentin, M., 2020. Detecting turbulent structures on single Doppler lidar large datasets: an automated classification method for horizontal scans. Atmos. Meas. Tech. 13, 6579–6592. https://doi.org/10.5194/amt-13-6579-2020

Chen, W., Kuze, H., Uchiyama, A., Suzuki, Y., Takeuchi, N., 2001. One-year observation of urban mixed layer characteristics at Tsukuba, Japan using a micro pulse lidar. Atmos. Environ. 35, 4273–4280. https://doi.org/10.1016/S1352-2310(01)00181-9

Cimini, D., De Angelis, F., Dupont, J.C., Pal, S., Haeffelin, M., 2013. Mixing layer height retrievals by multichannel microwave radiometer observations. Atmos. Meas. Tech. 6, 2941–2951. https://doi.org/10.5194/amt-6-2941-2013

Cimini, D., Hewison, T.J., Martin, L., Güldner, J., Gaffard, C., Marzano, F.S., 2006. Temperature and humidity profile retrievals from ground-based microwave radiometers during TUC. Meteorol. Zeitschrift 15, 45–56. https://doi.org/10.1127/0941-2948/2006/0099

Clausi, D.A., Zhao, Y., 2002. Rapid extraction of image texture by co-occurrence using a hybrid data structure. Comput. Geosci. 28, 763–774. https://doi.org/10.1016/S0098-3004(01)00108-X

Coen, M.C., Collaud Coen, M., Praz, C., Haefele, A., Ruffieux, D., Kaufmann, P., Calpini, B., 2014. Absorbing aerosols and fate of the indian glaciers-AAFIG View project A long-term black-carbon observation in Central Tibet View project Determination and climatology of the planetary boundary layer height above the Swiss plateau by in situ and remote sensing measurements as well as by the COSMO-2 model. Atmos. Chem. Phys 14, 13205–13221. https://doi.org/10.5194/acp-14-13205-2014

Cover, T.M., Hart, P.E., 1967. Nearest neighbour pattern classification. IEEE Trans. Inf. Theory 13, 21–27. https://doi.org/10.1109/TIT.1967.1053964

Davies, F., Collier, C.G., Bozier, K.E., 2005. Errors associated with dual-Doppler-lidar turbulence measurements, in: Journal of Optics A: Pure and Applied Optics. IOP

Publishing, 7, S280. https://doi.org/10.1088/1464-4258/7/6/005

Davis, E. V., Rajeev, K., Mishra, M.K., 2020. Effect of clouds on the diurnal evolution of the atmospheric boundary-layer height over a tropical coastal station. Boundary-Layer Meteorol. 175, 135–152. https://doi.org/10.1007/s10546-019-00497-6

De Tomasi, F., Marcello Miglietta, M., Rita Perrone, M., 2011. The growth of the planetary boundary layer at a coastal site: a case Ssudy 139, 521–541. https://doi.org/10.1007/s10546-011-9592-6

Deardorff, J.W., 1972. Numerical investigation of neutral and unstable planetary boundary layers. J. Atmos. Sci. 29, 91–115. https://doi.org/10.1175/1520-0469(1972)029<0091:nionau>2.0.co;2

Deniz, E., Şengür, A., Kadiroğlu, Z., Guo, Y., Bajaj, V., Budak, Ü., 2018. Transfer learning based histopathologic image classification for breast cancer detection. Heal. Inf. Sci. Syst. 6, 1–7. https://doi.org/10.1007/s13755-018-0057-x

Dieudonné, E., 2012. Analyse multi-instrumentale de l'influence de la variabilité de la hauteur de couche limite sur la distribution verticale des oxydes d'azote en région parisienne. http://www.theses.fr.

Doviak, R.J., Zrnic, D.S., 1993. Doppler Radar and Weather Observations, Doppler Radar and Weather Observations. Elsevier, Amsterdam, The Netherlands. https://doi.org/10.1016/c2009-0-22358-0

Drobinski, P., Brown, R.A., Flamant, P.H., Pelon, J., 1998. Evidence of organized large eddies by ground-based Doppler lidar, sonic anemometer and sodar. Boundary-Layer Meteorol. 88, 343–361. https://doi.org/10.1023/A:1001167212584

Drobinski, P., Carlotti, P., Newsom, R.K., Banta, R.M., Foster, R.C., Redelsperger, J.-L., 2004. The structure of the near-neutral atmospheric surface layer. J. Atmos. Sci. 61, 699–714. https://doi.org/10.1175/1520-0469(2004)061<0699:TSOTNA>2.0.CO;2

Drobinski, P., Foster, R.C., 2003. On the origin of near-surface streaks in the neutrally-stratified planetary boundary layer. Boundary-Layer Meteorol. 108, 247–256. https://doi.org/10.1023/A:1024100125735

Eberhard, W.L., Cupp, R.E., Healy, K.R., 1989. Doppler lidar measurement of profiles of turbulence and momentum flux. J. Atmos. Ocean. Technol. 6, 809–819. https://doi.org/10.1175/1520-0426(1989)006<0809:dlmopo>2.0.co;2

Etling, D., Brown, R.A., 1993. Roll vortices in the planetary boundary layer: A review. Boundary-Layer Meteorol. 65, 215-248. https://doi.org/10.1007/BF00705527

Eymard, L., Weill, A., 1988. Dual Doppler radar investigation of the tropical convective boundary layer. J. Atmos. Sci. 45, 853–864. https://doi.org/10.1175/1520-0469(1988)045<0853:DDRIOT>2.0.CO;2

Farrell, B.F., Ioannou, P.J., 2012. Dynamics of streamwise rolls and streaks in turbulent wall-bounded shear flow. J. Fluid Mech. 708, 149–196. https://doi.org/10.1017/jfm.2012.300

Fernando, H.J.S., Weil, J.C., 2010. Whither the stable boundary layer? Bull. Am. Meteorol. Soc. 91, 1475–1484. https://doi.org/10.1175/2010BAMS2770.1

Ferrare, R.A., Schols, J.L., Eloranta, E.W., Coulter, R., 1991. Lidar observations of banded convection during BLX83. J. Appl. Meteorol. 30, 312–326. https://doi.org/10.1175/1520-0450(1991)030<0312:LOOBCD>2.0.CO;2

Flamant, C., Pelon, J., Flamant, P.H., Durand, P., 1997. Lidar determination of the entrainment zone thickness at the top of the unstable marine atmospheric boundary layer. Boundary-Layer Meteorol. 83, 247–284. https://doi.org/10.1023/A:1000258318944

Fochesatto, G.J., Drobinski, P., Flamant, C., Guedalia, D., Sarrat, C., Flamant, P.H., Pelon, J., 2001. Evidence of dynamical coupling between the residual layer and the developing convective boundary layer. Boundary-Layer Meteorol. 99, 451–464. https://doi.org/10.1023/A:1018935129006

Fritts, D.C., Alexander, M.J., 2003. Gravity wave dynamics and effects in the middle atmosphere. Rev. Geophys. 41. https://doi.org/10.1029/2001RG000106

Gal-Chen, T., Xu, M., Eberhard, W.L., 1992. Estimations of atmospheric boundary layer fluxes and other turbulence parameters from Doppler lidar data. J. Geophys. Res. 97, 18409–18423. https://doi.org/10.1029/91jd03174

Garratt, J.R., 1994. Review: the atmospheric boundary layer. Earth Sci. Rev. 37, 89–134. https://doi.org/10.1016/0012-8252(94)90026-4

Gisch, D.L., Bodmann, B.E.J., De Vilhena, M.T.M.B., 2018. Analytical model study for the pollutants dispersion with coherent structures presence. Proceeding Ser. Brazilian Soc. Comput. Appl. Math. 6.

Gu, Z.Q., Duncan, C.N., Renshaw, E., Mugglestone, M.A., Cowan, C.F.N., Grant, P.M., 1989. Comparison of techniques for measuring cloud texture in remotely sensed satellite meteorological image data. IEE proceedings. Part F. Commun. radar signal Process. 136, 236–248. https://doi.org/10.1049/ip-f-2.1989.0037

Guo, P., Kuo, Y.H., Sokolovskiy, S. V., Lenschow, D.H., 2011. Estimating atmospheric boundary layer depth depth COSMIC radio occultation data. J. Atmos. Sci. 68, 1703–1713. https://doi.org/10.1175/2011JAS3612.1

Haralick, R.M., 1979. Statistical and structural approaches to texture. Proc. IEEE 67, 786–804. https://doi.org/10.1109/PROC.1979.11328

Haralick, R.M., Dinstein, I., Shanmugam, K., 1973. Textural Features for Image Classification. IEEE Trans. Syst. Man Cybern. SMC-3, 610–621. https://doi.org/10.1109/TSMC.1973.4309314

Harnby, N., Edwards, M.F., Nienow, A.W., 1985. Mixing in the process industries. Butterworths & Co. (Publishers) Ltd, Oxford, United Kingdom. https://doi.org/10.1016/b978-0-7506-3760-2.x5020-3

Hastie, T., Tibshirani, R., Friedman, J., 2009. The elements of statistical learning: Data mining, inference, and prediction, Springer Series in Statistics.

Heinsohn, R.J., Kabel, R.L., 1998. Sources and control of air pollution: Engineering principles.

Holli, K., Lääperi, A.L., Harrison, L., Luukkaala, T., Toivonen, T., Ryymin, P., Dastidar, P., Soimakallio, S., Eskola, H., 2010. Characterization of breast cancer types by texture

analysis of magnetic resonance images. Acad. Radiol. 17, 135–141. https://doi.org/10.1016/j.acra.2009.08.012

Huld, T., Müller, R., Gambardella, A., 2012. A new solar radiation database for estimating PV performance in Europe and Africa. Sol. Energy 86, 1803–1815. https://doi.org/10.1016/j.solener.2012.03.006

Hussain, A.K.M.F., 1983. Coherent structures - Reality and myth. Phys. Fluids 26, 2816–2850. https://doi.org/10.1063/1.864048

Hutchins, N., Marusic, I., 2007. Evidence of very long meandering features in the logarithmic region of turbulent boundary layers. J. Fluid Mech. 579, 1–28. https://doi.org/10.1017/S0022112006003946

Iwai, H., Ishii, S., Tsunematsu, N., Mizutani, K., Murayama, Y., Itabe, T., Yamada, I., Matayoshi, N., Matsushima, D., Weiming, S., Yamazaki, T., Iwasaki, T., 2008b. Dual-Doppler lidar observation of horizontal convective rolls and near-surface streaks. Geophys. Res. Lett. 35, L14808. https://doi.org/10.1029/2008GL034571

James, G., Witten, D., Hastie, T., Tibshirani, R., 2000. An introduction to Statistical Learning, Current medicinal chemistry. https://doi.org/10.1007/978-1-4614-7138-7

Jiang, P., Wen, Z., Sha, W., Chen, G., 2017. Interaction between turbulent flow and sea breeze front over urban-like coast in large-eddy simulation. J. Geophys. Res. Atmos. 122, 5298–5315. https://doi.org/10.1002/2016JD026247

Johnson, A.H., Siccama, T.G., 1983. Acid deposition and forest decline. Environ. Sci. Technol. 17, 294-305. https://doi.org/10.1021/es00113a001

Kayitakire, F., Hamel, C., Defourny, P., 2006. Retrieving forest structure variables based on image texture analysis and IKONOS-2 imagery. Remote Sens. Environ. 102, 390–401. https://doi.org/10.1016/j.rse.2006.02.022

Kelly, R.D., 1982. A single Doppler radar study of horizontal-roll convection in a lake-effect snow storm (Lake-Michigan). J. Atmos. Sci. 39, 1521–1531. https://doi.org/10.1175/1520-0469(1982)039<1521:ASDRSO>2.0.CO;2

Khanna, S., Brasseur, J.G., 1998. Three-dimensional buoyancy and shear-induced local structure of the atmospheric boundary layer. J. Atmos. Sci. 55, 710–743. https://doi.org/10.1175/1520-0469(1998)055<0710:TDBASI>2.0.CO;2

Kim, S.W., Yoon, S.C., Won, J.G., Choi, S.C., 2007. Ground-based remote sensing measurements of aerosol and ozone in an urban area: A case study of mixing height evolution and its effect on ground-level ozone concentrations. Atmos. Environ. 41, 7069–7081. https://doi.org/10.1016/j.atmosenv.2007.04.063

Klein, A., Ravetta, F., Thomas, J.L., Ancellet, G., Augustin, P., Wilson, R., Dieudonné, E., Fourmentin, M., Delbarre, H., Pelon, J., 2019. 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. https://doi.org/10.1016/j.atmosenv.2018.10.009

Klein, P., Bonin, T.A., Newman, J.F., Turner, D.D., Chilson, P.B., Wainwright, C.E., Blumberg, W.G., Mishra, S., Carney, M., Jacacobsen, E.P., Wharton, S., Newsom, R.K., 2015. LABLE: A multi-institutional, student-led, atmospheric boundary layer experiment. Bull. Am. Meteorol. Soc. 96, 1743–1764. https://doi.org/10.1175/BAMS-

D-13-00267.1

Kline, S.J., Reynolds, W.C., Schraub, F.A., Runstadler, P.W., 1967. The structure of turbulent boundary layers. J. Fluid Mech. 30, 741–773. https://doi.org/10.1017/S0022112067001740

Konrad, T.G., 1968. The alignment of clear air convective cells, in: Proceedings International Conference of Cloud Physics. Toronto, Canada, pp. 539–543.

Kotthaus, S., Grimmond, C.S.B., 2018. Atmospheric boundary-layer characteristics from ceilometer measurements. Part 1: A new method to track mixed layer height and classify clouds. Q. J. R. Meteorol. Soc. 144, 1525–1538. https://doi.org/10.1002/qj.3299

Kropfli, R.A., Kohn, N.M., 1978. Persistent horizontal rolls in the urban mixed layer as revealed by dual-Doppler radar. J. Appl. Meteorol. 17, 669–676. https://doi.org/10.1175/1520-0450(1978)017<0669:phritu>2.0.co;2

Kubat, M., 2017. An introduction to machine learning. Springer International Publishing. https://doi.org/10.1007/978-3-319-63913-0

Kuettner, J., 1959. The band structure of the atmosphere. Tellus 11, 267–294. https://doi.org/10.3402/tellusa.v11i3.9319

Kuettner, J.P., 1971. Cloud bands in the earth's atmosphere: Observations and Theory. Tellus 23, 404–426. https://doi.org/10.3402/tellusa.v23i4-5.10519

Kunkel, K.E., Eloranta, E.W., Weinman, J.A., 1980. Remote determination of winds, turbulence spectra and energy dissipation rates in the boundary layer from lidar measurements. J. Atmos. Sci. 37, 978–985. https://doi.org/10.1175/1520-0469(1980)037<0978:rdowts>2.0.co;2

Kuznetsova, A., Rom, H., Alldrin, N., Uijlings, J., Krasin, I., Pont-Tuset, J., Kamali, S., Popov, S., Malloci, M., Kolesnikov, A., Duerig, T., Ferrari, V., 2018. The Open Images Dataset V4: Unified image classification, object detection, and visual relationship detection at scale. Int. J. Comput. Vis. 128, 1956–1981. https://doi.org/10.1007/s11263-020-01316-z

Larsen, S.E., 2013. The atmospheric boundary layer over land and sea: Focus on the off-shore Southern Baltic and Southern North Sea region. DTU Wind Energy. UR:L http://www.southbaltic-offshore.eu/reports-studies.html (accessed 18.12.20).

LeMone, M., 1972. The structure and dynamics of the horizontal roll vortices in the planetary boundary layer. J. Atmos. Sci. 30, 1077–1091. https://doi.org/10.1175/1520-0469(1973)030<1077:tsadoh>2.0.co;2

LeMone, M.A., 1973. The Structure and Dynamics of Horizontal Roll Vortices in the Planetary Boundary Layer. J. Atmos. Sci. 30, 1077–1091. https://doi.org/10.1175/1520-0469(1973)030<1077:TSADOH>2.0.CO;2

Lemonsu, A., Masson, V., 2002. Simulation of a summer urban breeze over Paris. Boundary-Layer Meteorol. 104, 463–490. https://doi.org/10.1023/A:1016509614936

Lenschow, D.H., Agee, E.M., 1976. Preliminary results from the Air Mass Transformation Experiment (AMTEX). Bull. Am. Meteorol. Soc. 57, 1346–1355. https://doi.org/10.1175/1520-0477-57.11.1346

Lhermitte, R.M., 1962. Note on wind variability with Doppler radar. J. Atmos. Sci. 19, 343–346.

Lin, C.L., Xia, Q., Calhoun, R., 2008a. Retrieval of urban boundary layer structures from Doppler lidar data. Part II: Proper orthogonal decomposition. J. Atmos. Sci. 65, 21–42. https://doi.org/10.1175/2007JAS2329.1

Lin, C.L., Xia, Q., Calhoun, R., 2008b. Retrieval of urban boundary layer structures from Doppler lidar data. Part II: Proper orthogonal decomposition. J. Atmos. Sci. 65, 21–42. https://doi.org/10.1175/2007JAS2329.1

Lohou, F., Druilhet, A., Campistron, B., 1998. Spatial and temporal characteristics of horizontal rolls and cells in the atmospheric boundary layer based on radar and in situ observations. Boundary-Layer Meteorol. 89, 407–444. https://doi.org/10.1023/A:1001791408470

Lokoshchenko, M.A., 2002. Long-term sodar observations in Moscow and a new approach to potential mixing determination by radiosonde data. J. Atmos. Ocean. Technol. 19, 1151–1162. https://doi.org/10.1175/1520-0426(2002)019<1151:LTSOIM>2.0.CO;2

Lothon, M., Lenschow, D.H., Mayor, S.D., 2006. Coherence and scale of vertical velocity in the convective boundary layer from a Doppler lidar. Boundary-Layer Meteorol. 121, 521–536. https://doi.org/10.1007/s10546-006-9077-1

Lund, B., Graber, H.C., Romeiser, R., 2012. Wind retrieval from shipborne nautical X-band radar data. IEEE Trans. Geosci. Remote Sens. 50, 3800–3811. https://doi.org/10.1109/TGRS.2012.2186457

MacDonald, M., Kurowski, M.J., Teixeira, J., 2020. Direct numerical simulation of the moist stably stratified surface layer: turbulence and fog formation. Boundary-Layer Meteorol. 175, 343–368. https://doi.org/10.1007/s10546-020-00511-2

Mann, J., Cariou, J.-P., Courtney, M.S., Parmentier, R., Mikkelsen, T., Wagner, R., Lindelöw, P., Sjöholm, M., Enevoldsen, K., 2008. Comparison of 3D turbulence measurements using three staring wind lidars and a sonic anemometer. IOP Conf. Ser. Earth Environ. Sci. 1, 012012. https://doi.org/10.1088/1755-1315/1/1/012012

Martin, T., Bakan, S., 1991. Airplane investigation of a case of convective cloud bands over the North Sea. Boundary-Layer Meteorol. 56, 359–380. https://doi.org/10.1007/BF00119212

Mehta, S.K., Ratnam, M.V., Sunilkumar, S. V, Rao, D.N., Murthy, B.V.K., 2017. Diurnal variability of the atmospheric boundary layer height over a tropical station in the Indian monsoon region. Atmos. Chem. Phys 17, 531–549. https://doi.org/10.5194/acp-17-531-2017

Menut, L., Flamant, C., Pelon, J., Flamant, P.H., 1999. Urban boundary-layer height determination from lidar measurements over the Paris area. Appl. Opt. 38, 945-954. https://doi.org/10.1364/ao.38.000945

Miller, S.T.K., 2003. Sea breeze: Structure, forecasting, and impacts. Rev. Geophys. 41, 1011. https://doi.org/10.1029/2003RG000124

Moeng, C.H., Sullivan, P.P., 1994. A comparison of shear and buoyancy-driven planetary boundary layer flows. J. Atmos. Sci. 51, 999–1022. https://doi.org/10.1175/1520-0469(1994)051<0999:acosab>2.0.co;2

Nakanishi, M., 2000. Large-eddy simulation of radiation fog. Boundary-Layer Meteorol. 94, 461–493. https://doi.org/10.1023/A:1002490423389

Newman, J.F., Klein, P.M., Wharton, S., Sathe, A., Bonin, T.A., Chilson, P.B., Muschinski, A., 2016. Evaluation of three lidar scanning strategies for turbulence measurements. Atmos. Meas. Tech. 9, 1993–2013. https://doi.org/10.5194/amt-9-1993-2016

Newsom, R., Calhoun, R., Ligon, D., Allwine, J., 2008. Linearly organized turbulence structures observed over a suburban area by Dual-Doppler lidar. Boundary-Layer Meteorol. 127, 111–130. https://doi.org/10.1007/s10546-007-9243-0

Newsom, R.K., Ligon, D., Calhoun, R., Heap, R., Cregan, E., Princevac, M., 2005. Retrieval of microscale wind and temperature fields from single- and dual-Doppler lidar data. J. Appl. Meteorol. 44, 1324–1344. https://doi.org/10.1175/JAM2280.1

Oke, T.R., 1988. Boundary Layer Climates. Halsted Press, New York.

Olaode, A., Naghdy, G., Todd, C., 2014. Unsupervised classification of images: A review, International Journal of Image Processing (IJIP), 8, 325-342.

Pal, S., Behrendt, A., Wulfmeyer, V., 2010. Elastic-backscatter-lidar-based characterization of the convective boundary layer and investigation of related statistics. Ann. Geophys. 28, 825–847. https://doi.org/10.5194/angeo-28-825-2010

Pal, S., Haeffelin, M., 2015. Forcing mechanisms governing diurnal, seasonal, and interannual variability in the boundary layer depths: Five years of continuous lidar observations over a suburban site near Paris. J. Geophys. Res. Atmos. 120, 11,936-11,956. https://doi.org/10.1002/2015JD023268

Partio, M., Cramariuc, B., Gabbouj, M., Visa, A., 2002. Rock texture retrieval using gray level co-occurrence matrix, in: Proceedings of 5th Nordic signal processing Symposium. Hurtigruten, Norway.

Pelon, J., Mallet, M., Mariscal, A., Goloub, P., Tanré, D., Bou Karam, D., Flamant, C., Haywood, J., Pospichal, B., Victori, S., 2008. Microlidar observations of biomass burning aerosol over Djougou (Benin) during African Monsoon Multidisciplinary Analysis Special Observation Period 0: Dust and Biomass-Burning Experiment. J. Geophys. Res. 113. https://doi.org/10.1029/2008jd009976

Philippopoulos, K., Deligiorgi, D., Mavrakou, T., Cheliotis, J., 2014. Winter atmospheric circulation patterns and their relationship with the meteorological conditions in Greece. Meteorol. Atmos. Phys. 124, 195–204. https://doi.org/10.1007/s00703-014-0310-z

Pichugina, Y.L., Banta, R.M., Kelley, N.D., Jonkman, B.J., Tucker, S.C., Newsom, R.K., Brewer, W.A., 2008. Horizontal-velocity and variance measurements in the stable boundary layer using Doppler lidar: Sensitivity to averaging procedures, in: Journal of Atmospheric and Oceanic Technology. American Meteorological Society, 25, 1307–1327. https://doi.org/10.1175/2008JTECHA988.1

Powell, D.C., Elderkin, C.E., 1974. Investigation of the application of Taylor's hypothesis to atmospheric boundary layer turbulence. J. Atmos. Sci. 31, 990–1002. https://doi.org/10.1175/1520-0469(1974)031<0990:AIOTAO>2.0.CO;2

Rabin, R.M., Doviak, R.J., Sundara-Rajan, A., 1982. Doppler radar observations of momentum flux in a cloudless convective layer with rolls. J. Atmos. Sci. 39, 851–863.

https://doi.org/10.1175/1520-0469(1982)039<0851:DROOMF>2.0.CO;2

Reinking, R.F., Doviak, R.J., Gilmer, R.O., 1981. Clear-air roll vortices and turbulent motions as detected with an airborne gust probe and dual-Doppler radar. J. Appl. Meteorol. 20, 678–685. https://doi.org/10.1175/1520-0450(1981)020<0678:CARVAT>2.0.CO;2

Roach, W.T., Brown, R., Caughey, S.J., Garland, J.A., Readings, C.J., 1976. The physics of radiation fog: I - a field study. Q. J. R. Meteorol. Soc. 102, 313–333. https://doi.org/10.1002/qj.49710243204

Roth, M., 2007. Review of atmospheric turbulence over cities. Q. J. R. Meteorol. Soc. 126, 941–990. https://doi.org/10.1002/qj.49712656409

Russell, P.B., Uthe, E.E., Ludwig, F.L., Shaw, N.A., 1974. A comparison of atmospheric structure as observed with monostatic acoustic sounder and lidar techniques. J. Geophys. Res. 79, 5555–5566. https://doi.org/10.1029/jc079i036p05555

Saint-Pierre, C., Becue, V., Diab, Y., Teller, J., 2010. Case study of mixed-use high-rise location at the Greater Paris scale. WIT Trans. Ecol. Environ. 129, 251–262. https://doi.org/10.2495/SC100221

Sandeepan, B.S., Rakesh, P.T., Venkatesan, R., 2013. Observation and simulation of boundary layer coherent roll structures and their effect on pollution dispersion. Atmos. Res. 120–121, 181–191. https://doi.org/10.1016/j.atmosres.2012.08.016

Sathe, A., Mann, J., 2013. A review of turbulence measurements using ground-based wind lidars. Atmos. Meas. Tech. 6, 3147–3167. https://doi.org/10.5194/amt-6-3147-2013

Sathe, A., Mann, J., Vasiljevic, N., Lea, G., 2015. A six-beam method to measure turbulence statistics using ground-based wind lidars. Atmos. Meas. Tech. 8, 729–740. https://doi.org/10.5194/amt-8-729-2015

Schmid, P., Niyogi, D., 2012. A method for estimating planetary boundary layer heights and its application over the ARM southern great plains site. J. Atmos. Ocean. Technol. 29, 316–322. https://doi.org/10.1175/JTECH-D-11-00118.1

Schowengerdt, R.A., 2007. Models and methods for image processing, Remote Sensing. Elsevier Inc., Amsterdam, The Netherlands. https://doi.org/10.1016/B978-0-12-369407-2.X5000-1

Seibert, P., Beyrich, F., Gryning, S.E., Joffre, S., Rasmussen, A., Tercier, P., 2000. Review and intercomparison of operational methods for the determination of the mixing height. Atmos. Environ. 34, 1001-10027. https://doi.org/10.1016/S1352-2310(99)00349-0

Seidel, D.J., Zhang, Y., Beljaars, A., Golaz, J.C., Jacobson, A.R., Medeiros, B., 2012. Climatology of the planetary boundary layer over the continental United States and Europe. J. Geophys. Res. Atmos. 117. https://doi.org/10.1029/2012JD018143

Shonk, J., 2013. Introducing meteorology: A guide to the weather. Dunedin Academic Press Ltd.

Smedman, A.S., 1991. Occurrence of roll circulations in a shallow boundary layer. Boundary-Layer Meteorol. 57, 343–358. https://doi.org/10.1007/BF00120053

Sokolov, A., Dmitriev, E., Gengembre, C., Delbarre, H., 2020. Automated classification of regional meteorological events in a coastal area using in-situ measurements. J. Atmos. Ocean. Technol. 37, 723-739. https://doi.org/10.1175/JTECH-D-19-0120.1

Soldati, A., 2005. Particles turbulence interactions in boundary layers. ZAMM 85, 683–699. https://doi.org/10.1002/zamm.200410213

Solheim, F., Godwin, J.R., Westwater, E.R., Han, Y., Keihm, S.J., Marsh, K., Ware, R., 1998. Radiometric profiling of temperature, water vapor and cloud liquid water using various inversion methods. Radio Sci. 33, 393–404. https://doi.org/10.1029/97RS03656@10.1002/(ISSN)1944-799X.MRRSE1

Sportisse, B., 2010. Atmospheric Boundary Layer, in: fundamentals in air Pollution. Springer Netherlands, Dordrecht, pp. 93–132. https://doi.org/10.1007/978-90-481-2970-6_4

Srivastava, D., Rajitha, B., Agarwal, S., Singh, S., 2018. Pattern-based image retrieval using GLCM. Neural Comput. Appl. 32, 1–14. https://doi.org/10.1007/s00521-018-3611-1

Stohl, A., Forster, C., Frank, A., Seibert, P., Wotawa, G., 2005. Technical note: The Lagrangian particle dispersion model FLEXPART version 6.2. Atmos. Chem. Phys. Discuss. 5, 2461–2474. https://doi.org/10.5194/acp-5-2461-2005

Strauch, R.G., Merritt, D.A., Moran, K.P., Earnshaw, K.B., De Kamp, D. Van, 1984. The Colorado wind-profiling network. J. Atmos. Ocean. Technol. 1, 37–49. https://doi.org/10.1175/1520-0426(1984)001<0037:tcwpn>2.0.co;2

Stull, R.B., 1988. An introduction to boundary layer meteorology. Kluwer Academic Publishers.

Taylor, H.S., 1938. The Balandin Multiplet Hypothesis of Dehydrogenation of Cycloparaffins. J. Am. Chem. Soc. 60, 627–632. https://doi.org/10.1021/ja01270a039

Träumner, K., Damian, T., Stawiarski, C., Wieser, A., 2015. Turbulent structures and coherence in the atmospheric surface layer. Boundary-Layer Meteorol. 154, 1-25. https://doi.org/10.1007/s10546-014-9967-6

Troude, F., Dupont, E., Carissimo, B., Flossmann, A.I., 2002. Relative influence of urban and orographic effects for low wind conditions in the Paris area. Boundary-Layer Meteorol. 103, 493–505. https://doi.org/10.1023/A:1014903627803

Tucker, S.C., Brewer, W.A., Banta, R.M., Senff, C.J., Sandberg, S.P., Law, D.C., Weickmann, A.M., Hardesty, R.M., 2009. Doppler lidar estimation of mixing height using turbulence, shear, and aerosol profiles. J. Atmos. Ocean. Technol. 26, 673-688. https://doi.org/10.1175/2008JTECHA1157.1

Tur, A. V., Levich, E., 1992. The origin of organized motion in turbulence. Fluid Dyn. Res. 10, 75–90. https://doi.org/10.1016/0169-5983(92)90009-L

Vapnik, V.N., 2000. The nature of statistical learning theory. Springer New York. https://doi.org/10.1007/978-1-4757-3264-1

Vasiljević, N., Lea, G., Courtney, M., Cariou, J.-P., Mann, J., Mikkelsen, T., 2016. Long-Range WindScanner System. Remote Sens. 8, 896.

https://doi.org/10.3390/rs8110896

Wang, C., Vandemark, D., Mouche, A., Chapron, B., Li, H., Foster, R.C., 2020. An assessment of marine atmospheric boundary layer roll detection using Sentinel-1 SAR data. Remote Sens. Environ. 250, 112031. https://doi.org/10.1016/j.rse.2020.112031

Wang, X., Wang, K., 2016. Homogenized variability of radiosonde-derived atmospheric boundary layer height over the global land surface from 1973 to 2014. J. Clim. 29, 6893–6908. https://doi.org/10.1175/JCLI-D-15-0766.1

Weckwerth, T.M., Grund, C.J., Mayor, S.D., 1997a. Linearly organized coherent structures in the surface layer, in: 12th Symposium on Boundary Layers and Turbulence. American Meteorological Society, Vancouver, BC, Canada, pp. 22–23.

Weckwerth, T.M., Grund, C.J., Mayor, S.D., 1997b. No Title, in: 12th Symposium on Boundary Layers and Turbulence. Vancouver, BC, Canada, pp. 22–23.

Weckwerth, T.M., Horst, T.W., Wilson, J.W., 1999. An observational study of the evolution of horizontal convective rolls. Mon. Weather Rev. 127, 2160–2179. https://doi.org/10.1175/1520-0493(1999)127<2160:AOSOTE>2.0.CO;2

Weckwerth, T.M., Parsons, D.B., 2006. A review of convection initiation and motivation for IHOP_2002. Mon. Weather Rev., 134, 5-22. https://doi.org/10.1175/MWR3067.1

Weitkamp, C., 2006. Lidar: range-resolved optical remote sensing of the atmosphere. Springer Science & Business, Berlin, Germany.

Wilson, R.B., Start, G.E., Dickson, C.R., Ricks, N.R., 1976. Diffusion under low windspeed conditions near Oak Ridge, Tennessee, NOAA Technical Memorandum ERL ARL-61, 83.

Woodcock, A.H., Wyman, J., 1947. Convective motion in air over the sea. Ann. N. Y. Acad. Sci. 48, 749–776. https://doi.org/10.1111/j.1749-6632.1947.tb38488.x

Xia, Q., Lin, C.L., Calhoun, R., Newsom, R.K., 2008. Retrieval of urban boundary layer structures from Doppler lidar data. Part I: Accuracy assessment. J. Atmos. Sci. 65, 3–20. https://doi.org/10.1175/2007JAS2328.1

Yagi, A., Inagaki, A., Kanda, M., Fujiwara, C., Fujiyoshi, Y., 2015. Analysis of spacing of streaky structures within surface layer above real urban, in: 9th International Conference on Urban Climate Jointly with 12th Symposium on the Urban Environment. Toulouse, France.

Yang, X., Tridandapani, S., Beitler, J.J., Yu, D.S., Yoshida, E.J., Curran, W.J., Liu, T., 2012. Ultrasound GLCM texture analysis of radiation-induced parotid-gland injury in head-and-neck cancer radiotherapy: An *in vivo* study of late toxicity. Med. Phys. 39, 5732–5739. https://doi.org/10.1118/1.4747526

Young, G.S., Kristovich, D.A.R., Hjelmfelt, M.R., Foster, R.C., 2002. Rolls, streets, waves, and more: A review of quasi-two-dimensional structures in the atmospheric boundary layer. Bull. Am. Meteorol. Soc. 83, 997-1002. https://doi.org/10.1175/1520-0477(2002)083<0997:RSWAMA>2.3.CO;2

Zhang, Y.H., Zhang, S.D., Huang, C.M., Huang, K.M., Gong, Y., Gan, Q., 2014. Diurnal variations of the planetary boundary layer height estimated from intensive radiosonde

observations over Yichang, China. Sci. China Technol. Sci. 57, 2172–2176. https://doi.org/10.1007/s11431-014-5639-5

Zilitinkevich, S., 1995. Non-local turbulent transport: pollution dispersion aspects Of coherent structure of connective flows, in: WIT Transactions on Ecology and the Environment. WIT Press, pp. 53–60. https://doi.org/10.2495/AIR950071