Machine Learning-Based Turbulence-Risk Prediction Method for the Safe Operation of Aircrafts

Shinya Mizuno (s.mzn.eng@gmail.com)
Shizuoka Institute of Science and Technology: Shizuoka Rikoka Daigaku https://orcid.org/0000-0001-6030-3589

Haruka Ohba
Shizuoka Institute of Science and Technology

Koji Ito
J.F. Oberlin University

Methodology

Keywords: turbulence, machine learning, PCA, k-means, SVC, prediction, open data

Posted Date: November 18th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1051299/v1

Version of Record: A version of this preprint was published at Journal of Big Data on March 7th, 2022. See the published version at https://doi.org/10.1186/s40537-022-00584-5.

Abstract

Customer comfort is an important requirement for airlines, and avoiding and mitigating aircraft shaking have always been crucial in this regard. In particular, managing aircraft operations during turbulence is a major issue for airlines. We propose a method for predicting the occurrence of turbulence to support the safe and comfortable operation of aircrafts. Our method integrates meteorological data from Japan and turbulence information provided by Fuji Dream Airlines. Because turbulence occurs rarely, we define a risk cluster that includes turbulence observation data and use it as turbulence training data. Hence, we first estimated the risk cluster, then performed a principal component analysis (PCA) on meteorological data to obtain a projection matrix W for reducing data dimensions. Using the turbulence-occurrence indicator and the meteorological data coordinates linear transformed by W, we calculated the risk cluster using the k-means method which, in turn, was used in conjunction with support vector classification (SVC) to predict the turbulence-risk dates based on meteorological data from 2019. The results revealed that the days with turbulence risks were accurately identified from the meteorological data; thus, we believe that this method can help support the safe operation of aircrafts. Furthermore, we believe this study will lead to the development of human resources by providing a guide for making safety decisions through the effective use of aviation data.

1. Introduction

Providing a comfortable space for customers is an important requirement for airlines, and the avoidance and mitigation of aircraft shaking have always been crucial. Turbulence is often the cause of aviation accidents [1, 2]. In addition, the potential increase in aircraft turbulence owing to the effects of global warming is a prevalent concern [3].

When a pilot reports encountering one or more instances of severe turbulence during a flight, the corresponding aircraft must undergo maintenance work to confirm its airworthiness. Therefore, turbulence remains a major issue for airlines. In addition, if the maximum acceleration recorded exceeds the operational acceleration limit of the aircraft, the scope of maintenance work considerably increases, which significantly impacts aircraft operation schedules. Therefore, severe turbulence should be avoided as much as possible. However, if turbulence reporting relies primarily on the opinions of pilots, which tend to vary, variations in reports provided by pilots are inevitable.

Most existing research concerning turbulence prediction has been performed from a meteorological viewpoint [4, 5]. In these studies, data were acquired in real time from many sensors and analyzed using a time-series approach [6]. Although turbulence forecasting with pinpoint accuracy is desirable, it is expensive and infeasible for airlines to prepare a suitable environment for the sensors necessary to achieve this. In recent years, owing to the accumulation of aviation data and improvements in computation rapidity, the concept of turbulence prediction via machine learning has been introduced [7, 8]. However, studies concerning this subject are limited. Furthermore, it is difficult to determine an optimal machine learning approach for turbulence prediction. There exists a need to utilize open data (such as

meteorological data) to improve analysis accuracy; this could aid in developing turbulence predictions that can be logically deduced from the data provided by the airlines.

In this paper, we propose a method for predicting turbulence occurrence, to contribute to the safe and comfortable operation of aircrafts. Figure 1 outlines this method, which involves the accumulation and aggregation of open data and quick access recorder (QAR) data [9, 10], and the prediction of turbulence using machine learning methods, the results of which are fed back to airlines and pilots. Flights to and from Matsumoto Airport in Japan, on E-170 aircrafts operated by Fuji Dream Airlines (FDA), frequently experience turbulence during the winter season. In this study, we consider the Matsumoto Airport as the model airport representing mountainous areas subject to turbulence. This technique can also be adapted to other airports.

For our study, we used meteorological data from Japan and turbulence information provided by FDA. Because turbulence is a relatively rare event, we first estimated the risk cluster. To this end, we performed a principal component analysis (PCA) of the meteorological data to obtain a projection matrix W to reduce the number of dimensions of the data to be analyzed. Subsequently, using the turbulence-occurrence indicator and the meteorological data transformed by W, we calculated the risk cluster using the k-means method. We used this risk cluster to predict the days with turbulence risk for meteorological data from the year 2019 through support vector classification (SVC). The results based on this meteorological data revealed that the prediction method accurately identified the days with a risk of turbulence.

We believe that the integration and utilization of open data [11] (such as meteorological data and aviation data accumulated by airlines) will be promoted through this study; we demonstrated the possibility of calculating logical criteria for determining turbulence using machine learning. We expect that our research will not only improve the safety of aircraft operations but also lead to the development of human resources by providing a guide for making safety decisions, thereby promoting the effective utilization of aviation data.

2. Basic Analysis Of Turbulence At Matsumoto Airport

In this section we describe a basic analysis of the data collected at Matsumoto Airport.

2.1. Examples of the effects of turbulence on flights from Matsumoto Airport

From the topographical characteristics shown in Figure 2, it can be inferred that flights operating from Matsumoto Airport are susceptible to mountain waves [12] from the Northern Alps, particularly on the route toward New Chitose Airport.

Table 1 summarizes the turbulence, presumably caused by mountain waves, reported by flights departing from the Matsumoto Airport. The authors were present on the December 27th, 2017 flight, to gain a real-

world understanding of the level of turbulence faced during a flight.

Table 1
Examples of the impact of turbulence on operations

Date	Flight route	Turbulence reported
12/12/2017	MMJ[1] to FUK[2]	Encountered moderate-plus turbulence while climbing.
12/27/2017	MMJ[1] to FUK[2]	Encountered moderate-plus turbulence while climbing.
01/23/2018	MMJ to CTS[3]	Encountered severe turbulence while climbing. Destination changed to NKM[4]. The maintenance inspection reported no problems.

[1] Matsumoto Airport; [2] Fukuoka Airport; [3] New Chitose Airport; [4] Nagoya Airfield.

2.2. Visualization of the wind direction, speed of mountain waves, and sway of aircrafts

We believe that visualizing turbulence and the resulting impact on operations is the first step toward solving the problem. To this end, we created a visualization of a severe turbulence scenario. We visualized the altitude changes during turbulence by modeling the flight of an E170 aircraft and depicting the aircraft altitude at every second using Google Earth Pro 7.3.4. Figure 3 shows the visual representation of a journey via FDA Flight 211 in January 2018, wherein the pilot encountered severe turbulence during ascent. The average wind directions during the turbulence are represented using red lines. The wind blew over the Northern Alps in a direction toward the aircraft (from the back to the front of the figure). Significant altitude changes were observed during this period.

2.3. Elementary analysis of turbulence occurrences using open data

To create our dataset, we obtained weather information from October 1, 2017, through March 31, 2018, from Sunny Spot [13], which is the website homepage of the Japan Meteorological Agency [14], and an environmental database provided by Iowa State University [15]. Subsequently, a dataset with 165 rows and 45 columns was created as an explanatory variable. Table 2 summarizes the items in this dataset. Using real-world QAR data from a pilot report provided by FDA, we obtained Yes/No values indicating whether any FDA flights that either departed from or landed at Matsumoto airport encountered a greater than moderate (“moderate-plus”) or higher level of daily turbulence during the observation period. Three instances of moderate-plus turbulence exist in the data used in this study. We describe these data based on “location-time-altitude-type.”

Table 2
List of data used

Weather data	Time (UTC)	Altitude (hPa)	Descriptive variables	
			Name	Description
fx106 (FXJP106) [16]	0000Z	500	spd	Wind speed (knots)
	0300Z	700	shear	Wind speed difference[5] (knots/feet)
fx502 (FXFE502) [16]	0000Z	500	low	Low pressure (yes = 1, no = 0)
			trough	Trough (yes = 1, no = 0)
			alt	Number of contour lines between Wajima and Tateno
MA (Matsumoto high-rise weather)	1200Z[6]	500	spd	Wind speed (knots)
		700	hum	Humidity (%)
			dir	Wind direction (°)
			temp	Temperature (°C)
WA (Wajima high-rise weather)	1200Z[6]	500	spd	Wind speed (knots)
		700	hum	Humidity (%)
			dir	Wind direction (deg)
			temp	Temperature (°C)
TA (Tateno high-rise weather)	1200Z[6]	500	spd	Wind speed (knots)
		700	hum	Humidity (%)
			dir	Wind direction (deg)
			temp	Temperature (°C)
MMJ meteorological terminal air report (METAR)[7]	2310Z		temp	Air temperature (°C)

[5] 500 hPa only.

[6] Previous day.

[7] Refers to the weather at Matsumoto Airport. We only use the names of METAR elements; the time and altitude are not described.

[8] Not available (NA) values in this field were replaced with 10000.

Weather data	Time (UTC)	Altitude (hPa)	Descriptive variables	
			Name	Description
			dwp	Dew point (°C)
			relh	Relative humidity (%)
			dir	Wind direction (deg)
			spd	Wind speed (knots)
			alt	Altimeter (inches)
			vsby	Visibility (miles)
			gust	Wind gust (knots)
			vis1[8]	Cloud height level 1 (feet)
			vis2[8]	Cloud height level 2 (feet)
			vis3[8]	Cloud height level 3 (feet)
[5] 500 hPa only.				
[6] Previous day.				
[7] Refers to the weather at Matsumoto Airport. We only use the names of METAR elements; the time and altitude are not described.				
[8] Not available (NA) values in this field were replaced with 10000.				

Figure 4 illustrates the boxplots of fx106-03-500-spd, Wajima-12-700-temp, Matsumoto-12-500-hum, and fx106-03-500-shear; here, all data are normalized. The circle, triangle, and square in each boxplot represent the instances of turbulence in the data. On the days when turbulence was observed, the wind speed and shear were high, and the temperature was low [17].

3. Methods

3.1 Turbulence-occurrence analysis using pca

Due to a lack of sufficient data to observe patterns in annual turbulence, it is difficult to predict its occurrence through supervised learning [18]. In addition, meteorological data consists of many explanatory variables, and determining which variables contribute to turbulence is complex. Thus, in this study, we reduce the limits of the explanatory variables in the PCA and use the weights to calculate the risk clusters via the k-means method. This risk cluster is used to predict the occurrence of turbulence

through SVC. We run this program in the Python 3.7.0 environment and we also use scikit-learn version 0.23.2. The algorithm is described as follows.

1. Creation of a dataset for turbulence predictions, using open data

2. Calculation of turbulence risk cluster

   a. A projection matrix W is created via PCA [19]

   b. The data are converted to principal component (PC) vector Z using W

   c. The risk clusters are generated based on Z using the k-means method [19]

3. Prediction of turbulence using risk clusters

   a. Prediction of turbulence-occurrence dates via SVC [19, 20]

   b. Validation of predicted turbulence-occurrence dates

3.2. Dimensionality reduction and coordinate transformation in PCA

We use PCA to determine the factors that actually cause turbulence. Figure 5 depicts a plot for each observation date, with PCs 1 and 2 forming the x-and y- axes, respectively; the points indicated by arrows represent the three actual instances of turbulence. Flights with turbulence are plotted in the upper-right part of the figure. Figure 6 illustrates a scatter plot of the elements comprising the first and second PC planes. The wind speed, contour lines, and trough elements are concentrated in the upper-right quadrant of the PC1-axis; the temperature elements, in the upper-left. In other words, the farther the PC1 lies on the right-hand side, the higher the wind speed and the lower the temperature (i.e., there are many contour lines). The humidity elements are present at the top of the plot, and wind direction and cloud height elements at the bottom. This indicates that when the humidity is high, the wind direction is negative. Thus, when most of the wind is from the west, when the turbulence occurs, there is a large influence of wind from the southwest on the aircraft. Furthermore, the cloud height is low toward the top of the PC2 axis. The values of these PC loads are listed in Table 3. Figure 5 reveals that in PC4, troughs and cloud height are major factors on the days when the turbulence occurred. It is observed that wind speed difference has a significant impact on PC5.

Table 3
Value of each PC load

PC	Item	Value
PC1(26.5%)	WA-12-700-temp	-0.26
	MA-12-700-temp	-0.2518
	fx106-00-500-spd	0.2423
	fx106-03-500-spd	0.2399
	WA-12-500-temp	-0.2354
PC2(16.7%)	vis2	-0.2675
	vis3	-0.2595
	WA-12-500-dir	-0.2319
	MA-12-700-hum	0.2268
	TA-12-500-hum	0.2209
PC3(7.57%)	TA-12-700-spd	0.3143
	WA-12-700-spd	0.2914
	WA-12-500-spd	0.2871
	TA-12-700-dir	0.2266
	MA-12-700-temp	0.2262
PC4(5.42%)	spd	0.4332
	relh	-0.3829
	fx502-00-trough	0.2476
	gust	0.2439
	vis1	0.2398
PC5(4.05%)	dir	-0.4083
	fx106-00-500-shear	0.406
	fx106-03-500-shear	0.3111
	WA-12-500-hum	-0.2732
	MA-12-500-hum	-0.2324

The cumulative contribution rate from PC1 to PC2 was 43.23%, and 13 components were required to achieve a cumulative contribution of at least 80%. Therefore, we consider 13 PCs. We use the PCs of 13

items to obtain the matrix W that performs coordinate transformations based on $\boldsymbol{Z=Wx}$. Here, $\boldsymbol{x}$ is the original data and $\boldsymbol{Z}$ is the coordinate after transformation.

3.3. Calculation of risk clusters by k-means method

Using the coordinate transformation matrix obtained from the PCA described in the previous section, we calculated the risk cluster via the k-means method, wherein the $\boldsymbol{Z}$ coordinate transformed by $\boldsymbol{W}$ was used. Figure 7 shows the resulting classification into six clusters. Clusters where turbulence is expected to occur are indicated in red This risk cluster includes almost all the dates on which turbulence was observed, as given in Table 1. While Cluster ID 5 might have been affected by turbulence, it did not significantly affect flight operations. It is also likely that the other clusters were less affected by turbulence.

Figure 8 compares the risk clusters with other clusters. We find that the risk clusters have faster wind speeds, lower temperatures, lower humidity, and larger wind speed differences.

4. Result And Discussion

4.1. Turbulence prediction for validation data

We will use the risk clusters described in the previous chapter to predict the occurrence of turbulence using 179 data points that were collected from Table 2 in the year 2019. After normalizing this data, we perform axis transformation using the transformation matrix W described in the previous section.

4.2. Calculation of risk date using the risk cluster via SVC

We used the risk cluster as the training data and predicted the turbulence dates for the 2019 data using SVC. Table 4 lists the validation data and SVC parameters.

Table 4
Usage data and SVC parameters

Item	Value
Usage data year	2019
Corresponding dates	01/01–03/31, 10/01–12/31
Number of data points	179
Kernel function	Gaussian kernel
Gamma	1 / (Number of data points × Variance of data)
C	1.0

Figure 9 compares the days for which turbulence is predicted with those in which it is not. We observe that the distributions of wind speed and temperature are similar to that of the risk cluster. For fx106-03-

500-shear, all values between 10/01/2019 and 12/31/2019 equaled 0. We consider that the days with predicted turbulence have strong wind speeds, low temperatures, and large wind speed differences.

4.3. Verification of forecasted turbulence dates via SVC

Through a weather map, we verified the days with the risk of turbulence predicted using the risk clusters and SVC; the results are summarized in Table 5. We assigned turbulence risk based on four levels, categorized in increasing order of risk as 1 (normal), 2 (caution), 3 (warning), and 4 (critical), to make it easier to propose to airlines. Notably, the highest risk was observed on January 9, 2019, when flight cancellations were considered. Even on the dates whose turbulence risk levels were at least two, passenger safety, if not flight cancellation, were seriously considered. Therefore, we believe that this analysis can adequately predict turbulence-risk days.

Table 5
Verification of the predicted turbulence days using the weather map

Date	Level	Overview of weather map information
01/05/2019	3. warning	9 kt / 1000 ft shear (wind speed change per altitude) is expected from FL200[9] to FL150 over Matsumoto, and there is a high probability of mountain waves.
01/08/2019	2. caution	There is a 200 kt jet over Kyushu, and it seems that the shaking will increase over the next day, but it is possible to operate at this level.
01/09/2019	4. critical	Under these conditions, we are considering the cancellation of the flight. Moderate to severe turbulence is expected for a wide range of altitudes (FL180–FL7000); therefore, maximum caution is required. At Hanamaki Airport, four FDA flights were canceled owing to strong winds near the airport.
01/20/2019	3. warning	Two strong jet streams are approaching. Accretion and shaking are expected on the Sea of Japan side, and the wind over the mountains seems to be strong. Aircrafts can be operated, but only with extreme caution.
02/04/2019	3. warning	Shear is expected below 10000 ft and requires considerable caution. After the cold front passes, the wind becomes stronger and mountain waves are expected.

[9] FL represents the flight level, i.e., aircraft altitude at standard air pressure expressed in 100s of feet

Figure 10 compares the per-minute average of the maximum standard deviations (SDs) of the vertical sway of the aircraft [21, 22] obtained from actual QAR data against those of the predicted turbulence date calculated via SVC and the other days. We excluded 02/04/2019 because QAR data could not be obtained for that date. From the left, the graph shows the average of the maximum SDs of vertical sway, and the values concerning its climb and descent. The graph shows that the vertical sway is generally larger on the days wherein turbulence is predicted. Notably, there is considerable shaking during descent.

5. Conclusion

In this study, we used open data to predict the occurrence of turbulence to make aircraft operations safer and more comfortable. Although turbulence occurs infrequently, it is a leading cause of aircraft damage and changes in flight schedules. We derived the turbulence-risk cluster through k-means clustering after reducing the dimensions of available data via PCA, instead of using the rare instances of turbulence as the training data. Using this turbulence-risk cluster as training data, we predicted turbulence occurrences for 2019 through SVC. We confirmed that the obtained results were sufficiently accurate for utilization by pilots.

Although the prediction of turbulence occurrences was conducted for Matsumoto Airport in this study, the same method can be employed to analyze turbulence for other airports. We believe that it is possible to combine daily aircraft data with open data, such as weather data, to improve the prediction accuracy of the proposed technique. We hope that the proposed method will not only help predict turbulence but also lead to increased systems expertise and technological advancements in combating turbulence, to compensate for future human resource shortages in aviation.

Abbreviations

PCA Principal component analysis

FDA Fuji Dream Airlines

SVC Support vector classification

QAR Quick access recorder

SD Standard deviation

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

All data generated or analyzed during this study are included in this published article.

Competing interests

The authors declare that they have no competing interests.

Funding

Not applicable

Author contributions

Ito, Mizuno conceptualized the study, Mizuno designed the study, Mizuno, Ohba, Ito were responsible for data acquisition and analysis, Mizuno, Ohba were responsible for data interpretation, Mizuno, Ohba contributed new methods or models or software, Mizuno was responsible for writing draft and revision.

ACKNOWLEDGMENTS

We would like to thank Fuji Dream Airlines Co., Ltd. for providing information useful for this study.

References

1. Digest of Aircraft Accident Analyses for Prevention of Accidents due to the Shaking of the Aircraft. In: JTSB Digest Vol. 15. Japan Transport Safety Board. 2015. https://www.mlit.go.jp/jtsb/bunseki-kankoubutu/jtsbdigests_e/jtsbdigests_No15/No15_pdf/jtsbdi-15_all.pdf. Accessed 7 Dec 2020.
2. Cabinet Office in Japan: White Paper on Traffic Safety in Japan 2018. https://www8.cao.go.jp/koutu/taisaku/h30kou_haku/english/wp2018-pdf.html. Accessed 7 Dec 2020.
3. Regular Airlines Association: Movements towards alternative aviation fuel use in the aviation industry. https://www.meti.go.jp/committee/kenkyukai/energy_environment/biojet/pdf/001_02_00.pdf. Accessed 12 Sept 2018
4. Japan Aerospace Exploration Agency: Smart flight technology. https://www.aero.jaxa.jp/eng/research/star/smart-flight/ Accessed 29 Aug 2018.
5. Japan Aerospace Exploration Agency: Demonstration of turbulence prevention airframe technology (SafeAvio). https://www.aero.jaxa.jp/eng/research/star/safeavio/news170313.html. Accessed 7 Dec 2020.
6. Oikawa H, Inokuchi H, Izumi K, Kikuchi Y, and Hayasaki N. Relation between resolution enhancement and accuracy in prediction of turbulence area which perturbs aircraft. Tenki. 2010; 57;9:669-680. (In Japanese).
7. Aviation Today: Delta develops artificial intelligence tool to address weather disruption, improve flight operations. https://www.aviationtoday.com/2020/01/08/delta-develops-ai-tool-address-weather-disruption-improve-flight-operations/. Accessed 7 Dec 2020.
8. Muñoz-Esparza D, Sharman RD, Deierling W. Aviation turbulence forecasting at upper levels with machine learning techniques based on regression trees. J Appl Meteorol Clim. 2020;59:1883-1899.

9. International Aircraft Development Fund: Trends in flight data analysis technology (FDM/FOQA). http://www.iadf.or.jp/document/pdf/24-7.pdf. Accessed 24 Dec 2020. (In Japanese).
10. Dubois P, AIRBUS - Airlines SMS & FDA Assistance: Flight Data Analysis. https://www.icao.int/SAM/Documents/2015-FDA/1.3%20AIRBUS%20PDubois%20FDA-Seminar_1%20Recording.pdf. Accessed 24 Dec 2020.
11. Mizuno S, Ohba H, Ito K. Analysis of critical factors of aircraft operations' efficiency. J Jpn Indl Manag Assoc. 2020;70:4E:211-221.
12. Carney TQ, Bedard Jr AJ, Brown JM, McGinley J, Lindholm T, Kraus MJ. Hazardous mountain winds and their visual indicators. US Department of Commerce, National Oceanic and Atmospheric Administration, 1995;55.
13. Sunny Spot Inc.: Weather and climate information site. https://www.sunny-spot.net/chart/chart_archive.html. Accessed 10 Dec 2020.
14. Japan Meteorological Agency: Historical weather data retrieval (high rise), https://www.data.jma.go.jp/obd/stats/etrn/upper/index.php. Accessed 10 Dec 2020.
15. Iowa State University. Iowa Environmental Mesonet: DownloadASOS/AWOS/METAR Data. https://mesonet.agron.iastate.edu/request/download.phtml. Accessed 10 Dec 2020.
16. Japan Meteorological Agency: Aviation Forecast. https://www.data.jma.go.jp/add/suishin/cgi-bin/catalogue/make_product_page.cgi?id=KokuYoho. Accessed 24 Dec 2020. (In Japanese).
17. Volcanic Ash Advisory Center: Characteristics of high-degree disturbances near Japan. Tenki. 1967;14:5:179-187. (In Japanese).
18. Fukui K. Machine learning learned with Python and examples: Identification / Prediction / Abnormality Detection. Ohmsha 2018. (In Japanese).
19. Yang XS. Optimization techniques and applications with examples. Wiley; 2018.
20. Bishop, CM. Pattern recognition and machine learning. Springer; 2006.
21. Sato M, Endo E. Spectrum analysis for turbulence and induced accelerations. J Jpn Soc Aeronaut Space Sci. 2008;56 Supp 653:293-295. (In Japanese).
22. Prince JB, Buck BK, Robinson PA, Ryan T. In-service evaluation of the turbulence auto-PIREP system and enhanced turbulence radar technologies. NASA, CR-2007-214887. 2007.

Figures

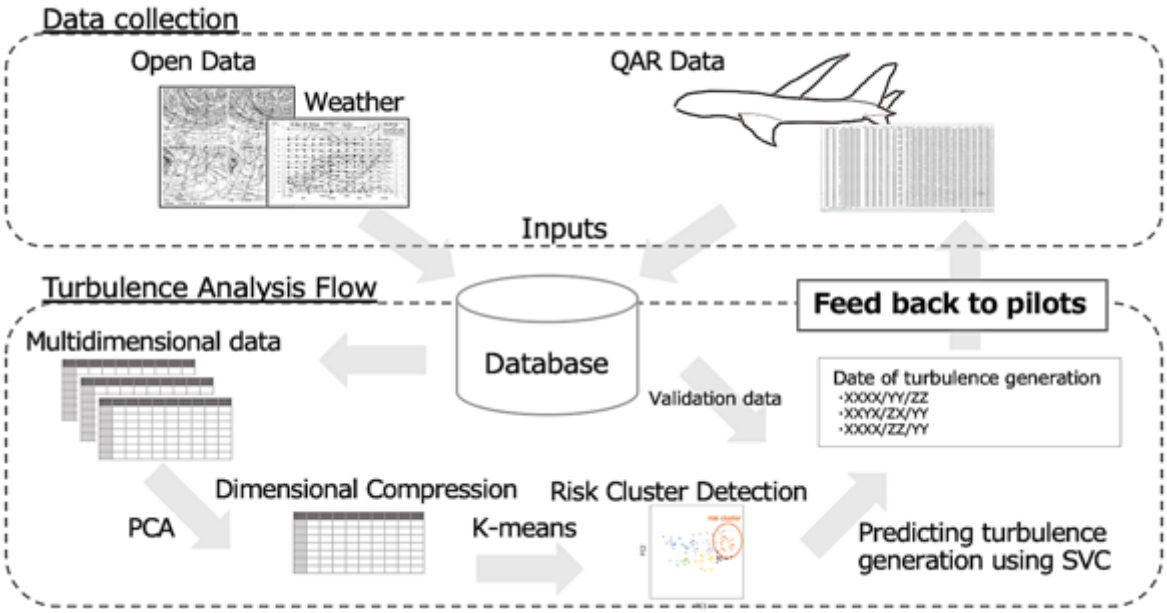

Figure 1

Framework of the proposed method

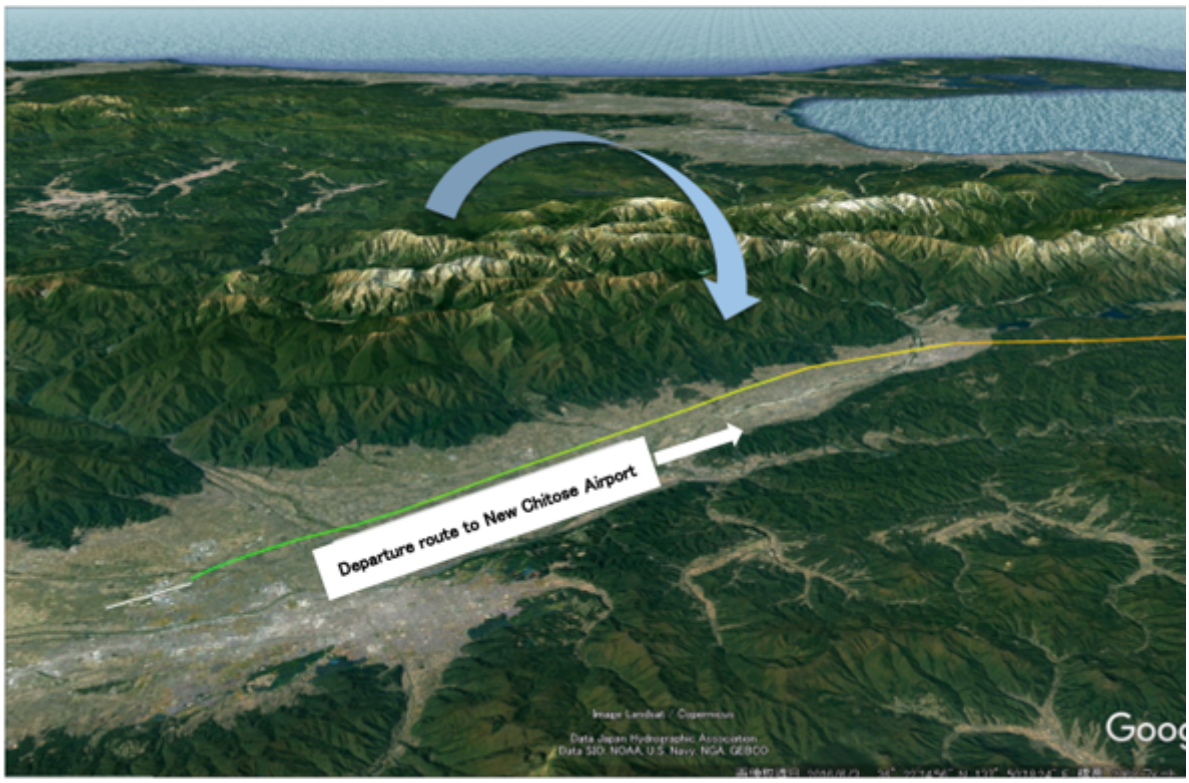

Figure 2

Impact of mountain waves on flights departing from Matsumoto Airport

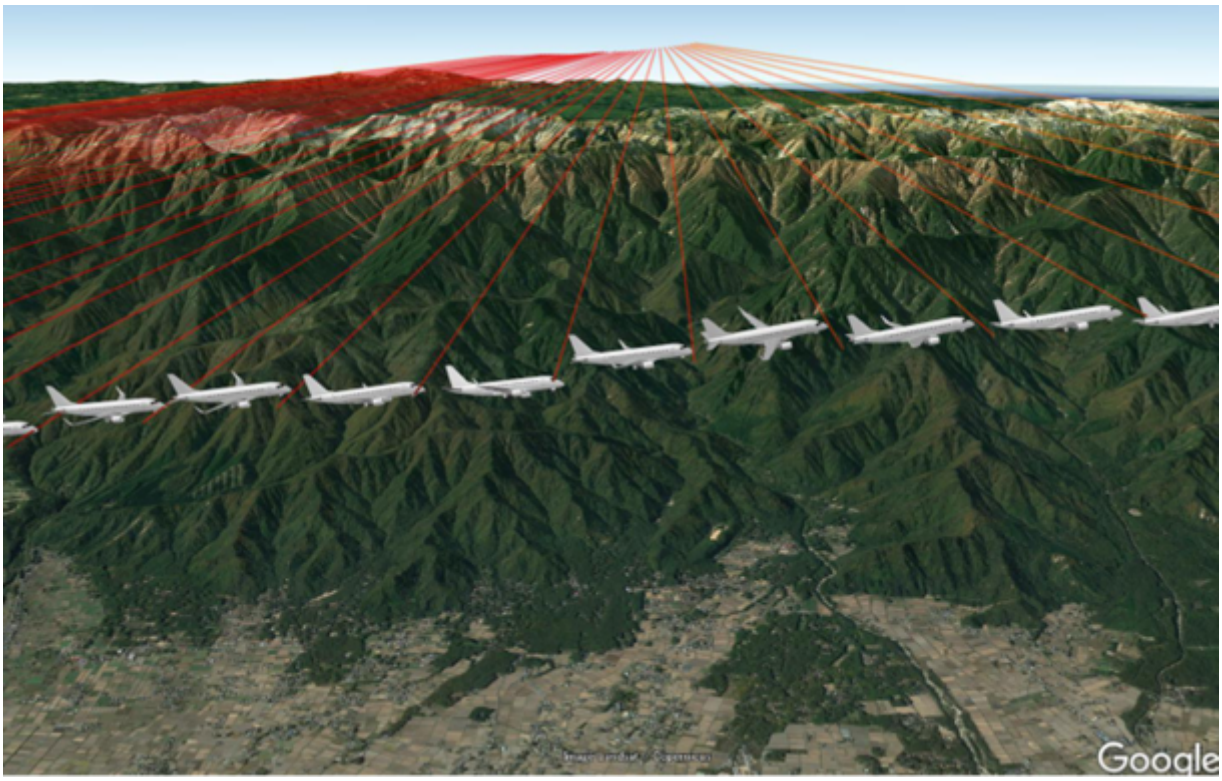

Figure 3

Visualization of altitude changes owing to turbulence

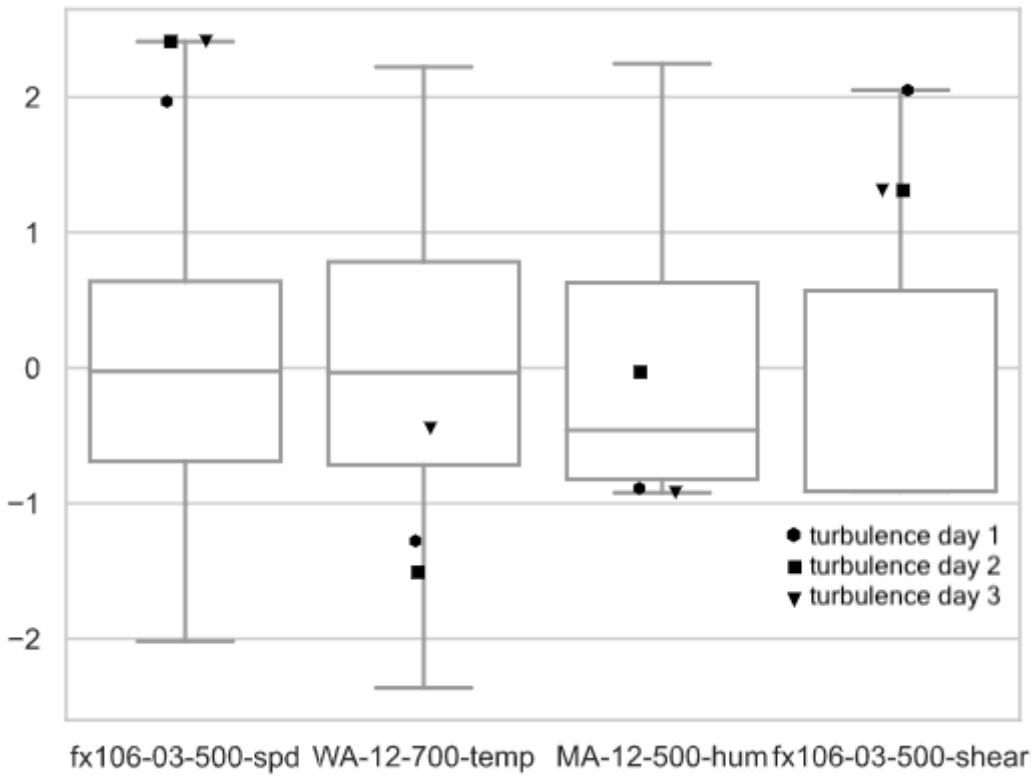

Figure 4

Dataset comparison by wind speed, temperature, humidity, and shear difference

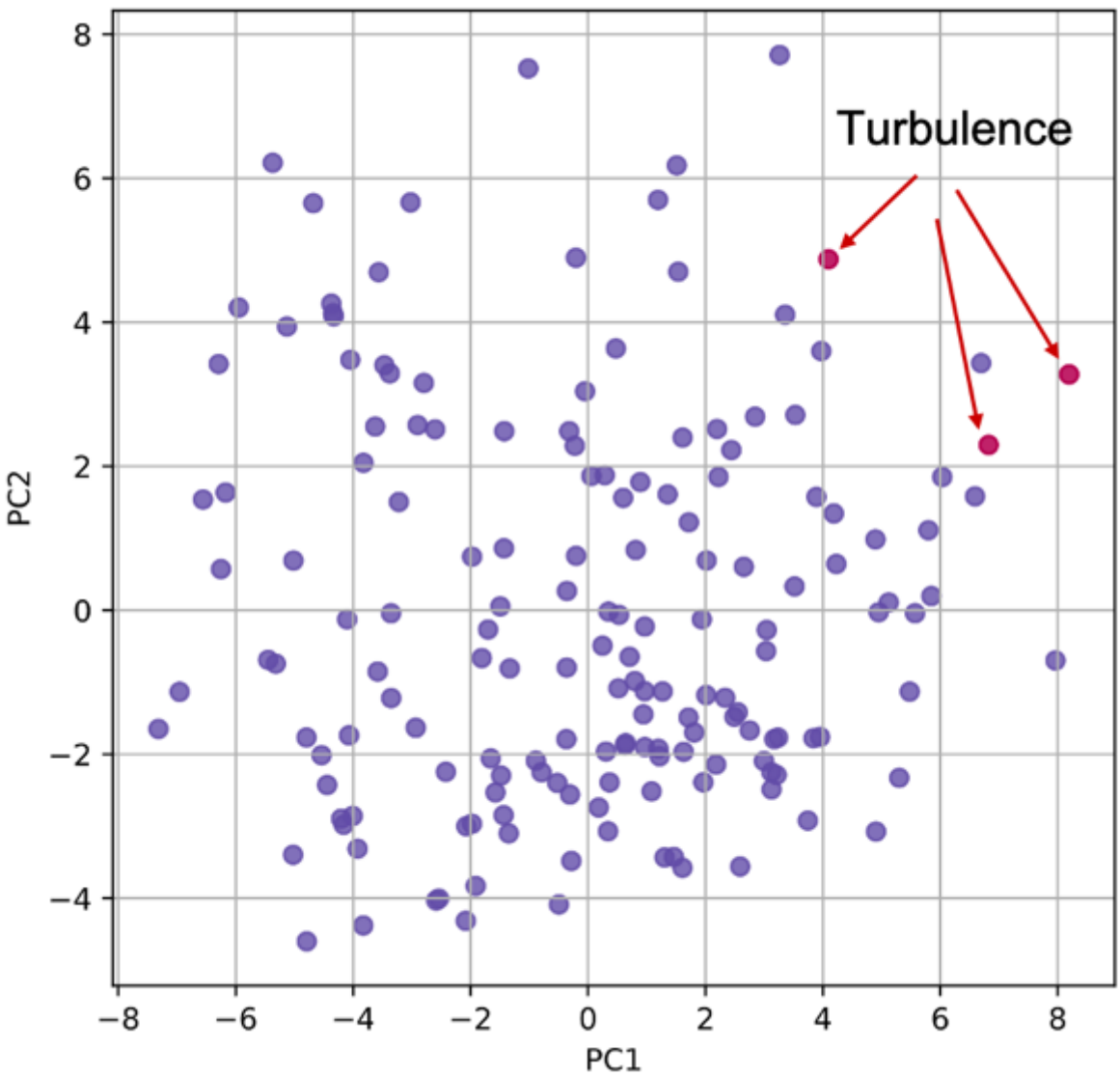

Figure 5

Scatter plot of PC1 and PC2

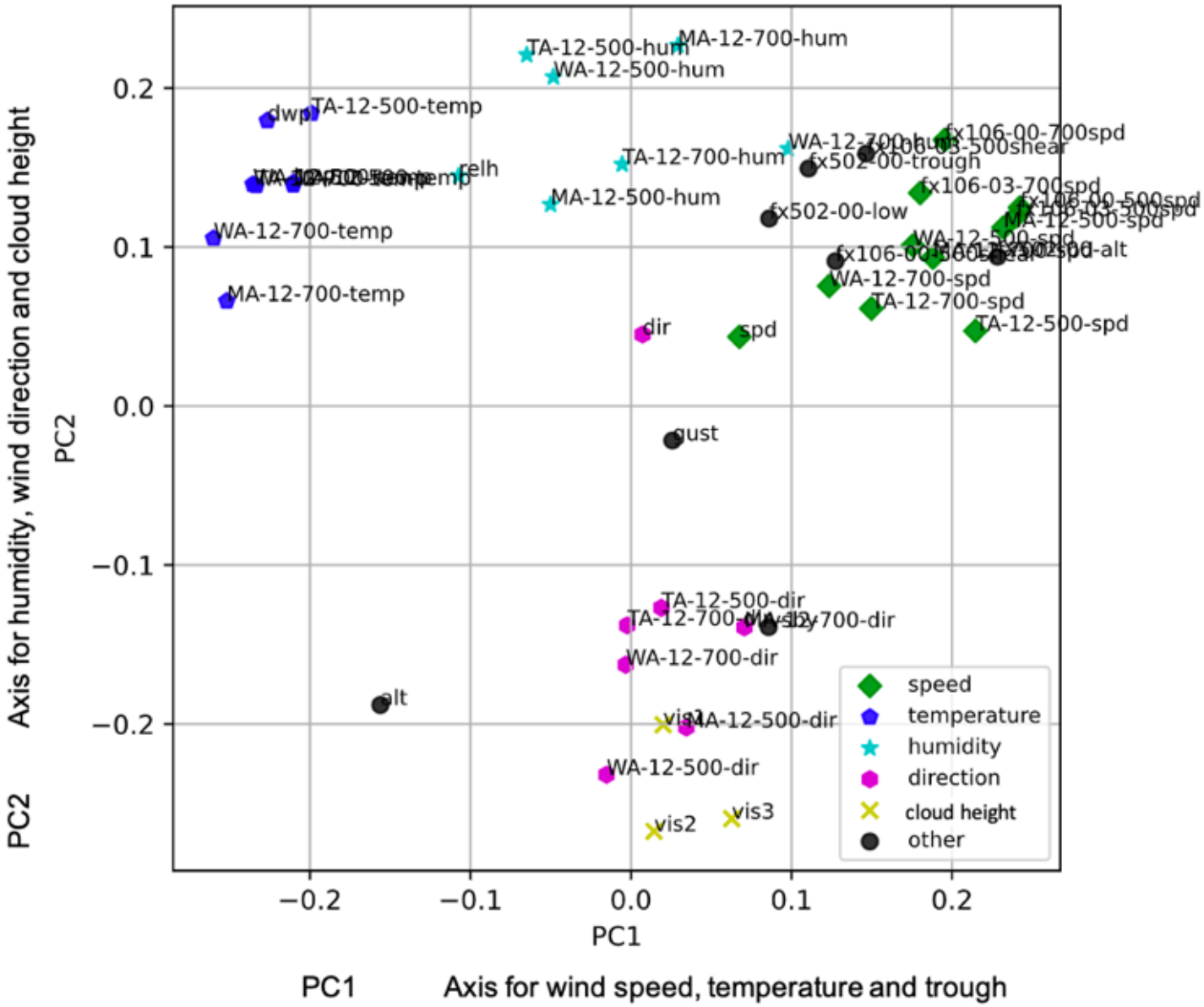

Figure 6

Plot of each element on PC1 and PC2 planes

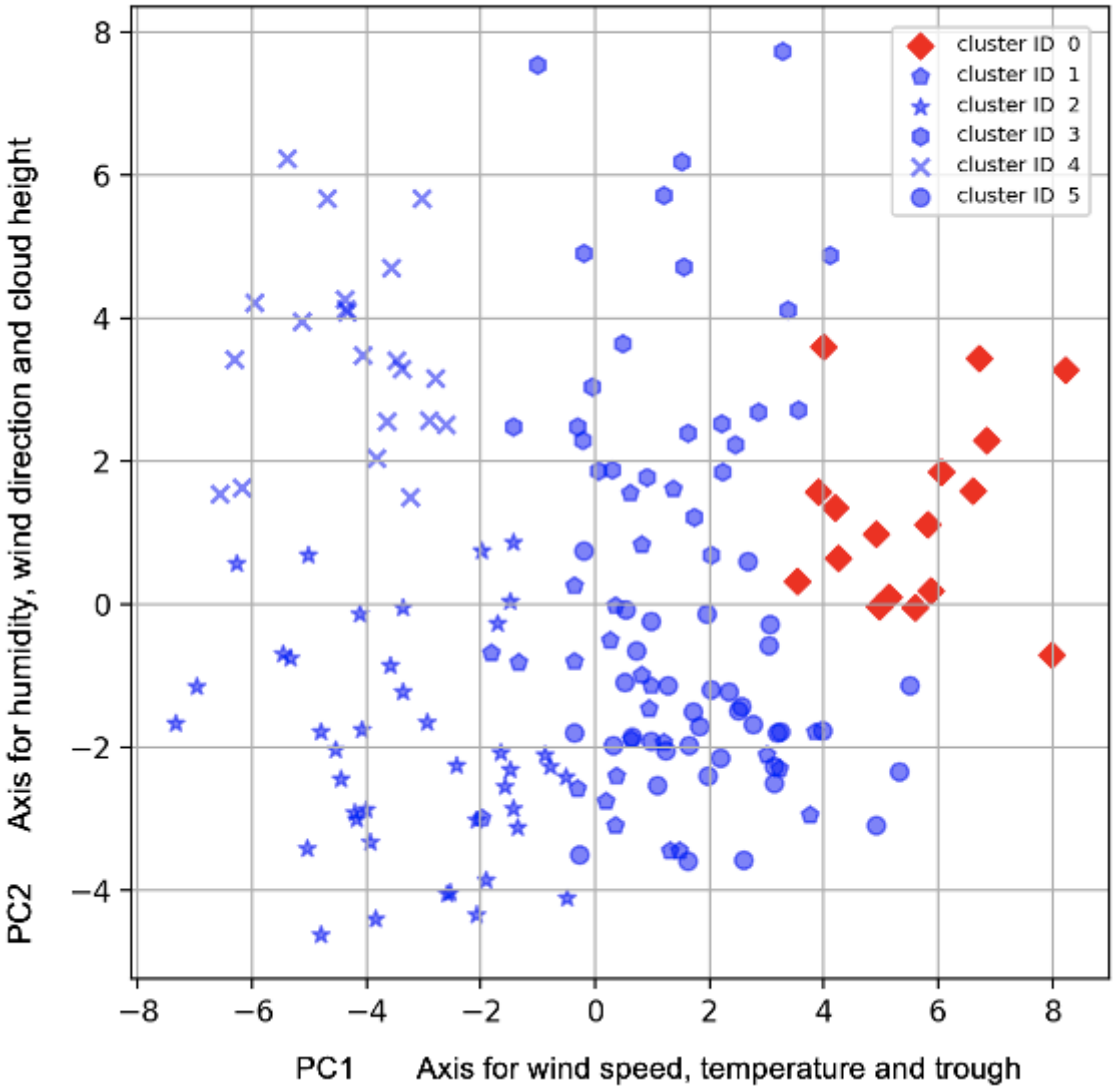

Figure 7

Six clusters obtained via the k-means method

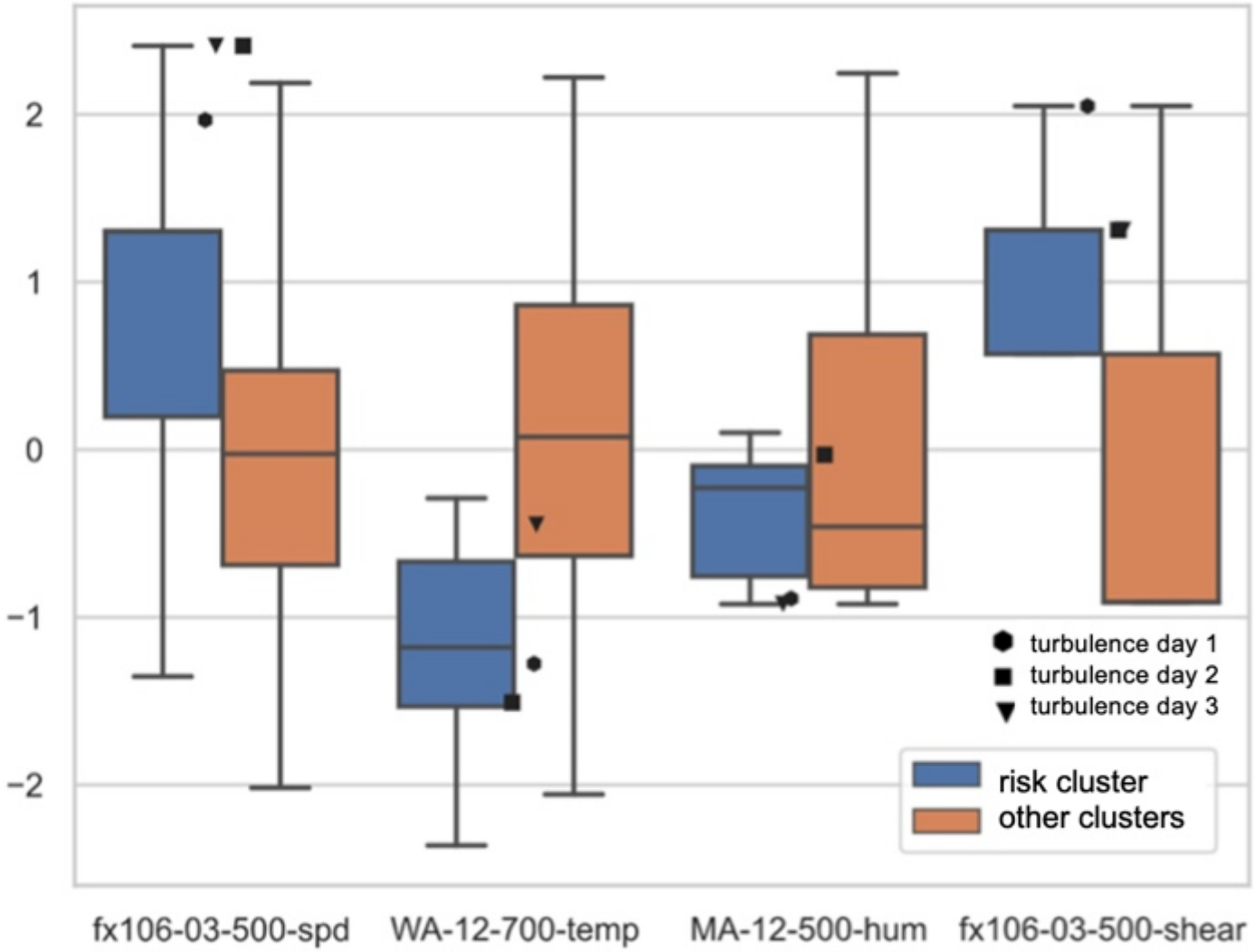

Figure 8

Comparison of risk cluster and other clusters

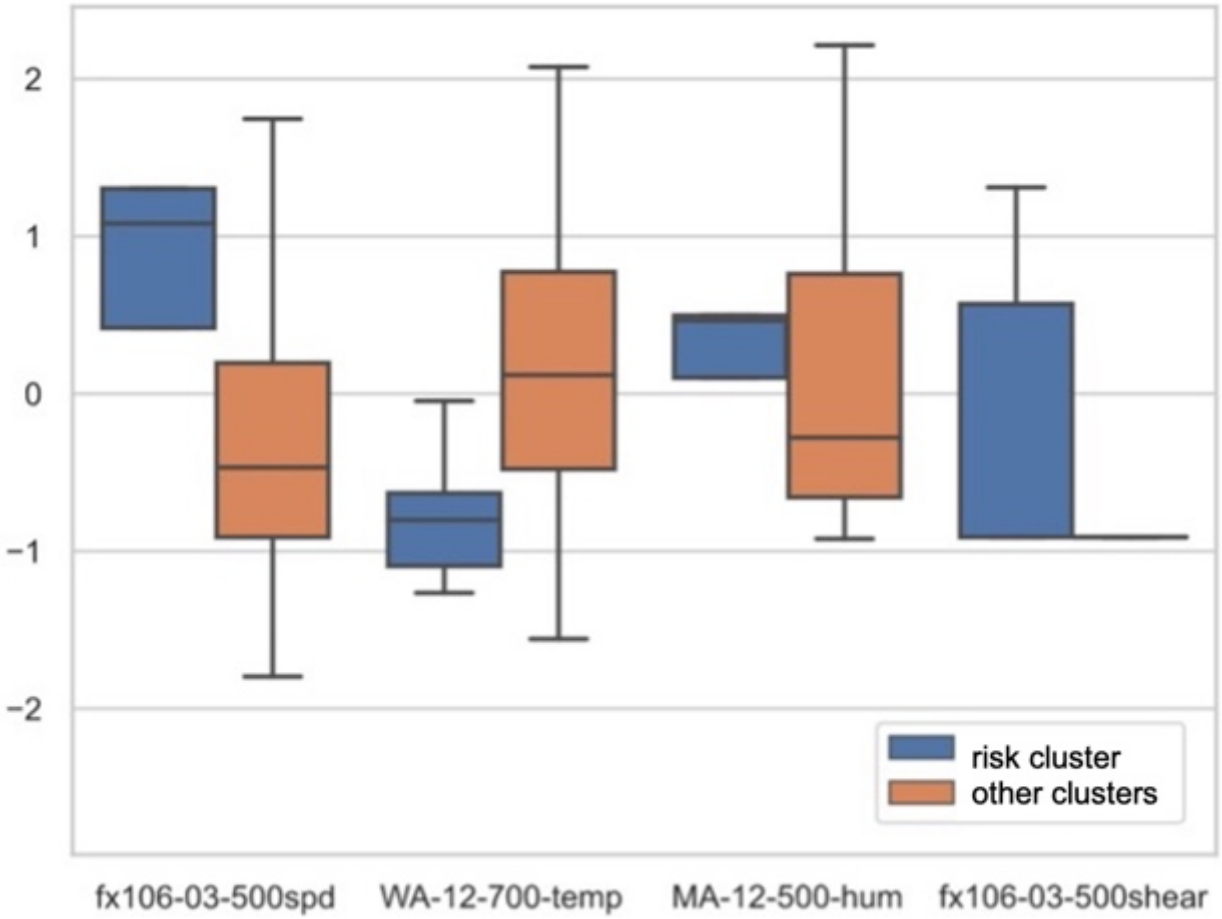

Figure 9

Comparison of days with and without turbulence predicted for the 2019 data

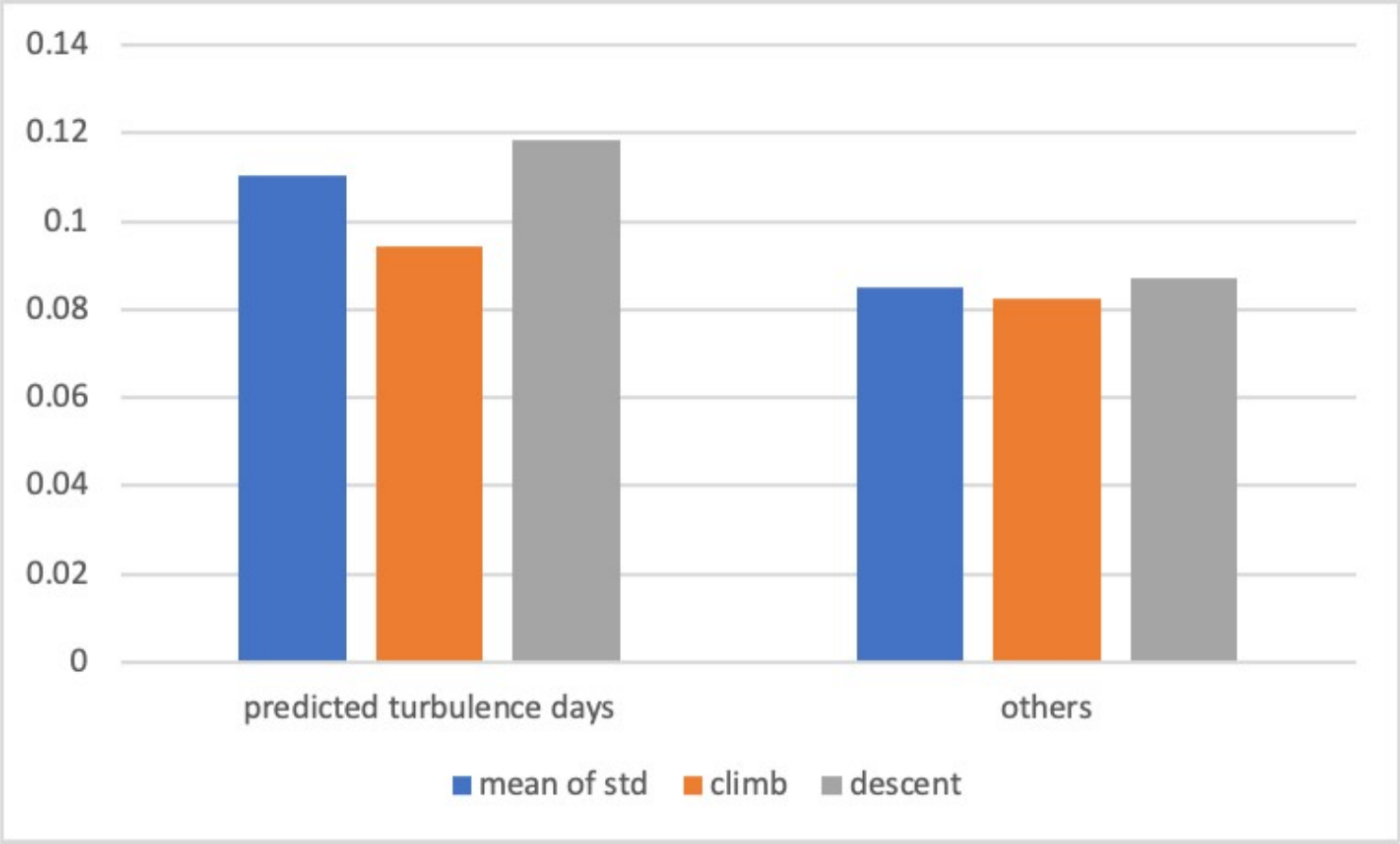

Figure 10

Comparison of the average of the maximum SDs of vertical sway for days with and without predicted turbulence from QAR data