Article

Wind Shear Prediction from Light Detection and Ranging Data Using Machine Learning Methods

Jingyan Huang [1], Michael Kwok Po Ng [2,*] and Pak Wai Chan [3]

1   Department of Statistics and Applied Probability, National University of Singapore,
     Singapore 119077, Singapore; jingyan_huang@u.nus.edu
2   Department of Mathematics, The University of Hong Kong, Hong Kong 999077, Hong Kong
3   Aviation Weather Services, Hong Kong Observatory, Hong Kong 999077, Hong Kong; pwchan@hko.gov.hk
*   Correspondence: michael.ng@hku.hk

Abstract: The main aim of this paper is to propose a statistical indicator for wind shear prediction from Light Detection and Ranging (LIDAR) observational data. Accurate warning signal of wind shear is particularly important for aviation safety. The main challenges are that wind shear may result from a sustained change of the headwind and the possible velocity of wind shear may have a wide range. Traditionally, aviation models based on terrain-induced setting are used to detect wind shear phenomena. Different from traditional methods, we study a statistical indicator which is used to measure the variation of headwinds from multiple headwind profiles. Because the indicator value is nonnegative, a decision rule based on one-side normal distribution is employed to distinguish wind shear cases and non-wind shear cases. Experimental results based on real data sets obtained at Hong Kong International Airport runway are presented to demonstrate that the proposed indicator is quite effective. The prediction performance of the proposed method is better than that by the supervised learning methods (LDA, KNN, SVM, and logistic regression). This model would also provide more accurate warnings of wind shear for pilots and improve the performance of Wind shear and Turbulence Warning System.

Keywords: machine learning methods; light detection and ranging data; prediction models; wind shear detection

1. Introduction

Wind shear is a kind of microscale meteorological phenomenon which refers to a sustained change in the wind direction and/or speed, resulting in a change in the headwind or tailwind encountered by an aircraft, see, for instance, [1,2]. Accurate wind shear detection is crucial for aviation safety in approach and landing. Typically, a sustained change of the headwind or tailwind of 15 knots (7.72 m/s) or more for more than a few seconds is significant wind shear. A significant wind shear occurring at low levels on approach and departure zones at airport might result in the difficulty in control for aircrafts. This losing of altitude could become very dangerous during an aircraft's landing and takeoff phases. Pilots require timely and appropriate corrective actions to ensure the safety of aircrafts when they encounter wind shear.

There were models and hardware launched at various airports to detect wind shear. Boilley and Mahfouf [3] studied a wind shear event using both observational and numerical data and demonstrated that a numerical model (Meso-NH) could be used to detect wind shear over the Nice airport. Weipert et al. [4] researched on low-level wind shear detection systems consisting of an Xband polarimetric scanning radar and an infrared scanning LIDAR (Light Detection and Ranging) to detect wind shear at the German international airports of Frankfurt and Munich. Since 2017, Airport Low-level Wind Information (ALWIN) system was operated at Tokyo International Airport (Haneda) and Narita International Airport to detect low-level wind shear, see [5].

Research on the low-level wind shear algorithm has begun using anemometers since 1970s and Doppler radars since 1980s. They are dedicated to some special weather phenomena such as gust front and microbursts/mesocyclones. The detailed information of Low-Level Windshear Alert System can be found in [6]. It is still used at some airports such as those in the United States and Taiwan, to deal with wind shear issues. The siting of the anemometers at critical locations is important for the success of the algorithm in capturing low level wind shear due to microburst. Currently, it is also effective in the detection and warning of low-level wind shear in rainy weather. However, there are limitations of the use of anemometer-based algorithms and Doppler weather radars in capturing wind shear due to dry microburst.

A method to capture low level wind shear in clear air situations is the use of wind profilers, such as the system used at Juneau, United States [7]. It is claimed to be capable of capturing wind shear due to terrain effects. However, wind profiler data consists of 10-min averages and its wind alerts may not be fast enough (i.e., not so frequently updated) to detect rapidly changing wind shear due to terrain effects, which may occur in a time interval of a few minutes or even within one minute. The detection of wind shear due to dry microburst and terrain effect points to the use of new remote sensing instrumentation for wind shear detection, namely, the Doppler weather LIDAR system. Recent advances in LIDAR techniques have spurred the development of several methods (see for instance [8–12]), for predicting wind shear from the LIDAR data. Some researchers have developed methods by using Doppler weather LIDAR to detect microburst wind shear, see [11]. In contrast, much fewer methods have been developed on the detection of terrain-induced wind shear by using Doppler weather LIDAR.

Wind shear at Hong Kong International Airport (HKIA) is terrain induced. HKIA lies to the north of the mountainous Lantau Island with highest peaks above 900 m and valleys as low as 300 m, see Figure 1. Sometimes, the flow of air across this hilly terrain could be disrupted to form mountain waves, gap outflow, etc., over the flight paths of HKIA, inducing wind shear. Most of the wind shear episodes at HKIA are related to airflow disturbances by the complex terrain near HKIA, see [1]. In addition, forecasters [8,9] in Hong Kong Observatory can issue warnings for imminent wind shear based on the broad prevailing meteorological conditions and the real time data from the Wind shear and Turbulence Warning System [2]. Currently, this warning system can generate "up-to-the-minute" alerts after wind shear is measured. The LIDAR alarm rate is about 0.76, see the results reported in [1]. In 2016, a high-resolution numerical aviation model was employed to forecast the terrain-induced wind shear at HKIA, see [12]. In 2018, a synthesized scheme based on the combination of improved signal smoothing and wind shear detection was proposed to detect wind shear. The scheme could capture 80.6% of wind shear according to pilot reports from HKIA, see [10]. The performance of this warning system should be further improved to issue timely warmings of wind shear for pilots. Therefore, further research is ongoing to explore the use of LIDAR data in detection of wind shear at HKIA.

Modern data mining techniques can be applied to various datasets used in many meteorological applications for prediction purposes. The main aim of this paper is to propose an indicator for wind shear prediction from Light Detection and Ranging (LIDAR) observational data. In the literature, the research on machine learning methods for wind shear detection with LIDAR data has received some attention. In 2012, a chaotic oscillatory-based neural network with a new learning algorithm was proposed to identify the wind shear occurrence, see [13]. Li et al. [14] designed a new method that employs a neural network as the position amendment module and exponential smoothing as the fluctuation compensation module. This method was applied to HKIA Doppler LIDAR data and a reasonable wind shear forecasting precision was obtained, see for example [15].

Different from the above traditional hardware-based and aviation methods, we propose in this paper to use supervised learning methods for wind shear detection. We will present a statistical indicator which is used to measure the variation of headwinds from multiple headwind profiles. The indicator is based on the maximum spread of measured

velocities along the range of the measurement beam with respect to different azimuth ranges. Because the indicator value is a nonnegative number, we employ a decision rule based on one-side normal distribution to distinguish a wind shear case and a non-wind shear case. Experimental results based on real data sets on two different periods obtained at Hong Kong International Airport runway are presented to demonstrate the proposed indicator is quite effective. In particular, the prediction accuracies of the proposed model are 98.0% for training data collected at 2015 and 92.0% for testing data collected at 2018, which are better than the other methods like LDA, KNN, SVM and logistic regression.

Figure 1. Map of Hong Kong International Airport and Lantau Island (height contours: 100 m) [1], with the location of the LIDAR (red square). Runway corridors are shown as pink arrows with the names marked alongside.

The outline of this paper is given as follows. In Section 2, we study the LIDAR observational data from HKIA. In Section 3, we present the proposed method. In Section 4, results are presented to demonstrate the effectiveness of the proposed method for wind shear prediction. Finally, some concluding remarks are given in Section 5.

2. LIDAR Observational Data

To measure the variable wind profile to be encountered by the aircraft at HKIA, Hong Kong Observatory developed Glide Path Scan (GPScan) strategy for the LIDAR [2]. A diagram of a glide path scan is given in Figure 2. For arrival corridors, three-degree glide paths originating from the touchdown points of the runways are considered. The laser beam of the LIDAR is used to slide along the glide paths and measure the winds. The azimuth and elevation motions of the LIDAR scanner must be configured so that the laser beam can slide smoothly. The radial wind measurements along a glide path are considered together to construct a headwind profile. During the scanning process, the wind velocities are measured at the locations represented by slant ranges (or range gates), azimuth angles, and elevation angles in three-dimensional space.

Since the aircraft on approach closely follows the three-degree glide slope in the final three nautical miles prior to touchdown on the runway, we only analyze the LIDAR of three-degree elevation angle. On the other hand, the glide path area in HKIA is measured by around 250-degree azimuth angle. Here we focus on the azimuth angles in between 220 degrees and 280 degrees. The slant ranges of the LIDAR data are from 350 to 10,000 m. However, the LIDAR data points with slant ranges over 4500 m are neglected since there are many missing Doppler velocities. In this paper, we focus on the LIDAR data at three-degree elevation angle, the azimuth angles in between 220 and 280 degrees, and the slant ranges in between 350 and 4500 m. At HKIA, the LIDAR scanner takes about four seconds

to obtain velocities at 1560 locations which refer to the combination of 39 azimuth angles and 40 slant ranges in the above setting.
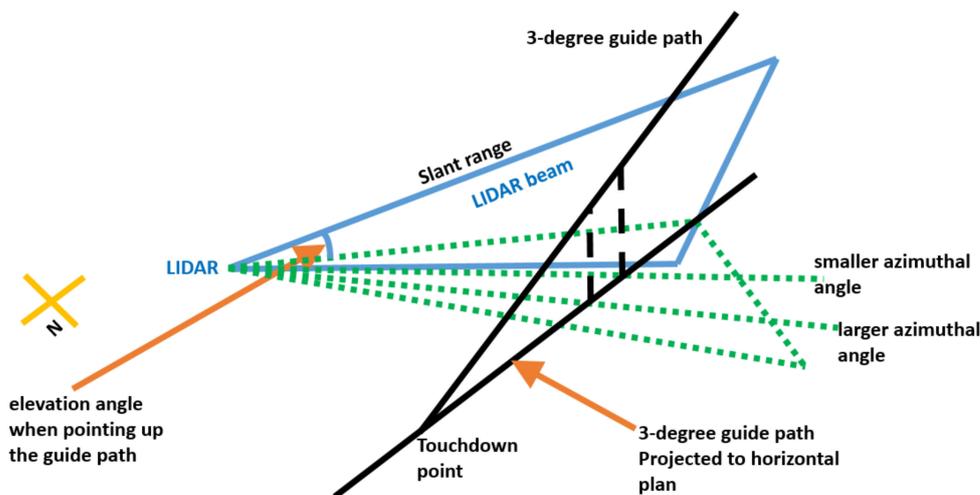

Figure 2. Diagram illustrating a LIDAR glide path scan along the 3-degree glide path for the western approach towards the north runway.

In Table 1, we show an example of the collected LIDAR data on 2 March 2015. The LIDAR cannot return actual locations in elevation angles, azimuth angles and slant ranges of measured velocities, and it only provides their minimum and maximum values. In Table 1, the differences between the maximum and minimum values in azimuth angles are about one-degree. However, the differences between the maximum and minimum values in slant ranges are from 10 to 105 m.

Table 1. An example of the collected LIDAR data at HKIA.

Time	Velocity (Knots)	Elevation Angle (Degrees)	Azimuth Angle (Degrees)	Slant Range (Meters)
2 March 2015 15:13:44	6.24	(3.00 3.00)	(220.06 221.63)	(464.49 569.49)
2 March 2015 15:13:44	4.59	(3.00 3.00)	(220.06 221.63)	(569.49 674.49)
⋮	⋮	⋮	⋮	⋮
2 March 2015 15:13:44	1.85	(3.00 3.00)	(220.06 221.63)	(4454.49 4559.49)
2 March 2015 15:13:44	7.12	(3.00 3.00)	(221.66 223.25)	(359.49 464.49)
⋮	⋮	⋮	⋮	⋮
2 March 2015 15:13:47	8.59	(3.00 3.00)	(280.88 282.44)	(4349.49 4454.49)
2 March 2015 15:13:47	8.44	(3.00 3.00)	(280.88 282.44)	(4454.49 4559.49)

3. The Detection Method

In this section, we show developed data mining techniques for wind shear prediction from Light Detection and Ranging (LIDAR) observational data. We aimed to convert the raw LIDAR dataset into headwind profiles as wind shear results from a sustained change of the headwind. Since the exact location and range of wind shear are unknown, considering LIDAR in one azimuth range from 220 to 280 degrees for each episode is not enough. For the wind shear episodes that happen within a narrow range, such as from 245 to 255 degrees, there would be a large noise by considering all LIDAR data ranging from 220 to 280 degrees. Moreover, we may lose information about wind shear if we consider a narrow azimuth range for wind shear occurring in a wide azimuth range. Therefore, for each episode, we consider seven different azimuth ranges where one headwind profile is constructed for each azimuth range. Then there are seven differ-

ent headwind profiles for each episode. The seven azimuth ranges are included as follows: $az_1 = 220°–280°$, $az_2 = 230°–270°$, $az_3 = 240°–260°$, $az_4 = 245°–255°$, $az_5 = 248°–252°$, $az_6 = 249°–250°$, $az_7 = 250°–251°$. Moreover, there are 40 measurements along the slant range. Each headwind profile is a vector of size 40. For convenience, we set $x_{i,j}$ to be the headwind profile for the *j*-th azimuth range at the *i*-th episode, where $j = 1, 2, \ldots, 7$ and $i = 1, 2, \ldots, N$. Here N is the size of the training dataset.

To detect a wind shear case, we study two statistical indicators based on seven headwind profiles to measure the variation of headwind for each episode, which are given as follows:

$$k_{i,1} = \max_{j = 1, \ldots, 7} \{\max(x_{i,j}) - \min(x_{i,j})\} \tag{1}$$

where $\max(x_{i,j})$ and $\min(x_{i,j})$ refer to the largest and smallest magnitudes of the values in $x_{i,j}$ with respect to the slant range in the headwind profile (i.e., all forty values in the vector); and

$$k_{i,2} = \max_{j = 1, \ldots, 7} \{\mathrm{var}(x_{i,j})\} \tag{2}$$

where $\mathrm{var}(x_{i,j})$ refers to the variance of the forty values in $x_{i,j}$. For $k_{i,1}$, the maximum spread of measured velocities along the range of the measurement beam (slant range) with respect to different azimuth ranges (az_j) is used. For $k_{i,2}$, the maximum variance of measured velocities along the range of the measurement beam (slant range) with respect to different azimuth ranges (az_j) is used. Both indicators are considered to check a change in the wind direction and/or speed across multiple headwind profiles.

In this paper, we propose to determine a threshold value of the indicator to identify a wind shear case or a non-wind shear case. In other words, if $k_{i,1}$ (or $k_{i,2}$) is larger than the threshold value, then a wind shear case is identified, otherwise a non-wind shear case is identified. In Figure 3, we report the histograms of $k_{i,1}$ and $k_{i,2}$ for a data set in March 2015 at Hong Kong International Airport. The detailed information of the data set will be described and used in Section 4. The values of $k_{i,1}$ and $k_{i,2}$ for wind shear cases and non-wind shear cases are shown in the histograms. We see from the figure that wind shear and non-wind shear cases cannot be distinguished clearly though $k_{i,2}$ is one measure of the variation of wind velocity. However, there is a clear boundary to separate wind shear cases and non-wind shear cases in the histogram of $k_{i,1}$. In the following discussion, we consider how to determine a decision boundary of $k_{i,1}$ for wind shear and non-wind shear cases.
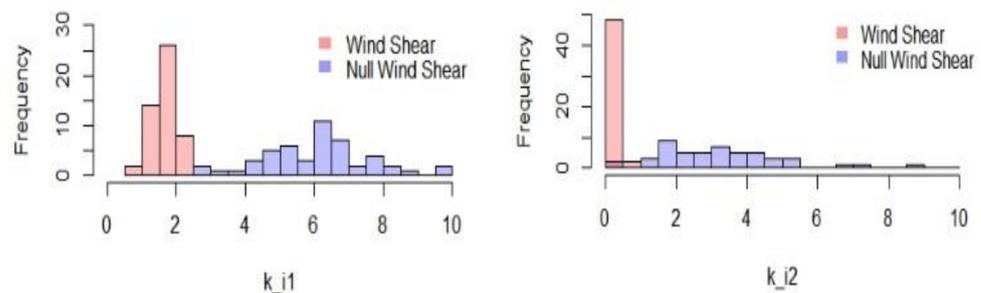

Figure 3. Histograms of $k_{i,1}$ and $k_{i,2}$ for wind shear and non-wind shear cases in March 2015. (**left**) $k_{i,1}$ for wind shear and non-wind shear cases; (**right**) $k_{i,2}$ for wind shear and non-wind shear cases.

According to (1), we know that the value of $k_{i,1}$ is nonnegative. Moreover, we use the same data set in Figure 3 to show a QQ Plot in Figure 4. We find that the values of $k_{i,1}$ for wind shear and non-wind shear cases are skewed and the measurements are not normal distributed. Because of these two issues, we may not set a boundary point using classical quadratic discriminant analysis which assumes that the measurements from each class are normally distributed (see for instance [16]).
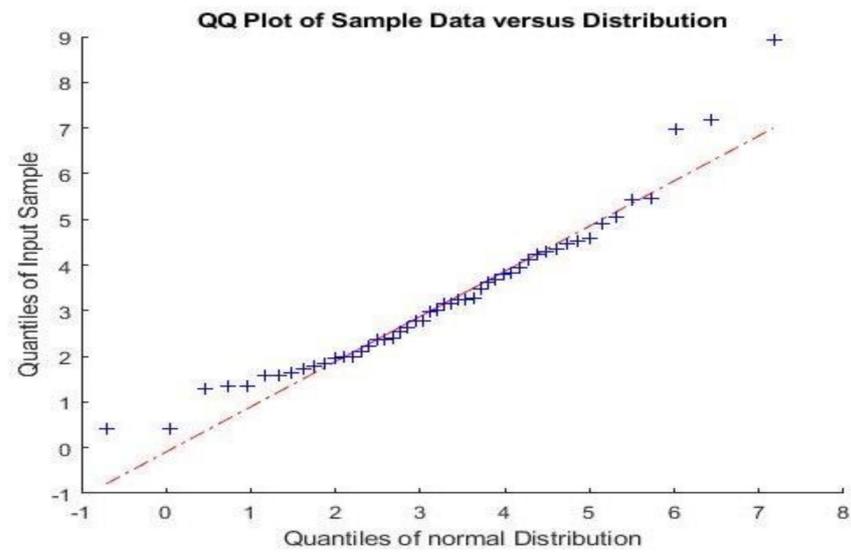
(**a**)
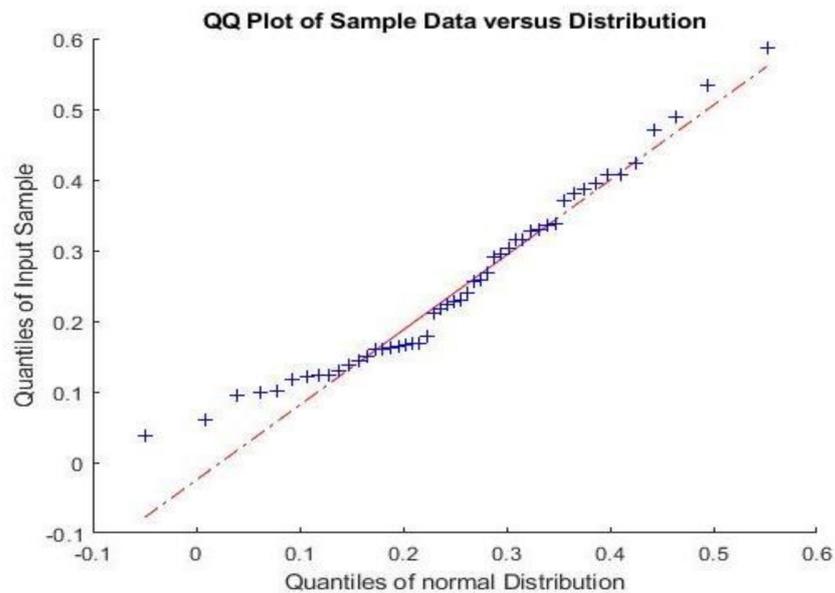
(**b**)

Figure 4. QQ Plot of $k_{i,1}$ for wind shear and null wind shear cases in March 2015. (**a**) $k_{i,1}$ for wind shear; (**b**) $k_{i,1}$ for non-wind shear cases.

In this paper, we fit one-sided normal distributions with truncation of the lower tail for $k_{i,1}$ of wind shear cases and non-wind shear cases. Then we calculate the means and standard deviations of the truncated normal distributions for wind shear cases and non-wind shear cases. The means of the truncated normal distributions for non-wind shear cases and wind shear cases are denoted as m_{non} and m_{wind}, respectively. The standard deviations of the truncated normal distributions for non-wind shear cases and wind shear cases are denoted as s_{non} and s_{wind}, respectively. A decision boundary, denoted as z, is simply selected such that the number of standard deviations from the mean of non-wind shear cases and that from the mean of wind shear cases, are equal, i.e.,

$$(z - m_{non})/s_{non} = (m_{wind} - z)/s_{wind} \tag{3}$$

Here the main idea is to find a threshold value for the change in the wind direction and/or speed (the unit is knot) that can distinguish wind shear cases and non-wind shear cases. In (3), the normalization is done within wind shear cases and non-wind shear cases. The threshold value is measured in terms of z-score in the standard normal distribution after normalization. In the next section, we validate the proposed method in our case study by using HKIA data sets.

4. Prediction Results and Discussion

The training and testing LIDAR datasets were collected for validation of our proposed method in Section 3. Both training and testing datasets were for the glide path at a specific time. Training data was collected from the LIDAR data in 2015, whereas testing data was collected from the LIDAR data in 2018. Pilot reports serve as ground truth for the occurrence of wind shear. Wind shear training data was collected according to the timestamps of wind shears from pilot reports in 2015. For each timestamp of wind shear reported, the recent timestamp from the LIDAR data was matched and their related headwind profiles were collected. Non-wind shear training data sets were collected on 2 March 2015 where there was no wind shear. Moreover, on 2 March 2015, there was no cross-mountain airflow in HKIA though the easterly winds were fresh (it was around 5 knots). Similarly, wind shear testing data was collected according to the timestamps of wind shears from pilot reports in 2018. Non-wind shear testing data was collected on the days in March 2018 when there was no wind shear under the suggestion from Hong Kong Observatory. In total, there are fifty training wind shear cases and thirty-nine testing wind shear cases. There are also fifty training non-wind shear cases and thirty-nine testing non-wind shear cases. For each case, the LIDAR scanner takes about four seconds to obtain velocities at 1560 locations (see Section 2).

In the comparison, we test supervised principal component analysis [17] that minimizes Bayes error for data classification. By applying to headwind profiles of episodes in 2015, supervised Principal Component Analysis (PCA) selects the three features $[r_1, r_2, r_3]$ that greatly separate a wind shear case and a non-wind shear case. Moreover, linear discriminant analysis (LDA), k-nearest neighbor (kNN), linear support vector machine (SVM), and logistic regression (see [16]) are used to classify wind shear cases and non-wind shear cases. LDA assumes that the covariance matrices of wind shear class and non-wind shear class are the same, and determines a linear combination of features (i.e., (r_1), (r_1, r_2), or (r_1, r_2, r_3) in our experiment) and a threshold. A wind shear case (a non-wind shear case) can be determined if the inner product between the linear combination weights and the input features of a testing example is larger than (less than or equal to) the threshold value. SVM maps input features (i.e., (r_1), (r_1, r_2), or (r_1, r_2, r_3) in our experiment) of wind shear examples and non-wind shear examples to points and maximizes the width of the gap between the wind shear cases and non-wind shear cases. The gap can be represented by the maximum-margin hyperplane. The class of a testing case based on its input features can be determined if the corresponding point lies on the one side of the hyperplane. Logistic regression assumes a linear relationship between the response variables (a wind shear case or a non-wind shear case) and the log-odds of the event of wind shear case. After the regression coefficients are estimated from input features of wind shear cases and non-wind shear cases, the logistic regression can be used to model the probability of a headwind profile existing such as a wind shear or a non-wind shear case based on input features (r_1), (r_1, r_2), or (r_1, r_2, r_3). kNN is a non-parametric classification method with the headwind profile being assigned to the wind shear case or non-wind shear case most common among its k nearest neighbors.

The proposed detection method in Section 3 is to compute a threshold about the change in the wind direction and/or speed in the boundary of two one-sided normal distributions with truncation of the lower tails for wind shear cases and non-wind shear cases. The proposed method assumes that the distributions of wind shear cases and non-wind shear cases are one-sided normal which is different from LDA. Moreover, we do not employ a

linear combination of weights in LDA, a hyperplane in SVM and a linear relationship in logistic regression. Instead of using neighbors in kNN, we consider the distributions of wind shear examples and non-wind shear examples in the proposed method. The proposed method is different from the above compared classification methods.

4.1. Results for 2015

By applying to headwind profiles of episodes in 2015, supervised PCA selects the three features $[r_1, r_2, r_3]$ that greatly separate a wind shear case and a non-wind shear case. Then we apply the features selected by supervised PCA to wind shear cases and non-wind shear cases for classification. The three selected features are tested with LDA, kNN, SVM, and logistic regression. Here a fivefold cross validation procedure is applied to each supervised learning method. In fivefold cross-validation, the wind shear cases (or non-wind shear cases) are randomly partitioned into five equal sized subsamples. Each subsample contains 10 windshear cases and 10 non-wind shear cases. Of the five subsamples, a single subsample is retained as the validation data for testing the model, and the remaining four subsamples are used as training data. The cross-validation process is then repeated five times, with each of the five subsamples used exactly once as the validation data. The five results can then be averaged to produce cross-validation accuracy results. In Tables 2–4, we show the cross-validation results based on three different combinations of supervised PCA features ((r_1), (r_1, r_2), (r_1, r_2, r_3)), see Supplementary Materials.

Table 2. Cross-validation results of wind shear cases and non-wind shear cases in March 2015 under four classification methods with supervised PCA with (r_1).

Classification Method	Prediction	Ground Truth Wind Shear	Ground Truth Non-Wind Shear	Cross-Validation Accuracy
Linear SVM	Wind shear Non-wind shear	43.0 7.0	3.0 47.0	0.900
LDA	Wind shear Non-wind shear	41.0 9.0	1.0 49.0	0.900
Logistic Regression	Wind shear Non-wind shear	44.0 6.0	1.0 49.0	0.930
kNN (k = 3)	Wind shear Non-wind shear	44.0 6.0	3.0 47.0	0.910

Table 3. Cross-validation results of wind shear cases and non-wind shear cases in March 2015 under four classification methods with supervised PCA with (r_1, r_2).

Classification Method	Prediction	Ground Truth Wind Shear	Ground Truth Non-Wind Shear	Cross-Validation Accuracy
Linear SVM	Wind shear Non-wind shear	42.0 8.0	3.0 47.0	0.890
LDA	Wind shear Non-wind shear	41.0 9.0	0.0 50.0	0.900
Logistic Regression	Wind shear Non-wind shear	Logistic model is not significant to be reported.		Nil
kNN (k = 3)	Wind shear Non-wind shear	44.0 6.0	1.0 49.0	0.930

For logistic regression based on (r_1, r_2) or (r_1, r_2, r_3), there are more than one group of parameters estimated not significant at the 0.05 level of significance. Therefore, logistic regression models are not significant in Tables 3 and 4. Tables 2–4 show that four classifica-

tion algorithms based on supervised PCA features could predict wind shear and non-wind shear with a cross-validation accuracy up to 0.930.

Table 4. Cross-validation results of wind shear cases and non-wind shear cases in March 2015 under four classification methods with supervised PCA with (r_1, r_2, r_3).

Classification Method	Prediction	Ground Truth Wind Shear	Ground Truth Non-Wind Shear	Cross-Validation Accuracy
Linear SVM	Wind shear Non-wind shear	45.0 5.0	2.0 48.0	0.930
LDA	Wind shear Non-wind shear	38.0 12.0	1.0 49.0	0.870
Logistic Regression	Wind shear Non-wind shear	Logistic model is not significant to be reported.		Nil
kNN (k = 3)	Wind shear Non-wind shear	44.0 6.0	1.0 49.0	0.930

In addition, we show the fivefold cross-validation results for linear SVM, LDA, kNN, and logistic regression based on the statistical indicators $k_{i,1}$ and $k_{i,2}$ in (1) and (2), see Supplementary Materials. We test on two combinations of statistical indicators: $(k_{i,1})$ and $(k_{i,1}, k_{i,2})$. The cross-validation results are summarized in Tables 5 and 6. We note from Tables 5 and 6 that kNN performs best in terms of the cross-validation accuracy for wind shear cases in March 2015, which are 0.960 based on $(k_{i,1})$ and 0.940 based on $(k_{i,1}, k_{i,2})$. For the general prediction of wind shear cases and non-wind shear cases in March 2015, kNN also reaches the most satisfactory cross-validation results which are 0.950 based on $(k_{i,1})$ and 0.940 based on $(k_{i,1}, k_{i,2})$. We observe in the four machine learning methods that the proposed statistical indicator $k_{i,1}$ is more effective than the supervised PCA features and the statistical indicator $k_{i,2}$ in terms of cross-validation accuracy for wind shear cases and non-wind shear cases.

Table 5. Cross-validation results of wind shear cases and non-wind shear cases in March 2015 under four classification methods with supervised PCA with $(k_{i,1})$.

Classification Method	Prediction	Ground Truth Wind Shear	Ground Truth Non-Wind Shear	Cross-Validation Accuracy
Linear SVM	Wind shear Non-wind shear	45.0 5.0	3.0 47.0	0.920
LDA	Wind shear Non-wind shear	42.0 8.0	1.0 49.0	0.910
Logistic Regression	Wind shear Non-wind shear	47.0 3.0	3.0 47.0	0.940
kNN (k = 3)	Wind shear Non-wind shear	48.0 2.0	3.0 47.0	0.950

On the other hand, we apply the proposed data analytics algorithm in Section 3 to training datasets in 2015. Fivefold cross-validation is applied to the statistical indicator $k_{i,1}$ of training datasets to set a decision boundary according to (3). The cross-validation results of wind shear and nonwind shear in March 2015 are presented in Table 7. The cross-validation accuracy is 0.980, which is better than those of other methods by at least 3.00%. In Table 8, we show the values of m_{non}, m_{wind}, s_{non}, and s_{wind} and the determined threshold value z in the five cross-validation tests. Note that the average threshold value is $z = 2.290$, i.e., if $k_{i,1}$ (the maximum spread of measured velocities along the range of the measurement beam) is greater than or equal to 2.290, then it is a wind shear case; otherwise, it is a non-wind shear case. Correspondingly, the number $(m_{wind} - 2.290)/s_{wind}$

of standard deviations of the truncated normal distribution of wind shear cases and the number $(2.290 - m_{non})/s_{non}$ of standard deviations of the truncated normal distribution of non-wind shear cases are about 2.458.

Table 6. Cross-validation results of wind shear cases and non-wind shear cases in March 2015 under four classification methods with supervised PCA with $(k_{i,1}, k_{i,2})$.

Classification Method	Prediction	Ground Truth Wind Shear	Ground Truth Non-Wind Shear	Cross-Validation Accuracy
Linear SVM	Wind shear Non-wind shear	43.0 7.0	3.0 47.0	0.900
LDA	Wind shear Non-wind shear	42.0 8.0	2.0 48.0	0.900
Logistic Regression	Wind shear Non-wind shear	45.0 5.0	3.0 47.0	0.920
kNN (k = 3)	Wind shear Non-wind shear	47.0 3.0	3.0 47.0	0.940

Table 7. Cross-validation results of wind shear and non-wind shear in March 2015 under the proposed data analytics algorithm.

Prediction	Ground Truth Wind Shear	Ground Truth Non-Wind Shear	Cross-Validation Accuracy
Wind shear Non-wind shear	50.0 0.0	2.0 48.0	0.980

Table 8. The values of m_{non}, m_{wind}, s_{non}, and s_{wind} and the determined threshold value z in the five cross-validation tests.

Mean (Wind Shear Case)	Standard Deviation (Wind Shear Case)	Mean (Non-Wind Shear Case)	Standard Deviation (Non-Wind Shear Case)	Determined Threshold Value
5.137	1.452	1.550	0.307	2.062
6.226	1.248	1.284	0.265	2.098
6.313	1.794	1.811	0.118	2.070
6.750	1.879	1.646	0.285	2.597
6.084	1.936	1.466	0.303	2.646

4.2. Results for 2018

In this subsection, we show the testing results for LIDAR in 2018 under the models trained by the LIDAR data in March 2015. There are thirty-nine wind shear cases and thirty-nine non-wind shear cases in March 2018. It is an out-of-training-sample test.

For supervised PCA with (r_1), we apply five training models under each classification method respectively to test all the seventy-eight cases in March 2018. In Table 9, we calculate the wind shear prediction accuracy and non-wind shear prediction accuracy by averaging the results over five training models for each classification method. The prediction accuracy is calculated by averaging wind shear prediction accuracy and non-wind shear prediction accuracy for each classification method. Similarly, for supervised PCA with (r_1, r_2) and (r_1, r_2, r_3), the prediction accuracy results are summarized in Tables 10 and 11. Furthermore, the prediction results based on the proposed statistical indicators $(k_{i,1})$ and $(k_{i,1}, k_{i,2})$ are shown in Tables 12 and 13, respectively. Moreover, we apply the determined threshold value $(z = 2.290)$ trained from the LIDAR data in March 2015 to the LIDAR data in March 2018 to detect wind shear cases and non-wind shear cases. The prediction accuracy results are shown in Table 14.

Table 9. Cross-validation results of wind shear cases and non-wind shear cases in March 2018 under four classification methods with supervised PCA with (r_1).

Classification Method	Wind Shear Average Prediction Accuracy	Non-Wind Shear Average Prediction Accuracy	Average Prediction Accuracy
Linear SVM	0.530	0.990	0.760
LDA	0.440	1.000	0.720
Logistic Regression	0.670	0.970	0.820
kNN (k = 3)	0.550	0.980	0.770

Table 10. Cross-validation results of wind shear cases and non-wind shear cases in March 2018 under four classification methods with supervised PCA with (r_1, r_2).

Classification Method	Wind Shear Average Prediction Accuracy	Non-Wind Shear Average Prediction Accuracy	Average Prediction Accuracy
Linear SVM	1.000	0.000	0.500
LDA	0.970	0.000	0.490
Logistic Regression	0.960	0.190	0.580
kNN (k = 3)	0.970	0.650	0.810

Table 11. Cross-validation results of wind shear cases and non-wind shear cases in March 2018 under four classification methods with supervised PCA with (r_1, r_2, r_3).

Classification Method	Wind Shear Average Prediction Accuracy	Non-Wind Shear Average Prediction Accuracy	Average Prediction Accuracy
Linear SVM	1.000	0.000	0.500
LDA	0.970	0.000	0.490
Logistic Regression	0.970	0.000	0.490
kNN (k = 3)	0.970	0.650	0.810

Table 12. Cross-validation results of wind shear cases and non-wind shear cases in March 2018 under four classification methods with supervised PCA with ($k_{i,1}$).

Classification Method	Wind Shear Average Prediction Accuracy	Non-Wind Shear Average Prediction Accuracy	Average Prediction Accuracy
Linear SVM	0.890	0.880	0.890
LDA	0.760	0.950	0.860
Logistic Regression	0.950	0.870	0.910
kNN (k = 3)	0.940	0.880	0.910

Table 13. Cross-validation results of wind shear cases and non-wind shear cases in March 2018 under four classification methods with supervised PCA with ($k_{i,1}$, $k_{i,2}$).

Classification Method	Wind Shear Average Prediction Accuracy	Non-Wind Shear Average Prediction Accuracy	Average Prediction Accuracy
Linear SVM	0.900	0.890	0.900
LDA	0.760	0.920	0.840
Logistic Regression	0.860	0.870	0.870
kNN (k = 3)	0.930	0.880	0.910

By comparing the prediction accuracy results in Table 14 with those in Tables 9–13, we observe that the proposed statistical indicator $k_{i,1}$ is more effective than supervised PCA features and statistical indicator $k_{i,2}$. The prediction accuracy of wind shear cases and non-wind shear cases in March 2018 by the proposed data analytics algorithm is 0.920, which is higher than the average prediction accuracy results of other machine learning

methods based on (r_1), (r_1, r_2), (r_1, r_2, r_3), $(k_{i,1})$ and $(k_{i,1}, k_{i,2})$. We also see from Tables 9–13 that the prediction accuracy of wind shear cases is very high, but the prediction accuracy of non-wind shear cases is very low. The models trained by the LIDAR data in March 2015 may be overfitting.

Table 14. Cross-validation results of wind shear cases and non-wind shear cases in March 2018 under the proposed data analytics algorithm.

Prediction	Ground Truth Wind Shear	Ground Truth Non-Wind Shear	Prediction Accuracy
Wind shear	38	5	0.920
Non-wind shear	1	34	

On the other hand, the prediction accuracy by the proposed data analytics algorithm is also higher than 0.806, the prediction accuracy of wind shear cases and non-wind shear cases by a recent synthesized scheme in [10]. Overall, our data analytics algorithm has a good performance in predictions of wind shear cases and non-wind shear cases, which could be applied to more the LIDAR data to predict wind shear.

5. Conclusions

In this paper, we construct headwind profiles from the LIDAR observations. With the unknown location and range of wind shear, seven different azimuth ranges are considered where one headwind profile is constructed for each azimuth range, which could gain a higher accuracy than one headwind profile for measuring a wind shear case. After that, a statistical indicator is proposed on the seven headwind profiles to measure the variation of headwinds for each episode. Then a decision rule is generated from the proposed statistical indicator. A wind shear case is identified when the indicator is larger than 2.290 that is obtained by using the LIDAR data in 2015, a non-wind shear case otherwise. The training and testing results in 2015 and 2018 respectively show that the proposed statistical indicator is more effective than supervised PCA features. Furthermore, the prediction accuracies of the proposed model are 98.0% for training data and 92.0% for testing data, which are better than the other methods like LDA, KNN, linear SVM and logistic regression.

In the future, the proposed model could be applied to more LIDAR datasets to predict wind shear. We plan to check the performance of the decision point to the LIDAR data in different locations and different periods. We would like to improve this model that can provide more accurate warnings of wind shear for pilots and improve the performance of Wind shear and Turbulence Warning System. On the other hand, doppler LIDARs are increasingly more commonly used for detecting airflow turbulence. The turbulence may arise from natural terrain, or even by man-made structures and buildings. The behavior of the turbulence from Terrain and buildings may be different. As a future research topic, the methods studied in the present paper may be applied to other meteorological conditions, such as tropical cyclones, thunderstorms, and other kinds of disturbed airflow, such as building effects on the low-level wind.

References

1. Chan, P.; Shun, C.; Wu, K. Operational Lidar-Based System for Automatic Windshear Alerting at the Hong Kong International Airport. Reprint 613, Hong Kong Observatory. Available online: www.hko.gov.hk/hko/publica/reprint/r613.pdf (accessed on 15 April 2021).
2. Windshear and Turbulence in Hong Kong—Information for Pilots. Available online: http://www.hko.gov.hk/en/aviat/articles/files/WS-turb-booklet-eng-3rd.pdf (accessed on 15 April 2021).
3. Boille, A.; Mahfouf, J. Wind shear over the Nice Cote d'Azur airport: Case studies. *Nat. Hazards Earth Syst. Sci.* **2013**, *13*, 2223–2238. [CrossRef]
4. Weipert, A.; Kauczok, S.; Hannesen, R.; Ernsdorf, T.; Stiller, B. Wind shear detection using radar and lidar at Frankfurt and Munich airports. In Proceedings of the Eighth European Conference on Rader in Meteorology and Hydrology, Garmisch-Partenkirchen, Garmisch-Partenkirchen, Germany, 1 September 2014; Abstract ID 058. 2014. Available online: http://www.pa.op.dlr.de/erad2014/programme/ExtendedAbstracts/058_Weipert (accessed on 15 April 2021).
5. Airport Low-Level Wind Information (ALWIN) Starts Operation at Haneda and Narita Airports. Available online: https://global.jaxa.jp/press/2017/04/20170419_alwin.html (accessed on 15 April 2021).
6. Low Level Wind Shear Alert System (LLWAS). Available online: https://ral.ucar.edu/projects/low-level-wind-shear-alert-system-llwas (accessed on 15 April 2021).
7. Juneau Airport Wind System (JAWS). Available online: https://ral.ucar.edu/projects/juneau-airport-wind-system-jaws (accessed on 15 April 2021).
8. Shun, C.; Chan, P. Applications of an infrared doppler lidar in detection of windshear. *J. Atmos. Ocean Technol.* **2007**, *25*, 637–694. [CrossRef]
9. Chan, P.; Hon, K. Observation and numerical simulation of terrain-induced windshear at the Hong Kong international airport in a planetary boundary layer without temperature inversions. *Adv. Meteorol.* **2016**, *2016*, 1–9. [CrossRef]
10. Meng, L.; Xu, J.; Xiong, X.; Ma, Y.; Zhao, Y. A novel ramp method based on improved smoothing algorithm and second recognition for windshear detection using lidar. *Curr. Opt. Photonics* **2018**, *2*, 7–14.
11. Keohan, C.; Keith, B.; Hannon, S. Evaluation of pulsed lidar wind hazard detection at Las Vegas international airport. In Proceedings of the 12th Conference on Aviation Range & Aerospace Meteorology, American Meteor Society, Atlanta, GA, USA, 29 January–2 February 2006. Available online: https://ams.confex.com/ams/pdfpapers/105481.pdf (accessed on 15 April 2021).
12. Chan, P.; Hon, K. Performance of super high resolution numerical weather prediction model in forecasting terrain disrupted airflow at the Hong Kong international airport: Case studies. *Meteor. Appl.* **2016**, *23*, 101–114. [CrossRef]
13. Liu, N.; Kwong, K.; Chan, P. Chaotic oscillatory-based neural network for wind shear and turbulence forecast with lidar data. *IEEE Trans. Syst. Man Cybern. Part C* **2012**, *42*, 1412–1423. [CrossRef]
14. Li, Y.; Hu, Q.; Liu, S. Wind-shear prediction with airport lidar data. In Proceedings of the 2012 IEEE International Geoscience and Remote Sensing Symposium, Munich, Germany, 22–27 July 2012; pp. 3704–3707.
15. Chan, P.; Shun, C. Application of a ground-based doppler lidar to automatic windshear alerting. In Proceedings of the 1st AIAA Atmospheric and Space Environments Conference, San Antonio, TX, USA, 22–25 June 2009; AIAA 2009-3870. Available online: https://arc.aiaa.org/doi/10.2514/6.2009-3870 (accessed on 15 April 2021).
16. Aggarwal, C. *Data Classification: Algorithms and Applications*; IBM T. J. Watson Research Center Yorktown Heights: New York, NY, USA, 2014.
17. De Carvalho, T.; Sibaldo, M.; Tsang, I.; da Cunha Cavalcanti, G. Principal component analysis for supervised learning: A minimum classification error approach. *J. Inf. Data Manag.* **2017**, *8*, 131–145.