https://doi.org/10.5194/amt-2020-82
Preprint. Discussion started: 30 April 2020

Atmospheric Measurement Techniques Discussions

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], Marc Fourmentin[1]

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

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

Correspondence to: Ioannis Cheliotis (ioannis.cheliotis@univ-littoral.fr)

Abstract

Turbulent structures can be observed using horizontal scans from single Doppler lidar or radar systems. Despite the ability to detect the structures manually on the images, this method would be time-consuming on large datasets, thus limiting the possibilities to perform studies of the turbulent 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) and installed atop a 75-m tower in Paris city centre (France) during a 2-months campaign (September-October 2014). The lidar recorded 4577 quasi-horizontal scans for which the turbulent component of the radial wind speed was determined using the velocity azimuth display method. Three turbulent structures types were identified by visual examination of the wind fields: unaligned thermals, rolls and streaks. A learning ensemble of 150 turbulent 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 (quadratic discriminate analysis). Using the 10-fold cross validation method, the classification error was estimated to be about 9.2% for the training ensemble and 3.3% in particular for streaks. The trained algorithm applied to the whole scan ensemble detected turbulent structures on 54 % of the scans, among which 34 % were coherent turbulent structures (rolls, streaks).

https://doi.org/10.5194/amt-2020-82
Preprint. Discussion started: 30 April 2020

Atmospheric Measurement Techniques Discussions

1. Introduction

The understanding of the connection between atmospheric physics and air pollutants' dispersion is a necessary step for improving regulation and monitoring of atmospheric pollution. The level of pollution is highly dependent on the weather and particularly on turbulence (Roth, 2007). The pollution peaks occur during weather conditions where the pollutants' dispersion is restrained e.g. stagnant conditions, low-altitude thermal inversion, low turbulence (Kallos et al., 1993).

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 instantaneously space and phase correlated vorticity of the turbulent fluid mass over the spatial extend of the flow structure. Furthermore, a coherent structure must maintain its form for a time period sufficient for time-averaged statistics calculations (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.

The high wind shear between the surface layer and 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). Turbulent streaks are aligned with the horizontal wind with alternating stripes of stronger horizontal wind associated with a subsidence and stripes of weaker horizontal wind associated with an ascendance (Khanna and Brasseur, 1998). 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 meters (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 few kilometers to few tens of kilometers. They are characterized by 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 turbulent structures. Their ability to scan large areas in a short period is advantageous compared to in situ measurements (Kunkel et al., 1980). Lhermitte (1962), Browning & Wexler (1968) were the first to implement the velocity azimuth display (VAD) technique, also known as 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 & 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 type 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. The PPI method can also be applied to Doppler lidars. Depending on the selected scanning method of the Doppler lidar it is possible to observe coherent turbulent structures in the atmospheric surface layer (Drobinski et al., 2004) as well as in the mixed-layer (Drobinski et al., 1998).

https://doi.org/10.5194/amt-2020-82
Preprint. Discussion started: 30 April 2020

Atmospheric Measurement Techniques Discussions

Newsom et al. (2008) and Iwai et al. (2008) introduced the dual-Doppler lidar method and revealed its benefits in the observation of coherent turbulent 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 turbulent structures including elongated areas resembling turbulent streaks. They combined a two-dimensional autocorrelation function with the observation of the scans by eye. However, the subjective classification by observing the images is a time-consuming approach. 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 turbulent structures on sonic anemometer time series. For instance, Barthlott et al. (2007), analysed 10 month of data from a meteorological tower located in the surface layer 20 km south of Paris, France and they observed coherent turbulent structures for 36% of the cases.

This study aims to identify the coherent turbulent structures on single Doppler lidar horizontal scans and develop an automatic classification process in order to handle large datasets. The structures are determined by combining texture analysis and machine learning technique. It 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/early fall. Section 2 presents the experimental set up of the study. The methodology for the identification and classification of the turbulent structures is demonstrated in Section 0. Subsequently, the results of the classification for the training ensemble as well as for the whole dataset are displayed in Section 4. Finally, the key points of the paper are summarized in Section 5.

2. Experimental set up

A two-months campaign (04/09-06/11/2014) was carried out in Paris 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 & urban heat island)] project (Klein et al., 2019). The Leosphere WLS100 Doppler lidar (www.leosphere.com) with a minimum range of observations at 100 m (Figure 1a) was installed atop a 75 m building in the Jussieu Campus, located in the centre of Paris city (Figure 1b) and was used for wind measurements. Table 1 shows the significant lidar properties during the VEGILOT campaign.

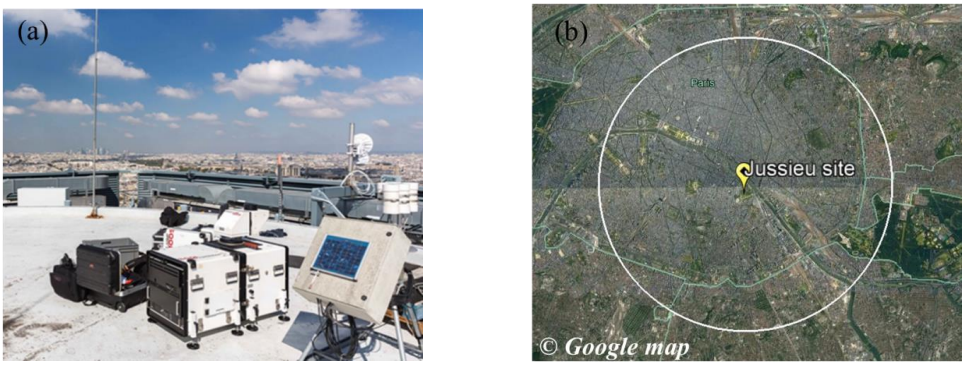


Figure 1: (a) The Doppler lidar installed on the tower roof during the VEGILOT campaign and (b) the measurement site in Paris with the circle of 10 km diameter demonstrating the maximum range of the PPI surface scan (Google earth satellite image).

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 Kumer et al. (2014) and Veselovskii et al. (2016). The lidar is sensitive only to 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 scanning sequence as it was implemented during the VEGILOT campaign. For the classification of the turbulent structures, 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° for a 2° resolution. This scenario was repeated every 18 minutes hence providing 4577 PPI scans during the whole campaign. The duration of each scan was 3 minutes which is sufficiently short for the observation of

https://doi.org/10.5194/amt-2020-82
Preprint. Discussion started: 30 April 2020

Atmospheric Measurement Techniques Discussions

structures. The maximum range of the scans reached 5 km (see white circle of Figure 1b) with a spatial resolution of 50 m. Due to the 1° elevation, the beam rise by about 87 m between the central point and the point at the 5 km.

Table 1: Properties of the lidar used for the observation of turbulent structures

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
Laser wavelength:	1.543 µm
Radial wind accuracy:	± 0.1 m/s
Accumulation time:	1 sec/beam

Table 2: Scanning sequence during VEGILOT

	Scanning area	**Purpose**	**Elevation & azimuth angle**	**Scan duration**
Range Height Indicator (RHI)	Vertical scans (North-South & East-West)	Estimation of atmospheric boundary layer height	Azimuth 0° & 90°, elevation 2° to 180° with 2° resol.	2 x 1,5 min
Plan Position Indicator (PPI)	Conical scanning	-	Elevation 25°, azimuth 0 to 360° with 2° resol.	3 min
Plan Position Indicator (PPI)	Almost horizontal scans near surface	**Identification of structures**	Elevation 1°, azimuth 0 to 360° with 2° resol.	3 min
Line of Sight (LOS)	Fixed direction shots	Identification of vertical wind profile	Elevation 75° & 90°, azimuth 0°, 90°, 180° & 270°	5+ min
Doppler Beam Swinging (DBS)	Combination of LOS	Identification of low level jet cases	Elevation 75° & 90°, azimuth 0°, 90°, 180° & 270°	2 x 15 sec

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 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 on Figure 2a) are displayed on Figure 2(b) 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 on Figure 2(c). The difference between the radial wind field u_r (Figure 2a) and the mean wind projected on the beam direction (Figure 2c) is the turbulent component of the radial wind field u_r' (Figure 2d). For a better visual representation

https://doi.org/10.5194/amt-2020-82
Preprint. Discussion started: 30 April 2020

Atmospheric Measurement Techniques Discussions

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 it is illustrated in the sign convention of Figure 2(b).

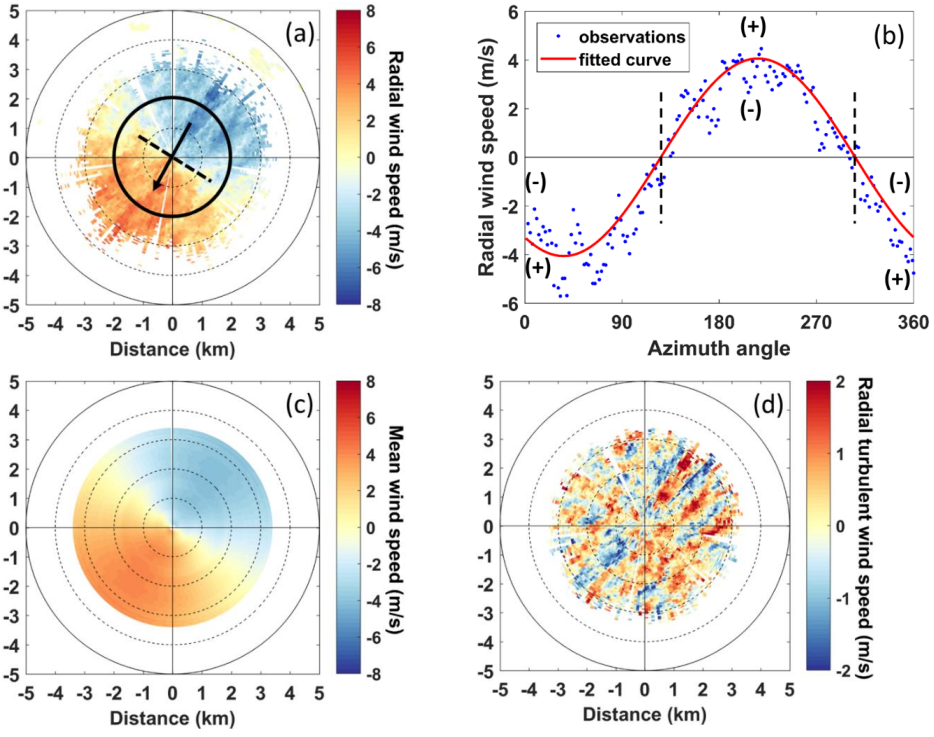


Figure 2: Observations recorded during a quasi-horizontal PPI scan on 08/09/2014 in Jussieu site, Paris at 09:26 UTC. (a) Radial wind speed along with the mean wind direction (black line) and the transverse direction perpendicular to it (black dotted line). (b) Radial wind speed (blue dots) as a function of the azimuth angle at a fixed 2 km distance from the lidar (black circle on panel a) along with the cosine fit function (red line). (c) Mean wind speed projected on the beam direction. (d) Turbulent component of the radial wind field.

Jussie site is located in an urban area nearby hills, hence the surface roughness or the orography can affect the homogeneity of the wind field. 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) do not have a defined direction (Wilson et al., 1976). Figure 3 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 (Figure 3b).

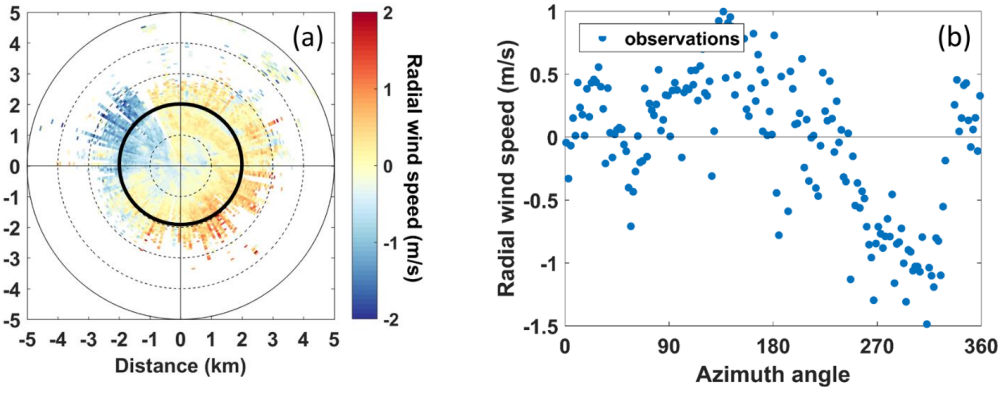


Figure 3: a case when the VAD method cannot be applied: (a)Radial wind field on 25/09/2014 at 23:42 UTC and (b) Radial wind speed (blue dots) as a function of the azimuth angle at a fixed 2 km distance from the lidar (black circle on panel a).

https://doi.org/10.5194/amt-2020-82
Preprint. Discussion started: 30 April 2020

Atmospheric Measurement Techniques Discussions

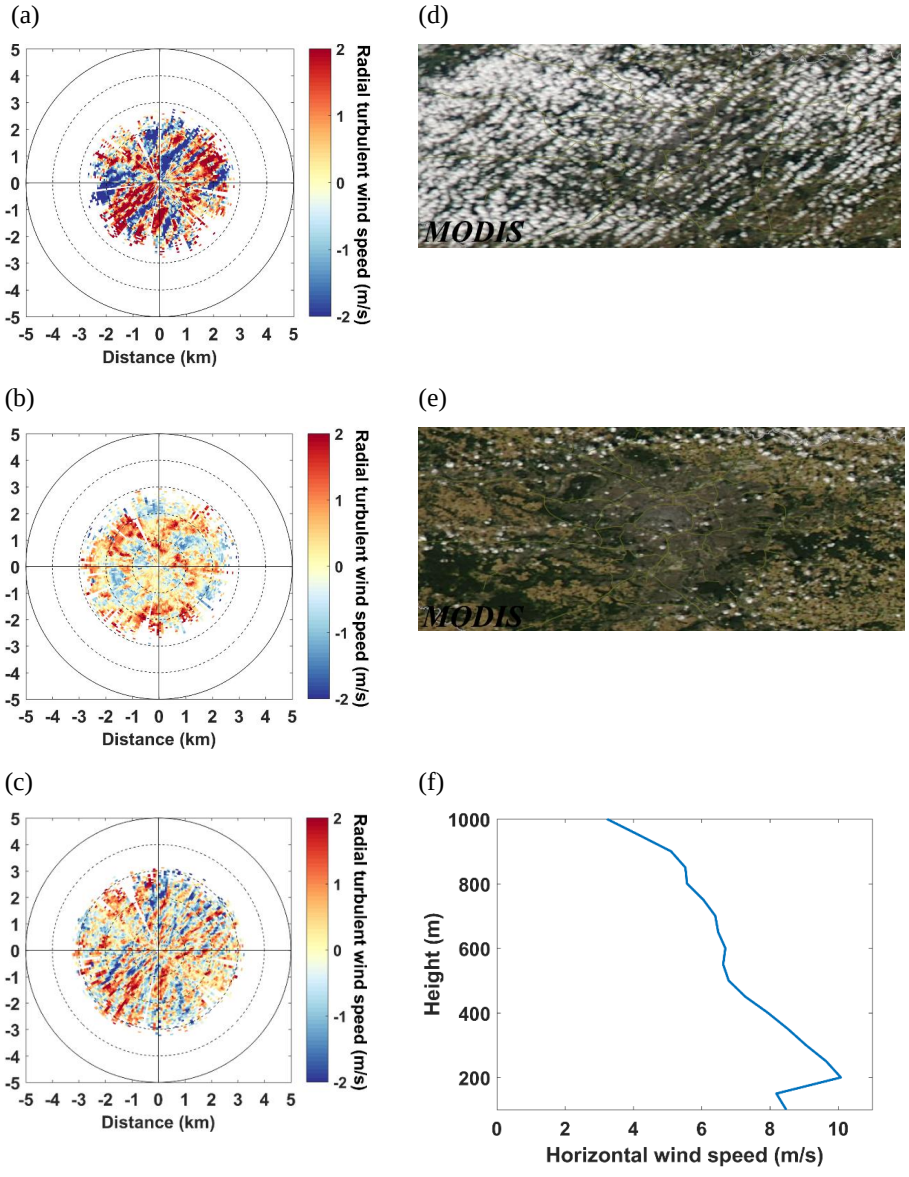
The visual examination of the turbulent radial wind fields led to the identification of three types of remarkable turbulent patterns. The first type was represented by large elongated areas of positive turbulent radial wind accompanied by large elongated areas of negative turbulent radial wind aligned with the mean wind (Figure 4a) during the day. In the atmosphere, these types of patterns are encountered concurrently with the existence of rolls, where strong ascending motions enhance the horizontal wind speed and descending motions reduce it. The second type of pattern was characterized by large enclosed areas of positive radial turbulent wind field attached to large enclosed areas of negative turbulent wind field (Figure 4b) during the day. The convergence zones formed between the positive and negative turbulent wind 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 radial turbulent wind and negative aligned with the mean wind (Figure 4c). These patterns resemble turbulent streaks as they are described in Section 1.

In order to train the classification algorithm (Sec. 4.1), it was necessary to build an ensemble of cases for which the presence of rolls, unaligned thermals or streaks was confirmed by other observations than the lidar measurements. Moderate resolution imaging spectroradiometer (MODIS) true colour images were used to detect the presence of cloud streets over Paris (Figure 4d) which confirmed the existence of rolls as stated in Section 1. Close to the moment when the cloud streets were present, rolls patterns were observed at the turbulent radial fields (Figure 4a). It is noteworthy to mention that, for the training ensemble, we selected only cases of rolls occurring 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. 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). For approximately the same time of the day, we observe the unaligned thermals patterns. Figure 4(b) 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 image at approximately the same time (Figure 4e).

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 wind shear was higher than 2 m/s, which is defined as the threshold for nocturnal low level jet events (Stull, 1988) (Figure 4f). For many cases, the wind shear was accompanied by turbulent streaks patterns (Figure 4c) so or the training ensemble, only night cases of streaks were selected to ensure that wind shear was the primary factor for the generation of turbulence. 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 (Figure 3).

https://doi.org/10.5194/amt-2020-82
Preprint. Discussion started: 30 April 2020

Atmospheric Measurement Techniques Discussions



170 Figure 4: The upper part shows the three types of remarkable turbulent radial wind fields to classify: (a) rolls observed on 13/10/2014 at 12:52 UTC, (b) unaligned thermals observed on 16/09/2014 at 12:52 UTC and (c) streaks observed on 09/09/2014 at 20:49 UTC. The lower part shows the ancillary observations used to ascertain the turbulence pattern type: (d) and (e) true color image recorded by MODIS Aqua on the same day as (a) and (b) at 12:50 UTC, (f) horizontal wind speed profile recorded by the Doppler lidar using the DBS technique on the same day as (c) at 20:51 UTC.

175 3.2 Computation of the co-occurrence matrices

In order to retrieve comparable texture analysis parameters from the turbulent wind field of the scans, the radial turbulent wind 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 180 purpose, the contrast of the images was increased by mapping the turbulent wind speed values into eight levels. One bin

https://doi.org/10.5194/amt-2020-82
Preprint. Discussion started: 30 April 2020

Atmospheric Measurement Techniques Discussions

included all the negative values below −0.5 m/s, six bins were equally distributed between −0.5 m/s and +0.5 m/s and one bin included all the positive values above +0.5 m/s (Figure 5b).

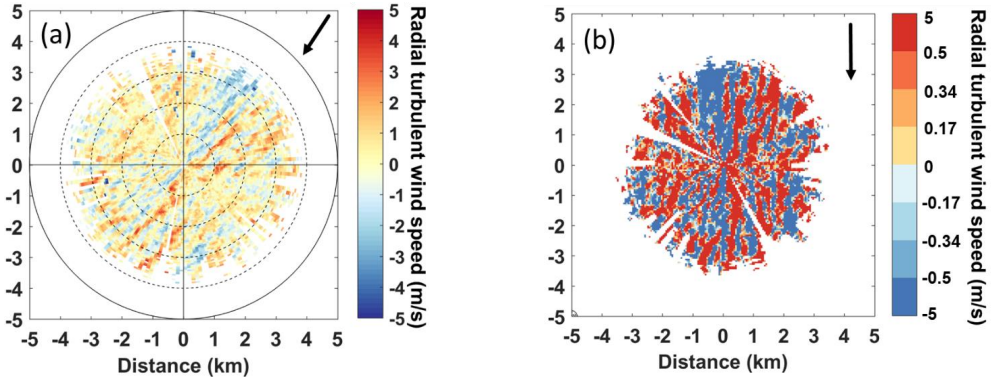


Figure 5: The radial turbulent wind field (a) before and (b) after image pre-processing with the arrow representing the mean wind direction on 10/09/2014 at 19:57.

For the automated classification of patterns, we need to map them to a space of corresponding numerical parameters. Each reconstructed turbulent wind field is represented by a matrix (cells corresponds to pixels) from which 8×8 co-occurrence matrices (CM) 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 number of occurrences of neighbour pairs with values corresponding to the row and column index. 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-right, up-down or diagonally) defining the cell pair orientation. In the same way a second, a third, etc. neighbour can be selected. Thus, the CM 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 counter clockwise direction) to +90° (transverse direction in the clockwise direction). Table 3 shows the cell values of the CM built from the image of Figure 5(b) 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.

Table 3: Co-occurrence matrix after the image pre-processing (Figure 5b) 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

https://doi.org/10.5194/amt-2020-82
Preprint. Discussion started: 30 April 2020

Atmospheric Measurement Techniques Discussions

Table 4: Co-occurrence matrix after the image pre-processing (Figure 5b) 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

On the other hand, Table 4 shows the CM of Figure 5(b) 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 on Figure 5(b), 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 (Haralick et al., 1973; Yang et al., 2012). In their study, the striped patterns resemble the elongated patterns of streaks and rolls that we observe in the radial turbulent wind field. Therefore, the same texture analysis parameters were selected for calculation in our dataset. More particularly, these parameters were computed by the Eq. (2), (3), (4) and (5):

Homogeneity:
$$Hom(\varphi, n) = \sum_{i,j} \frac{p(i,j)}{1 + |i - j|} \quad (2)$$

Contrast:
$$Con(\varphi, n) = \sum_{i,j} p(i,j)|i - j|^2 \quad (3)$$

Correlation:
$$Cor(\varphi, n) = \sum_{i,j} \frac{(i - \mu_i)(j - \mu_j)p(i,j)}{\sigma_i \sigma_j} \quad (4)$$

Energy:
$$En(\varphi, n) = \sum_{i,j} p(i,j)^2 \quad (5)$$

where $p(i,j) = \frac{CM(i,j)}{\sum_{i,j} CM(i,j)}$ 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 standard deviations $\sigma_i = \sqrt{\sum_i \sum_j (i - \mu_i)^2 \cdot p(i,j)}$, $\sigma_j = \sqrt{\sum_i \sum_j (j - \mu_j)^2 \cdot p(i,j)}$.

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 on Figure 6). The streaks and rolls have a more prominent peak in the longitudinal

https://doi.org/10.5194/amt-2020-82
Preprint. Discussion started: 30 April 2020

Atmospheric Measurement Techniques Discussions

direction ($\varphi = 0°$) compared to the unaligned thermals and "others" patterns. As streaks and rolls are aligned with the mean wind (azimuth $\varphi = 0°$), those peaks result from the elongated shapes of these patterns.

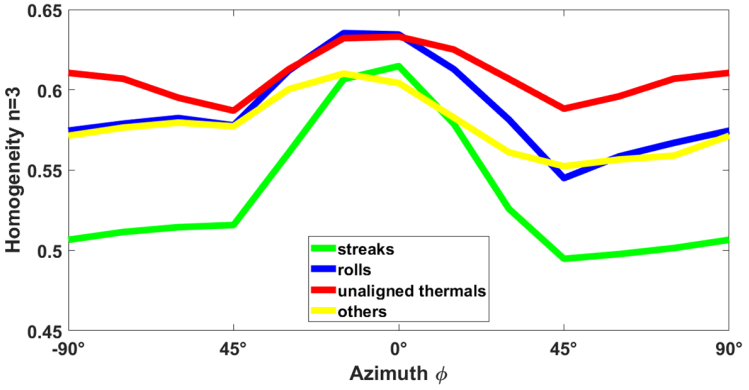


Figure 6: Third neighbour homogeneity as a function of azimuth for one selected scan of each type.

Three parameters of the curve in Figure 6 were selected in order to distinguish the different types of structures. For instance, fore the homogeneity curves, these parameters are defined by the Eq. (6), (7) and (8):

Amplitude: $$Hom.Amp(n) = \max\big(Hom(\varphi,n)\big) - \min(Hom(\varphi,n)) \quad (6)$$

Integral: $$Hom.Int(n) = \sum_{\varphi} Hom(\varphi,n) \quad (7)$$

Symmetry: $$Hom.Sym(n) = \sum_{\varphi} |Hom(\varphi,n) - Hom(-\varphi,n)| \quad (8)$$

These three curve parameters were calculated for the four texture analysis parameters and for each of the thirty neighbour orders, which gives 360 parameters. In addition to these parameters, the UTC hour (close to solar time in Paris), the average mean wind speed and the root-mean-square error of the cosine fit (Figure 2b) were included in the classification parameters. The total number of classification parameters associated with each scan was therefore 363.

4. Classification using supervised machine learning

4.1 Algorithm training and classification error

In order to classify the turbulent structures according to the aforementioned texture analysis parameters, the supervised machine learning methodology was applied. The Quadratic Discriminant Analysis (QDA) algorithm was used, that minimizes the total error probability of the classification, assuming that features of each class have a multidimensional Gaussian distribution. The QDA was trained (Hastie et al., 2009; Sokolov et al., 2020) with the 150-case ensemble described in Section 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 more cases since it is expected to be the dominant category in the classification, as it includes the chaotic radial turbulent wind fields and the cases where the turbulent wind field was not computed successfully by the VAD method.

The classification error of the QDA technique could be estimated for the training ensemble by means of the 10-fold cross validation. In this method, the algorithm is trained using 90% of the training ensemble (135 cases), then it is applied to the remaining 10% (15 cases) and the resulting (output) classes are compared to the expected (target) classes. The process is

https://doi.org/10.5194/amt-2020-82
Preprint. Discussion started: 30 April 2020

Atmospheric Measurement Techniques Discussions

repeated 10 times, each time extracting a different 10% sample for test, until the entire training ensemble has been tested. The classification error is the total number of misclassified cases, expressed as a fraction of the training ensemble.

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 Figure 7. The classification error reached a minimum of about 9.2% when five parameters were used; taking more into account increased the classification error.

In particular, these parameters were the amplitude and symmetry of the homogeneity curve, the integral and the amplitude of the contrast curve and the integral from the correlation curve. As explained in Section 0, elongated structures show a more prominent peak in the homogeneity (or the opposite for contrast) curve and as a result, they are characterized by a higher amplitude compared to unaligned thermals and "others" patterns. Similarly, the values of the integral of the curves with prominent peaks differentiate from quasi constant curves that occur for patterns with large enclosed areas (thermals) or for chaotic ones ("others"). The significant parameters cover various distances from the 2nd neighbour (100 m) to the 18th (900 m), hence demonstrating 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.

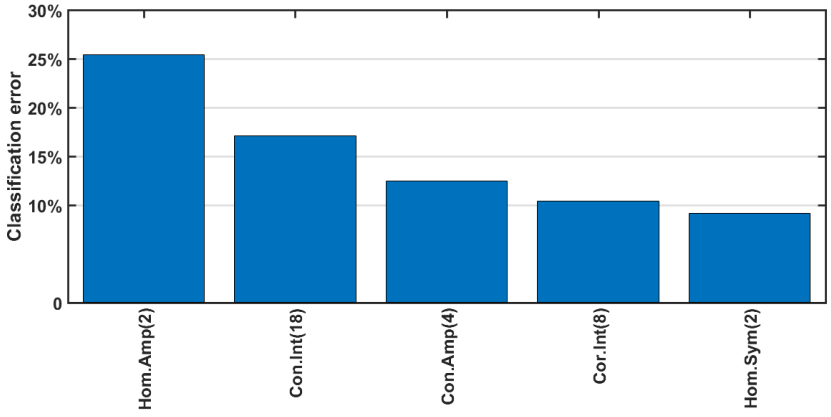


Figure 7: Parameters selected to minimize the classification error by the QDA method.

The detailed results of the cross-validation of QDA classification for the algorithm with five predictors are displayed in Table 5. 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 "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 3 cases were 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 2 cases as "others" showing a classification error of 20%.

https://doi.org/10.5194/amt-2020-82
Preprint. Discussion started: 30 April 2020

Atmospheric Measurement Techniques Discussions

Table 5: Confusion matrix calculated for the training dataset. The "target class" corresponds to the eye-made 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%

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

The whole dataset, consisting in 4577 scans, was classified according to the five parameters showcased in Figure 7. The results are displayed in Figure 8.

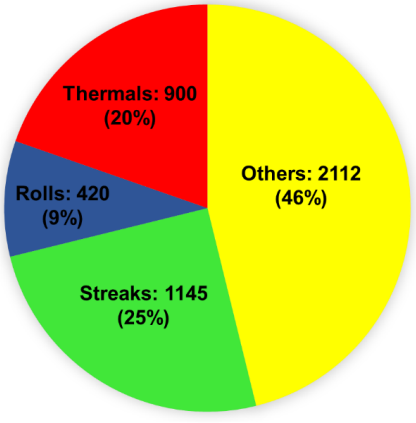


Figure 8: Classification of the whole ensemble using the QDA method according to the parameters of Figure 7.

The algorithm classifies 54% of the two-month dataset as containing turbulent structures and 34% in particular as coherent turbulent structures (streaks, rolls). The most frequent cases of turbulent structures were streaks (25%) and the least frequent were rolls (9%). It is important to note that, in our classification, we consider only thermals and rolls during daytime. Figure 9 illustrates the number of occurrences for each type of structure at a particular time of the day during the two months of the campaign. It is evident that despite time was a much less significant classifier compared to the curves parameters, the structures show a time preference. There were scarcely any rolls cases observed at night, though a few unaligned thermals were classified at night. This stems from the training process, where some cases of thermal were improperly classified as "other" and the reverse. On the contrary "others" cases were mostly observed during the night. This was expected since the cases of low winds

https://doi.org/10.5194/amt-2020-82
Preprint. Discussion started: 30 April 2020

Atmospheric Measurement Techniques Discussions

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 is a main driving factor for the formation of streaks.

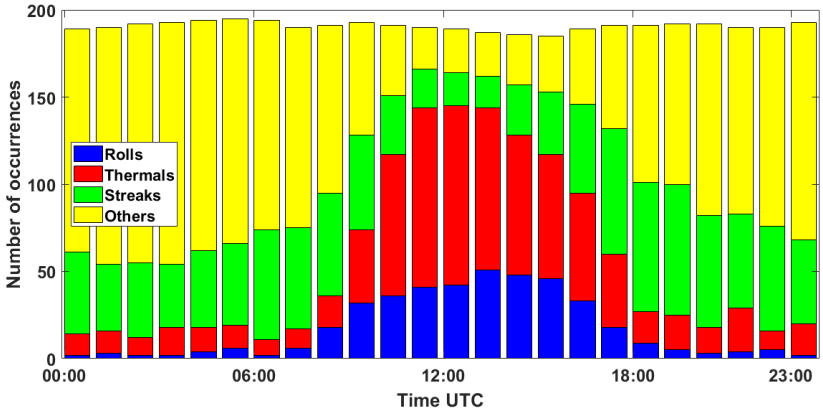


Figure 9: Histogram of the number of occurrences of the different types of structures with respect to time UTC.

5. Conclusions

The current study showcases that it is possible to identify and classify turbulent structures such as streaks, rolls and unaligned thermals with horizontal scans from a single Doppler lidar. More particularly a two-months campaign in Paris (VEGILOT) provided radial wind speed observation for 4577 horizontal scans with a 1° elevation. The VAD method was used to retrieve the radial turbulent wind field. By analysing patterns of the radial turbulent wind field, it was evident that the elongated patterns (streaks, rolls) showed a prominent peak for the curves of the texture analysis parameters compared to more chaotic or enclosed patterns (unaligned thermals).

In order to apply the supervised machine learning methodology for the classification of the ensemble, 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. For this purpose, the QDA algorithm with stepwise forward selection of the features was applied and its performance was estimated using the cross-validation technique.

The algorithm allowed classifying correctly about 91% of the dataset using five texture analysis parameters as predictors. The algorithm performed best for the category of turbulent 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 structures among which 34% were coherent turbulent structures (streaks, 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 turbulent structures 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. MODIS satellite images are publicly available following NASAs' open data policy (https://earthdata.nasa.gov/collaborate/open-data-services-and-software).

https://doi.org/10.5194/amt-2020-82
Preprint. Discussion started: 30 April 2020

Author contribution

IC, EDE, HD and AS conceptualized this study and developed the methodology. 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 AS and EDM for Section 4. IC wrote the original draft of the manuscript, with contributions from EDE and AS. All authors participated in the review and editing of the manuscript and agreed to this version.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

The authors thank F. Ravetta, J. Pelon, G. Plattner and A. Klein of the LATMOS, Sorbonne University, Paris for organizing and carrying out the VEGILOT campaign.

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).

We 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).

References

Barthlott, C., Drobinski, P., Fesquet, C., Dubos, T., Pietras, C.: 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, 2007.

Browning, K.A., 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.

Chin-Hoh Moeng, 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.

Drobinski, P., Brown, R.A., Flamant, P.H., Pelon, J.: 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, 1998.

Drobinski, P., Carlotti, P., Newsom, R.K., Banta, R.M., Foster, R.C., 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.

Drobinski, P., Foster, R.C.: 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, 2003.

Eberhard, W.L., Cupp, R.E., Healy, K.R.: 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, 1989.

Haralick, R.M., Dinstein, I., Shanmugam, K.: Textural features for image classification. IEEE Trans. Syst. Man Cybern. SMC-

3, 610–621. https://doi.org/10.1109/TSMC.1973.4309314, 1973.

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

Huld, T., Müller, R., 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.

Iwai, H., Ishii, S., Tsunematsu, N., Mizutani, K., Murayama, Y., Itabe, T., Yamada, I., Matayoshi, N., Matsushima, D., Weiming, S., Yamazaki, T., 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.

Kallos, G., Kassomenos, P., Pielke, R.A.: Synoptic and mesoscale weather conditions during air pollution episodes in Athens, Greece. Springer Netherlands, pp. 163–184. https://doi.org/10.1007/978-94-011-2749-3_9, 1993.

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., 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., 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., 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.

Kumer, V.M., Reuder, J., Furevik, B.R.: A comparison of LiDAR and radiosonde wind measurements. Energy Procedia, Elsevier Ltd, pp. 214–220. https://doi.org/10.1016/j.egypro.2014.07.230, 2014.

Kunkel, K.E., Eloranta, E.W., 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. https://doi.org/10.1175/1520-0469(1973)030<1077:tsadoh>2.0.co;2, 1972.

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

Lohou, F., Druilhet, A., Campistron, B.: 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, 1998.

Newsom, R., Calhoun, R., Ligon, D., Allwine, J.: 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, 2008.

Reinking, R.F., Doviak, R.J., 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.

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

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

Srivastava, D., Rajitha, B., Agarwal, S., Singh, S.: Pattern-based image retrieval using GLCM. Neural Comput. Appl. 1–14.

https://doi.org/10.1007/s00521-018-3611-1, 2018.

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

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

Tur, A. V., 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.

Veselovskii, I., Goloub, P., Podvin, T., Bovchaliuk, V., Derimian, Y., Augustin, P., Fourmentin, M., Tanre, D., Korenskiy, M., Whiteman, D.N., Diallo, A., Ndiaye, T., Kolgotin, A., Dubovik, O.: Retrieval of optical and physical properties of African dust from multiwavelength Raman lidar measurements during the SHADOW campaign in Senegal. Atmos. Chem. Phys 16, 7013–7028. https://doi.org/10.5194/acp-16-7013-2016, 2016.

Weckwerth, T.M., Horst, T.W., 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.

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

Wilson, R.B., Start, G.E., Dickson, C.R., Ricks, N.R.: Diffusion under low windspeed conditions near Oak Ridge, Tennessee, 1976.

Yang, X., Tridandapani, S., Beitler, J.J., Yu, D.S., Yoshida, E.J., Curran, W.J., 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., Foster, R.C.: Rolls, streets, waves, and more: A review of quasi-two-dimensional structures in the atmospheric boundary layer. Bull. Am. Meteorol. Soc. https://doi.org/10.1175/1520-0477(2002)083<0997:RSWAMA>2.3.CO;2, 2002.