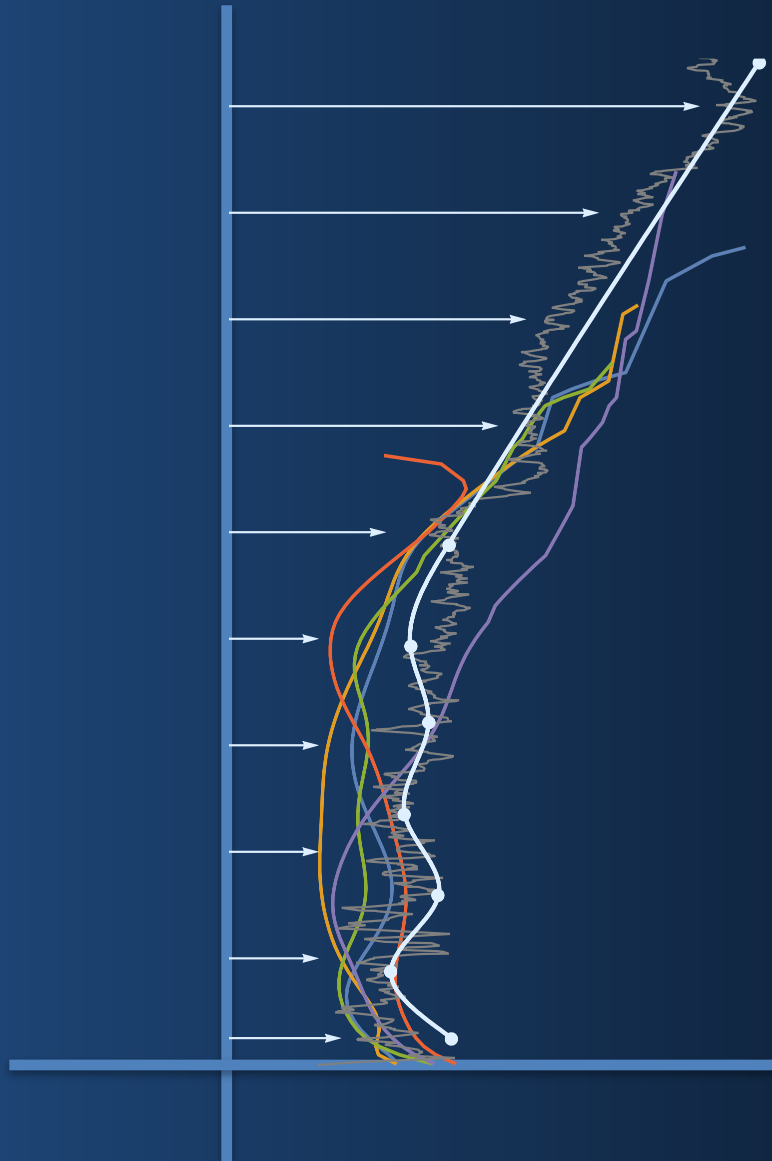


Developing a WiPPR Contact Classifier

Marshall Bradley and Mark Henderson
May 2020

LogLinear Group

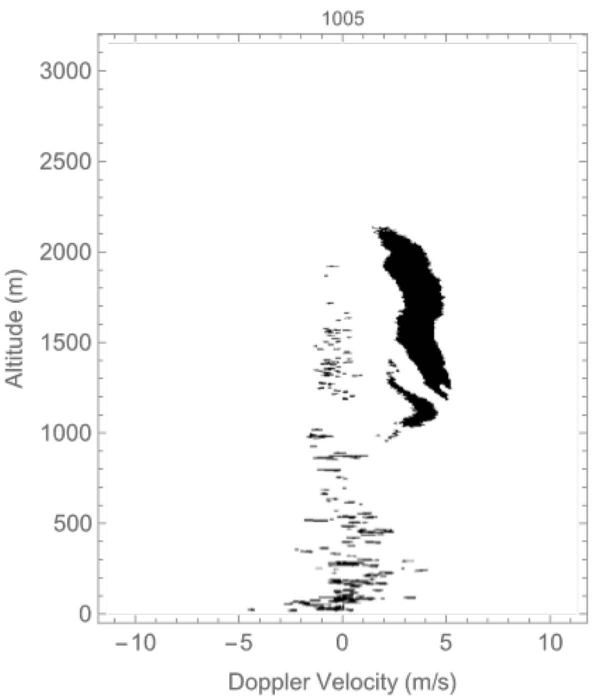
Background

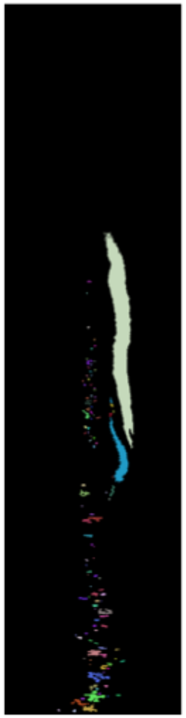
- The Wind Profiling Portable Radar (WiPPR) measures wind velocity profiles in the convective boundary layer (0-2000 m altitude range).
- In order for WiPPR to function in an automated fashion falling objects must be identified and flagged for special processing. This requires some type of artificial intelligence, specifically a classifier.
- In temperate climates the WiPPR occasionally detects virga, falling hydrometers that evaporate before they reach the ground. Virga is not spatially homogeneous so it is not detected on each of the four radar beams at the same time. This makes it difficult to estimate wind speed since clear air scatter is moving horizontally while virga is moving both horizontally and vertically. Virga always has higher Doppler velocity than clear air scatter even though they are being blown along at the same horizontal wind velocity.
- If the falling object is wide spread rain or snow and it is seen on all four beams then there is not a problem. The processing software will correctly estimate the fall velocity as well as the horizontal wind velocity. Falling objects detected only on one or two beams will lead to an incorrect estimate of wind velocity.
- This presentation briefly describes a classifier that has been developed to distinguish between virga and clear air scatter.
- This classifier represents a key step forward in developing a WiPPR system that can operate without a man in the loop.

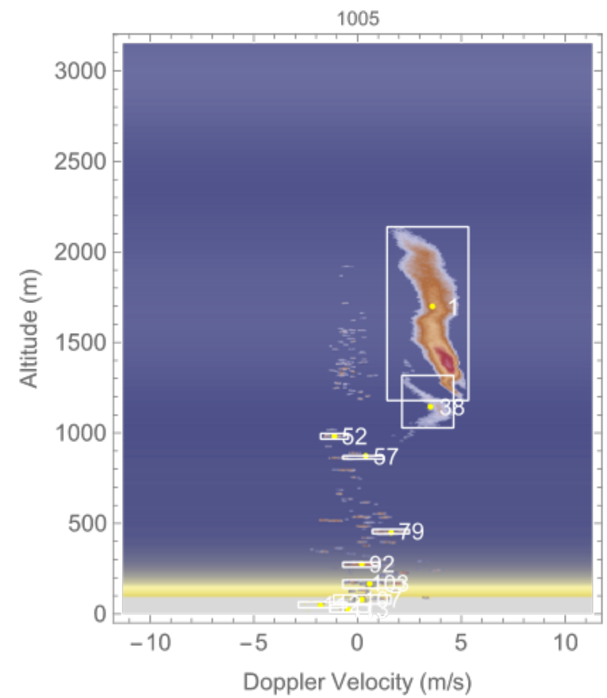
Morphology background

- Morphology is the study of the shape and characteristics of pixel connected regions in an image.
- The morphological properties of virga differ from clear air scatter in at least four ways. 1) Virga has higher pixel count. 2) Virga has higher peak echo. 3) Virga contacts have smoother edges. 4) Virga contacts have smaller aspect ratio. Aspect ratio is the width of the bounding box divided by its height.
- The selection of these four key features and finding an algorithm that rapidly determines edge smoothness turned out to be the key step in the development of this classifier.
- The classifier is based upon data collected with the WiPPR system in June 2017 at YPG and August 2017 at the Stennis airport near Kiln MS.

Example of WiPPR data with two virga contacts





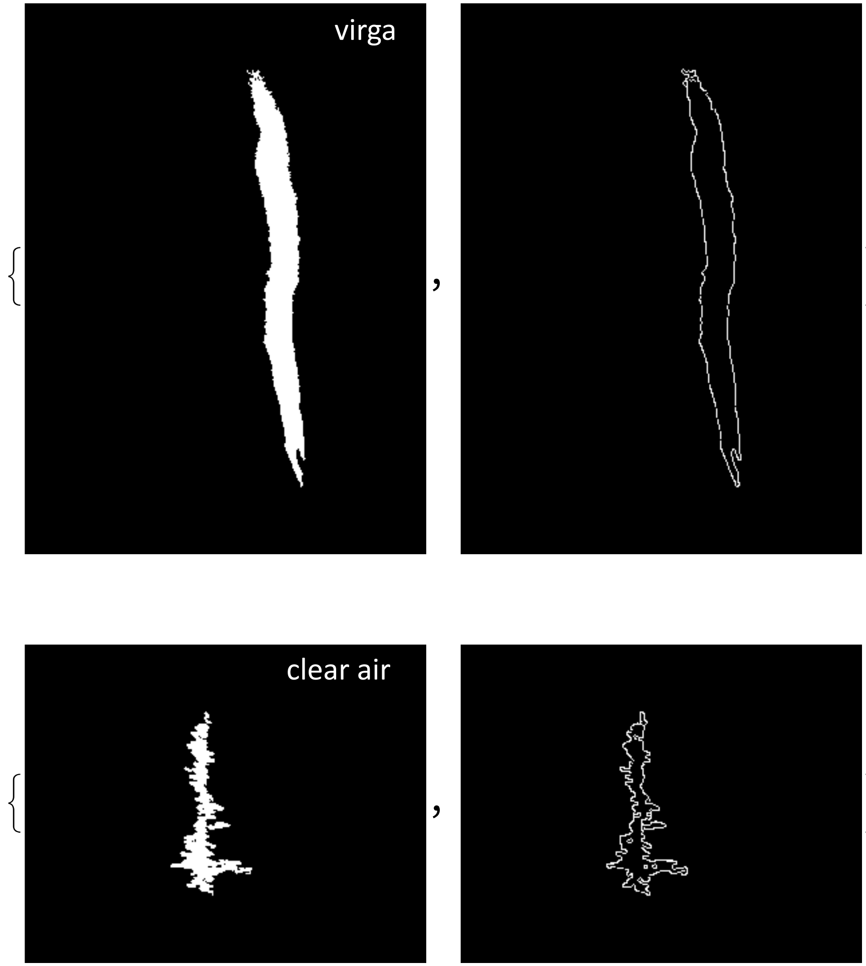


Morphological features 1 and 38 in the above figure to the right are virga. Everything else is clear air scatter. Bounding boxes are drawn in white. Relative to clear air scatter, virga has 1) high pixel count, 2) high peak echo, 3) smooth edges and 4) small aspect raio (bounding box width divided by height). Figure generated in" *Developing a GWPPR Contact Classifier Prototyp*e.nb".

Key objective

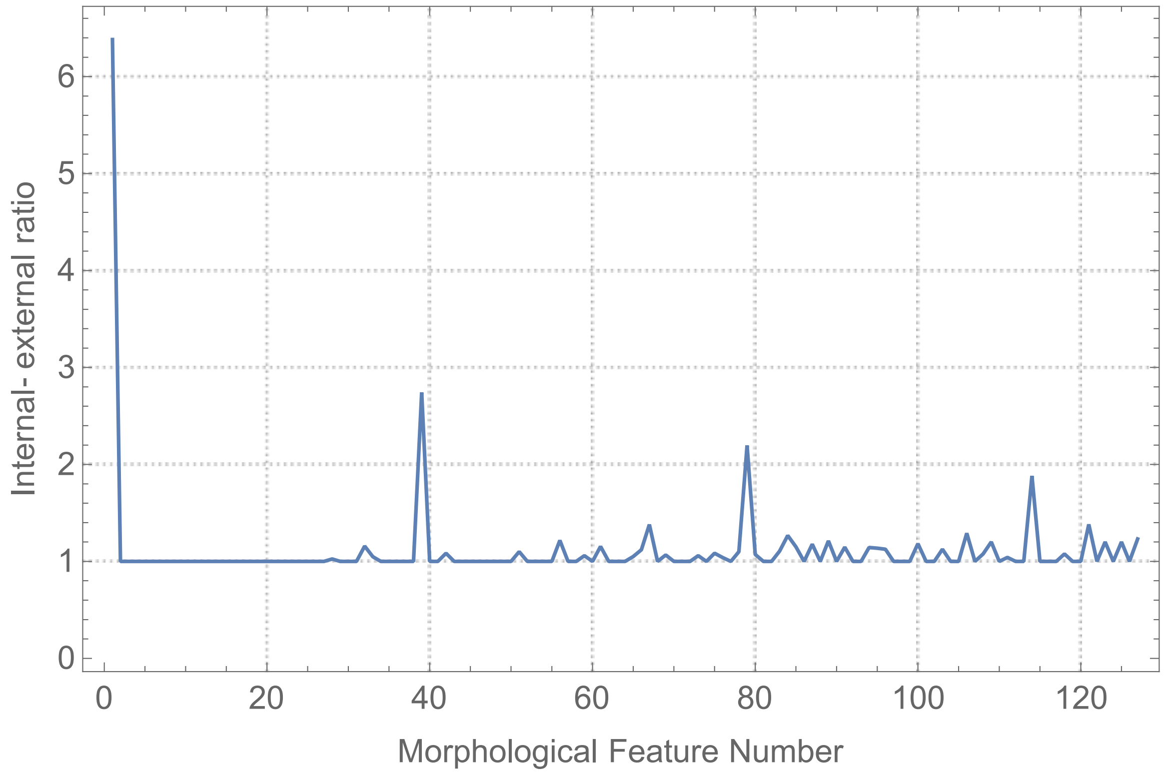
- The next slide shows a WiPPR range velocity matrix containing two high pixel count virga contacts, one high pixel count clear air scatter contact as well as numerous low pixel count clear air contacts.
- The thing that we are trying to do is decide which contacts arise from virga and which come from clear air scatter.
- The key feature that distinguishes high pixel count virga contacts from high pixel count clear air contacts is the roughness of the edges. Virga has smooth edges and the clear air scatter (actually caused by turbulence) has rough edges.
- In the next slide features 1 and 39 are virga and feature 79 is high count clear air scatter. There are also numerous other low pixel-count clear air scatter. Each of these has a morphological component number as well. These have not been shown in the images for clarity's sake.
- Morphological features numbers are nothing more than labels. They convey no physical meaning. We could just as well have used a, b, c,…. We use numbers simply because they are easier to work with.

Smooth and rough borders



interior exterior

Internal-external ratio



The proper way to characterize the roughness of a border is to compute the Hausdorff fractal exponent of the border. Rougher borders have bigger Hausdorff exponents. This is an expensive process and does not generalize well to the many small features from clear air scatter in WiPPR images. A cheap alternative is to compute the internal-external ratio. This is the ratio of total pixels in a feature to to the perimeter length of the feature measured in pixels. In the above example morphological features 1 and 39 are virga. Features 79 and 144 are high count clear air scatter. All other features are also clear air scatter. The virga has a higher external-internal ratio than does the clear air scatter. The idea for using the external-internal ratio came from reading the paper “The fractal dimension of quality habitats” by Irme and Bogaert.

Construction of training and test data sets

- The construction of classifiers requires training and test data.
- The classifier presented here is based upon data collected with the WiPPR system on 26 June 2017 at YPG and 24 August 2017 at the Stennis airport near Kiln MS.
- On 26 June there was no virga observed so it is very easy to obtain clear air training and test data for this day.
- On 24 August about 8% of the data files had virga contacts. These files were identified by hand.
- Identifying the morphological feature number of each of the virga contacts in the 24 August files was the most labor intensive portion of the classifier development presented here. A great deal of special purpose graphics software had to be developed to insure the correctness of the virga training and test data.
- The ratio of clear air to virga data in the training and test sets is about 100 to 1. It is quite likely that this is not optimal.

Characteristics of training and test data

The training virga data set contains 39 contacts.

The test virga data set contains 28 contacts.

The training clear air data set contains 4915 contacts.

The test clear air data set contains 4770 contacts.

Each contact is characterized by 4 parameters.

4 parameters generate 6 possible combinations.

The only good way to view multidimensional data is to look at the marginals.

Example virga training data

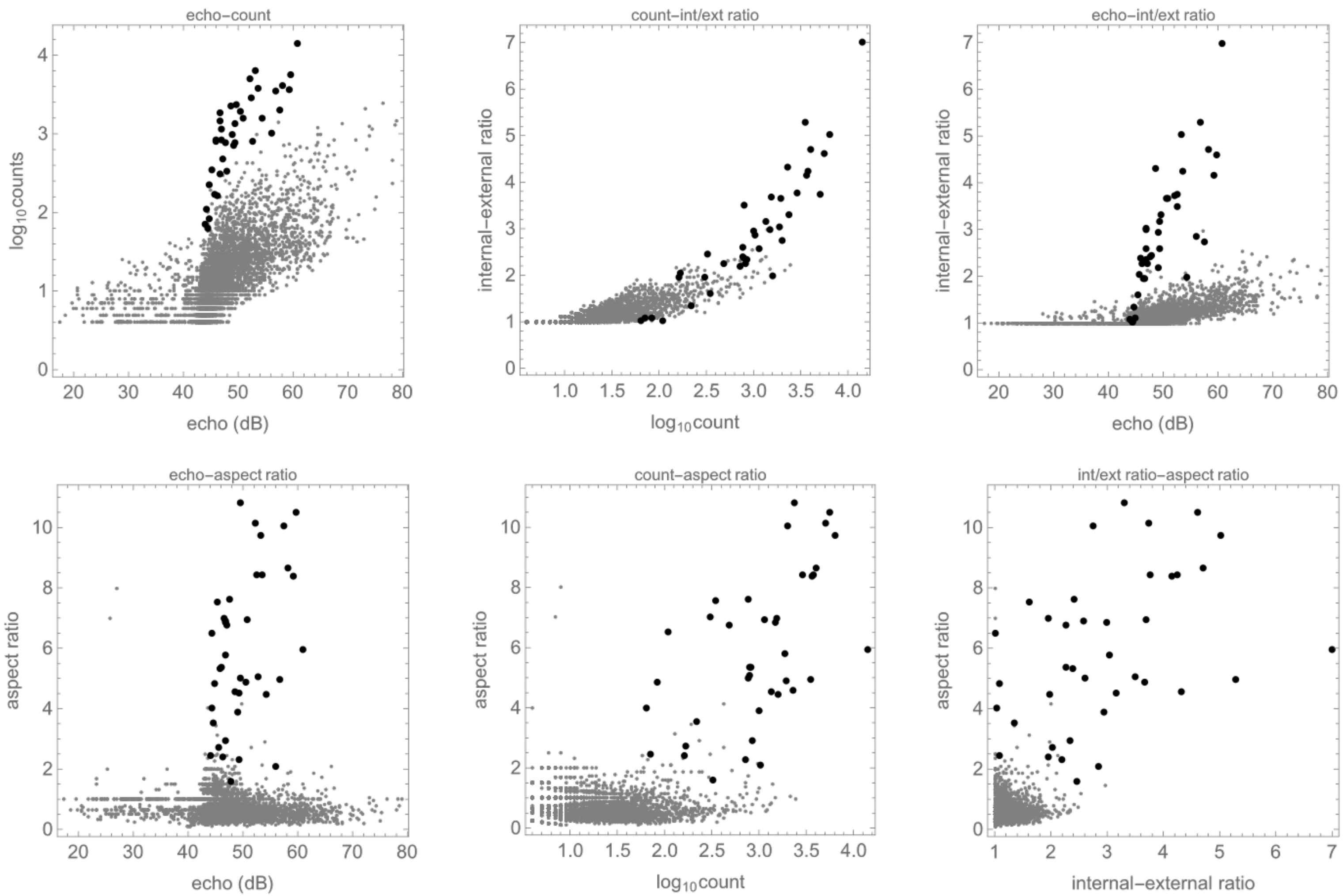
count	max echo(dB)	int/ext ratio	aspect ratio
486	47.083	2.26047	6.75
222	44.6651	1.34545	3.52941
5054	52.158	3.73817	10.1515
14 376	60.8113	6.99562	5.95604
1882	46.7228	3.03548	5.78571
3769	53.5455	4.24437	8.41667
796	45.8693	2.37612	5.33333
1360	49.3578	3.16279	4.52
6351	53.1631	5.02453	9.72093
163	46.2281	1.96386	2.41667

Example clear air training data

count	max echo(dB)	int/ext ratio	aspect ratio
24	45.8903	1.04348	0.555556
15	44.0863	1.	0.666667
5	43.2839	1.	2.
10	44.1115	1.	0.8
7	43.9337	1.	0.75
8	44.3181	1.	0.75
14	45.9135	1.07692	0.8
9	44.6532	1.	1.5
6	43.6432	1.	1.
4	42.9191	1.	1.5

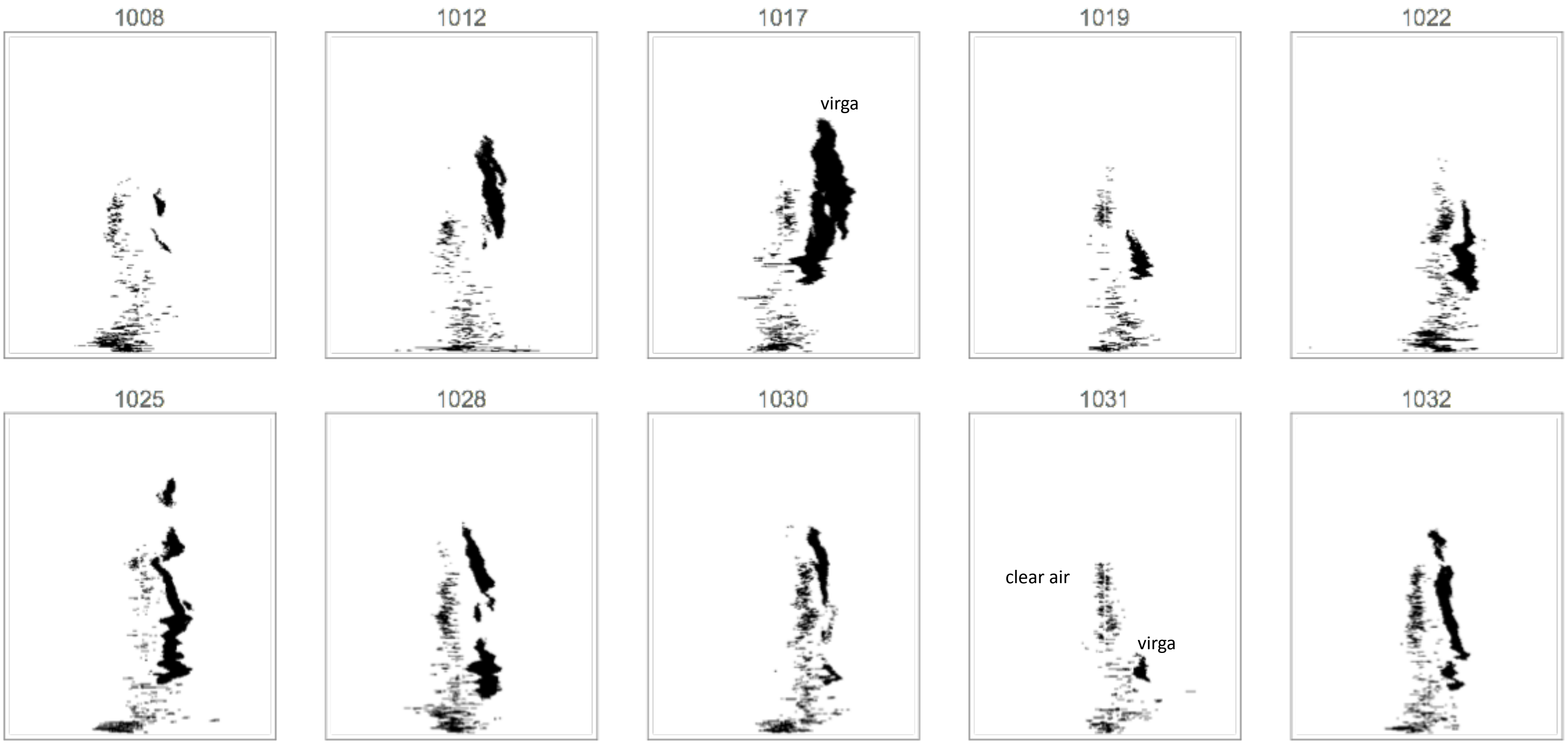
Six marginal distributions of the training data

LogLinear Group



Small gray points are clear air scatter. Large black points are virga. The classifier works because the gray and black points are well separated from one another This is the key step. It required physical not mathematical insight.

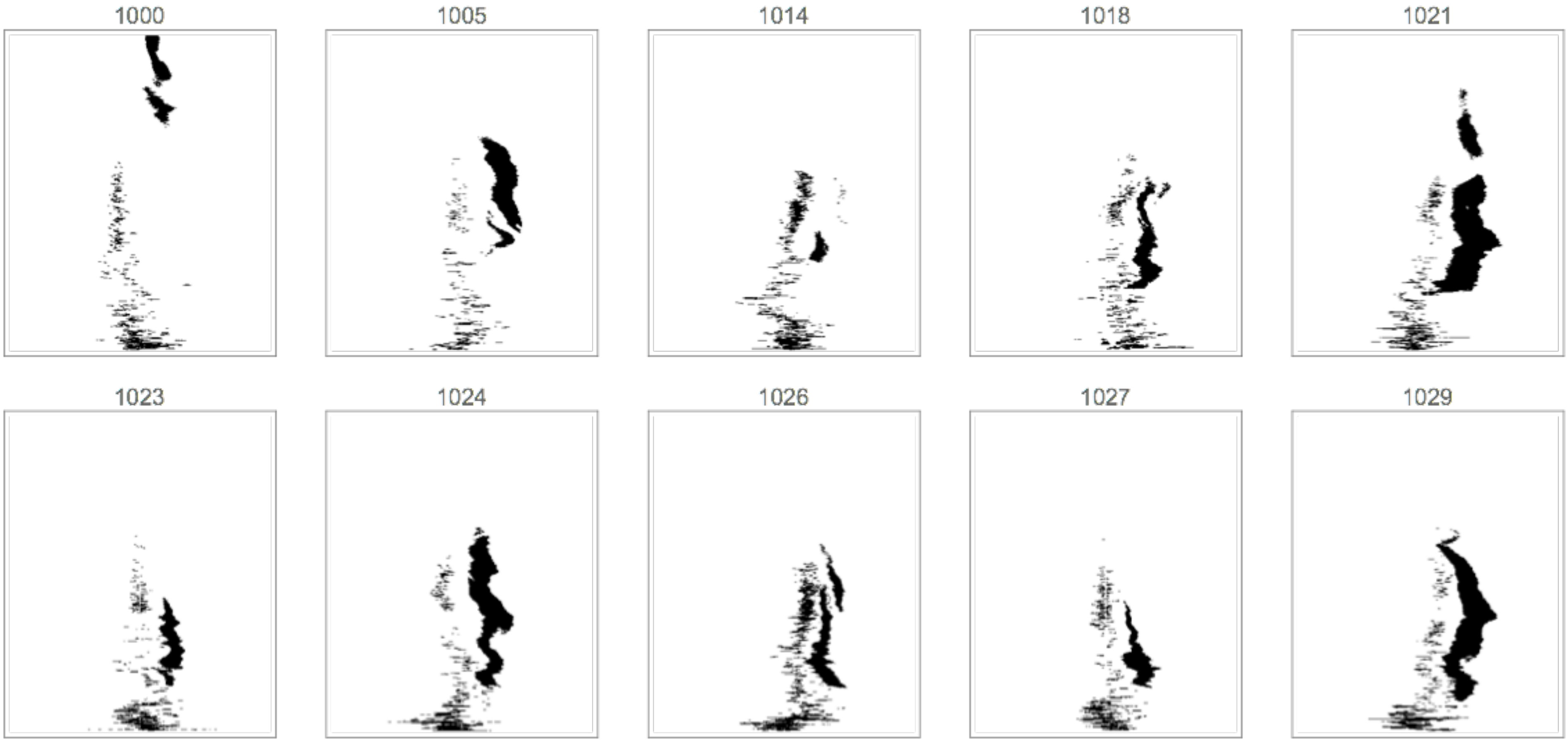
First 10 files used in making virga training data



Horizontal axes are Doppler velocity. Vertical axes are altitude. Virga is the black blobs usually to the right of the clear air scatter.

First 10 files used in making virga test data

LogLinear Group



Choice of classifier

- Once the training and test datasets were developed, the Mathematica *Classify* function was use to compute a classifier function. Reference: https://blog.wolfram.com/2017/10/10/building-the-automated-data-scientist-the-new-classify-and-predict/
- The *Classify* function chose the “LogisticRegression” option for our data. The method that the function uses to make this choice is not documented.
- Other options available to the *Classify* function include “Markov”, “RandomForrest”, “NeuralNetwork”,SupportVectorMachine” , “RandomForrest” and “NearestNeighbors”.
- We decided to stay with the “LogisticRegression” option since our focus was on prototype development and not classifier type optimization. We also understood how logistic regression works at a theoretical level.
- In a post analysis not shown here we found that the “NearestNeighbors” option performed very well and may have been a better choice. The performance of the “NeuralNetwork” option was terrible. This is surprising since logistic regression is a one node neural network.

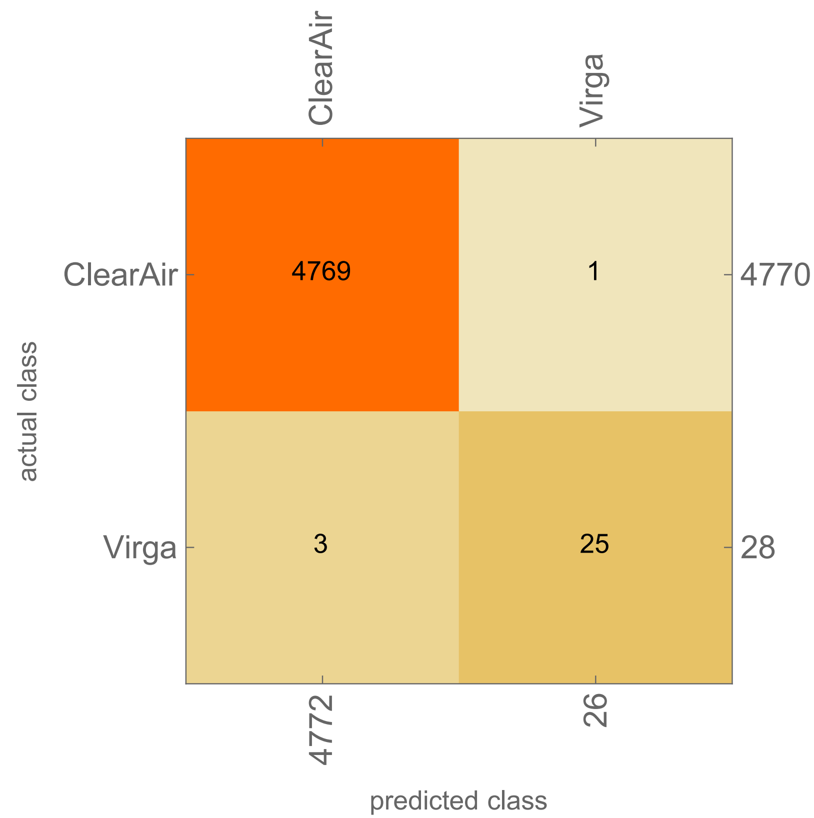
Properties of the logistic regression classifier

LogLinear Group

Classifier information

Input type	Mixed (number: 4)
Classes	ClearAir, Virga
Method	LogisticRegression
Accuracy	$100.\% \pm 1.3\times10^{-14}\%$
Loss	$3.33\times10^{-614} \pm 2.6\times10^{-308}$
Evaluation time	4.1 μs/example
Classifier memory	325. kB
Training examples used	4954 examples
Training time	11.8 s

Confusion matrix



The classifier was trained using 4954 observations of which 4915 were clear air and 39 were virga.

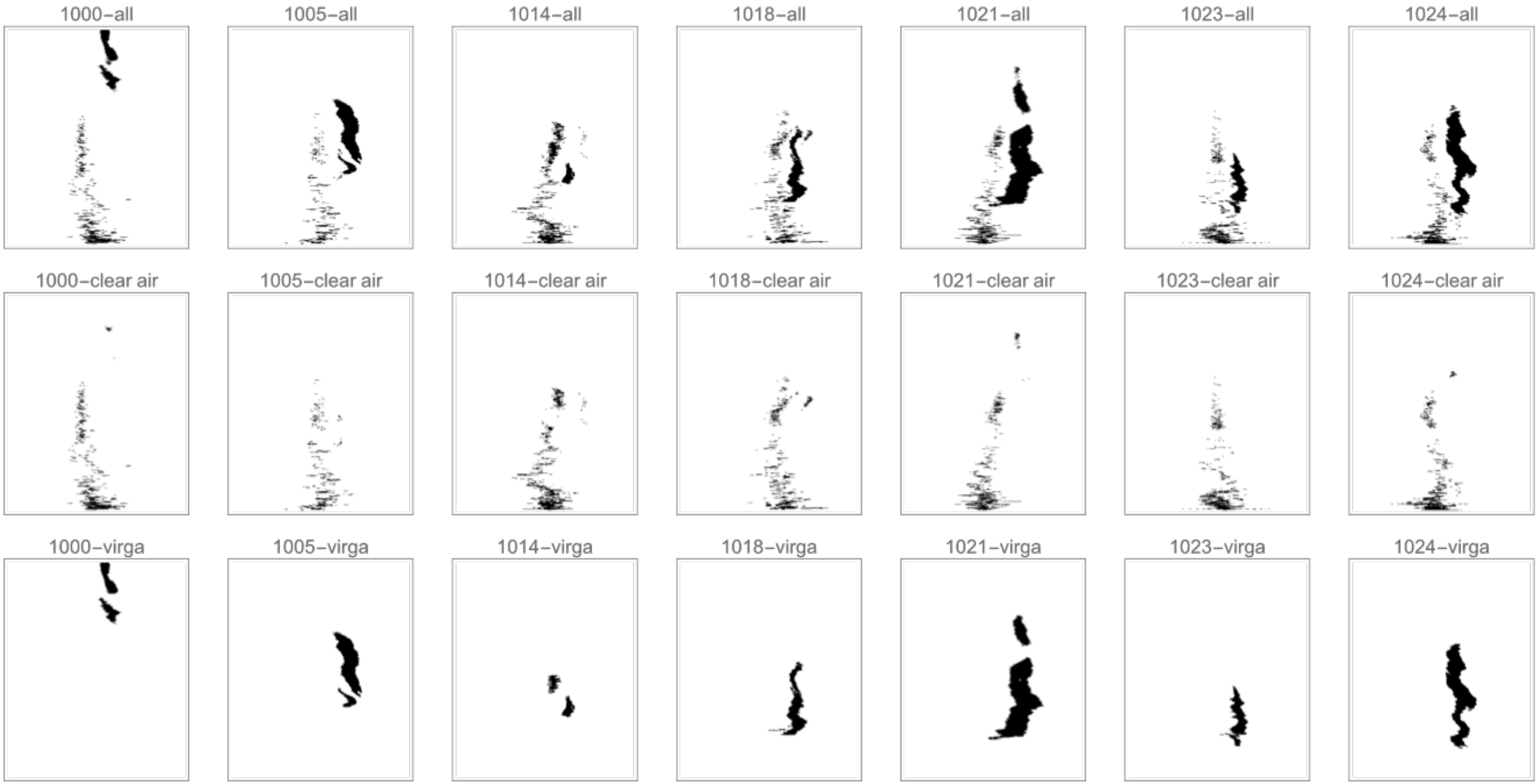
The classifier was tested using 4798 observations of which 4772 were clear air and 28 were virga.

Logistic classifier application

- The next three slides illustrate the performance of the logistic classifier.
- The first row depicts the measured WiPPR data in a compressed graphical form. The horizontal axis is Doppler velocity and the vertical axis is altitude.
- The second row shows those contacts determined to be clear air by the classifier.
- The third row shows those contacts determined to be virga by the classifier.

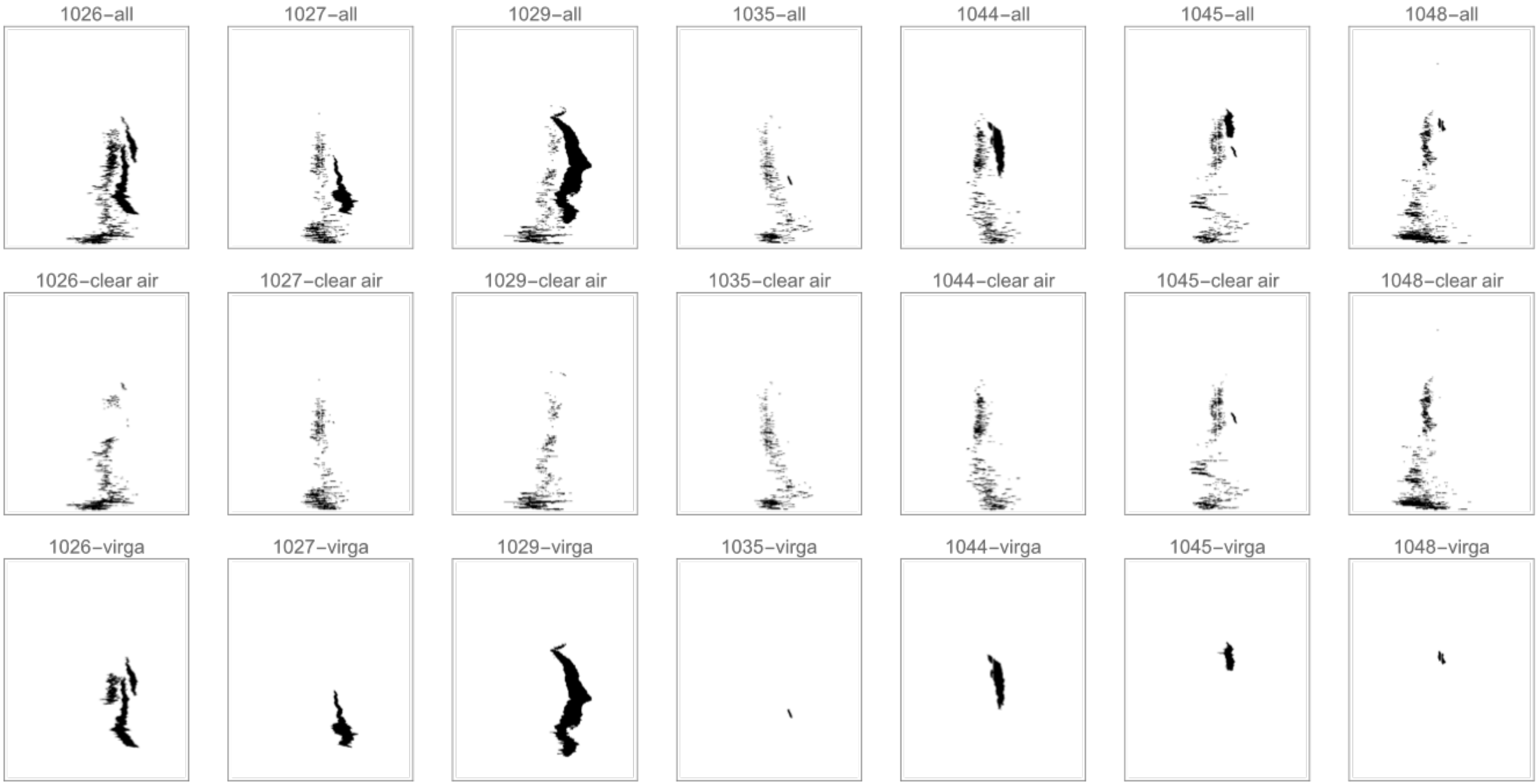
Classifier application- first 7 training files containing virga

LogLinear Group



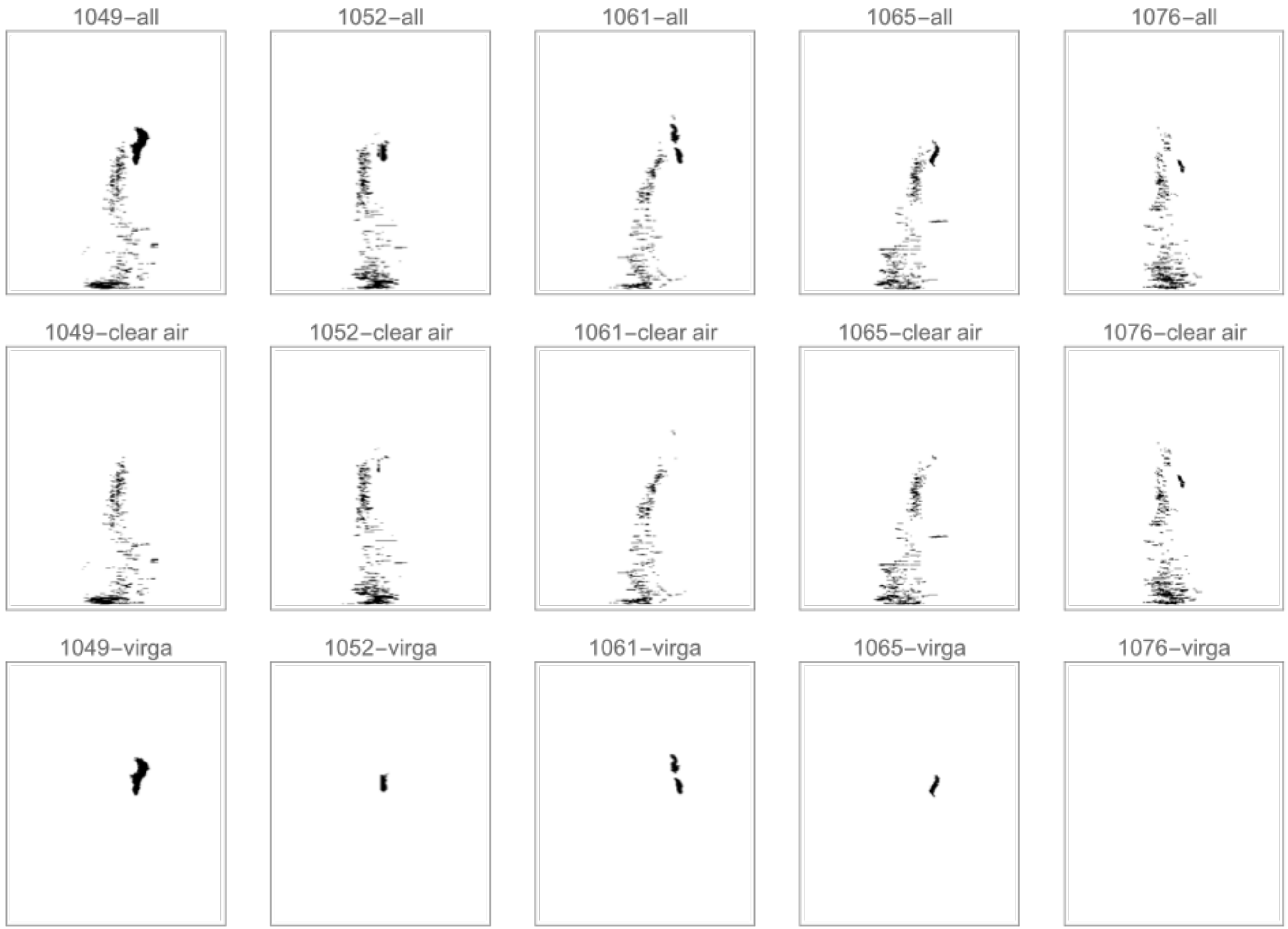
Classifier application- second 7 training files containing virga

LogLinear Group



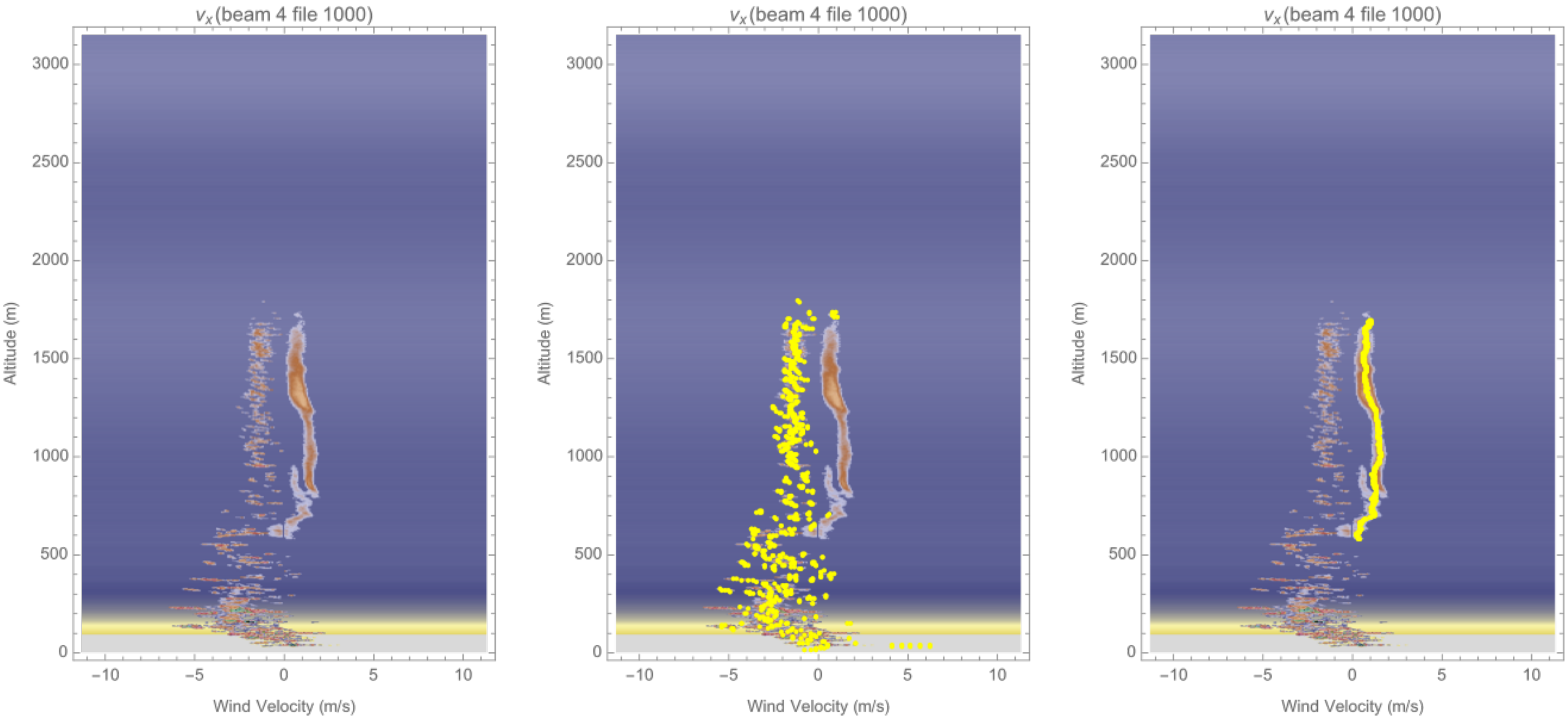
Classifier application- last 5 training files containing virga

LogLinear Group



Velocity estimation with classifier

LogLinear Group



Left) WiPPR range velocity matrix. Center) Yellow points indicated velocity estimates of contacts determined to be clear air. Right) Yellow indicates velocity estimates of contacts determined to be virga. Figure generated in "*GWPPR Contact Catalog with Classification.nb*".

Improved WiPPR catalog format

LogLinear Group

First 10 lines in a large catalog developed for improved WiPPR applications.

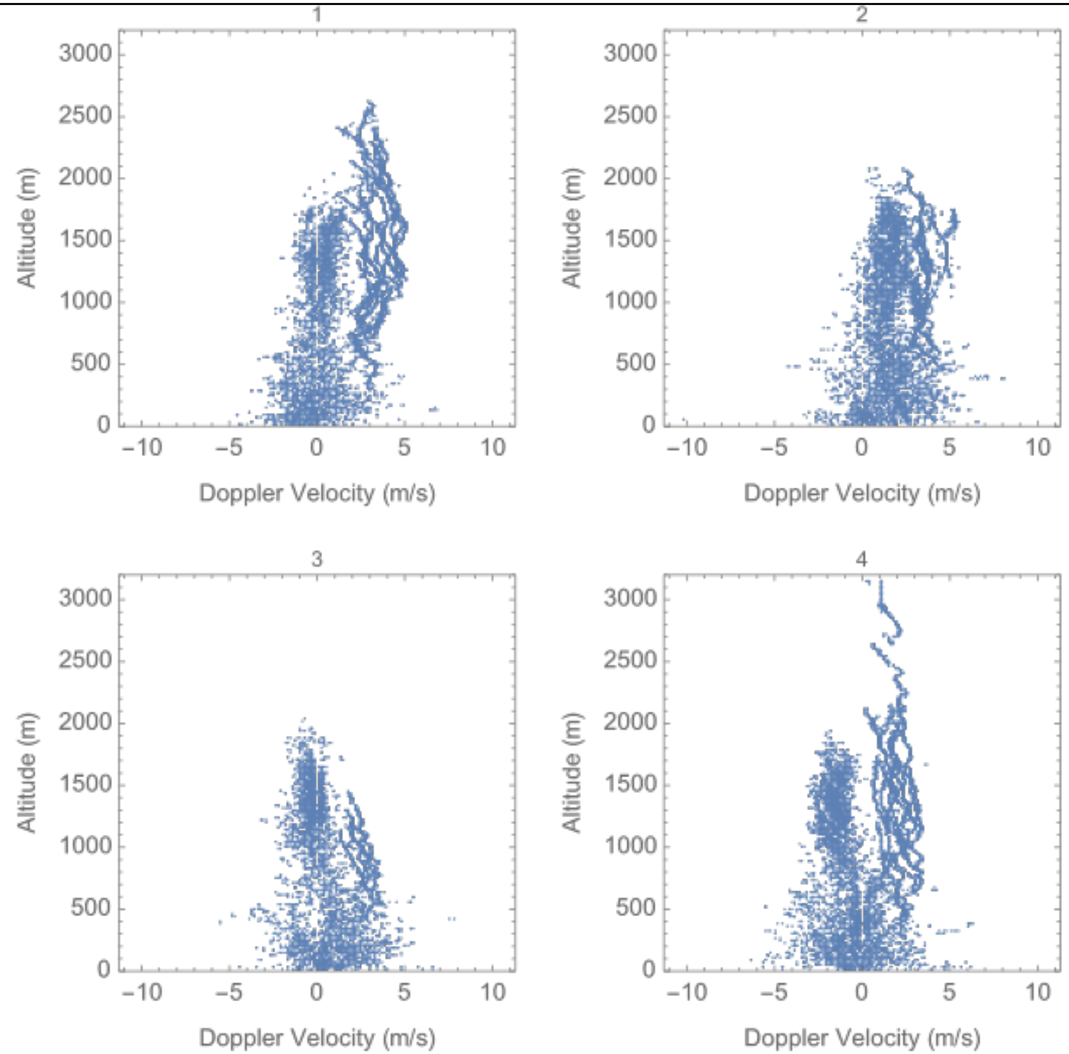
Developed for AWiPPR | Developed and added for WiPPR 2020

z(m)	nfile	Vpdf(m/s)	cosX	cosY	cosZ	beam	cmpnt#	count	max echo (dB)	int/ext ratio	aspect ratio	class
15.3876	221	-0.566199	0	0.173648	0.984808	1	39	32	27.5149	1.18519	0.230769	ClearAir
18.4651	221	-0.635167	0	0.173648	0.984808	1	39	32	27.5149	1.18519	0.230769	ClearAir
21.5427	221	-0.707139	0	0.173648	0.984808	1	39	32	27.5149	1.18519	0.230769	ClearAir
101.558	221	-1.2472	0	0.173648	0.984808	1	38	40	66.7333	1.48148	0.454545	ClearAir
104.636	221	-1.26152	0	0.173648	0.984808	1	38	40	66.7333	1.48148	0.454545	ClearAir
107.713	221	-1.27411	0	0.173648	0.984808	1	38	40	66.7333	1.48148	0.454545	ClearAir
110.791	221	-1.44794	0	0.173648	0.984808	1	38	40	66.7333	1.48148	0.454545	ClearAir
113.868	221	-1.50461	0	0.173648	0.984808	1	38	40	66.7333	1.48148	0.454545	ClearAir
116.946	221	-0.0946928	0	0.173648	0.984808	1	37	14	45.5993	1.	0.5	ClearAir
120.023	221	-0.0976502	0	0.173648	0.984808	1	37	14	45.5993	1.	0.5	ClearAir

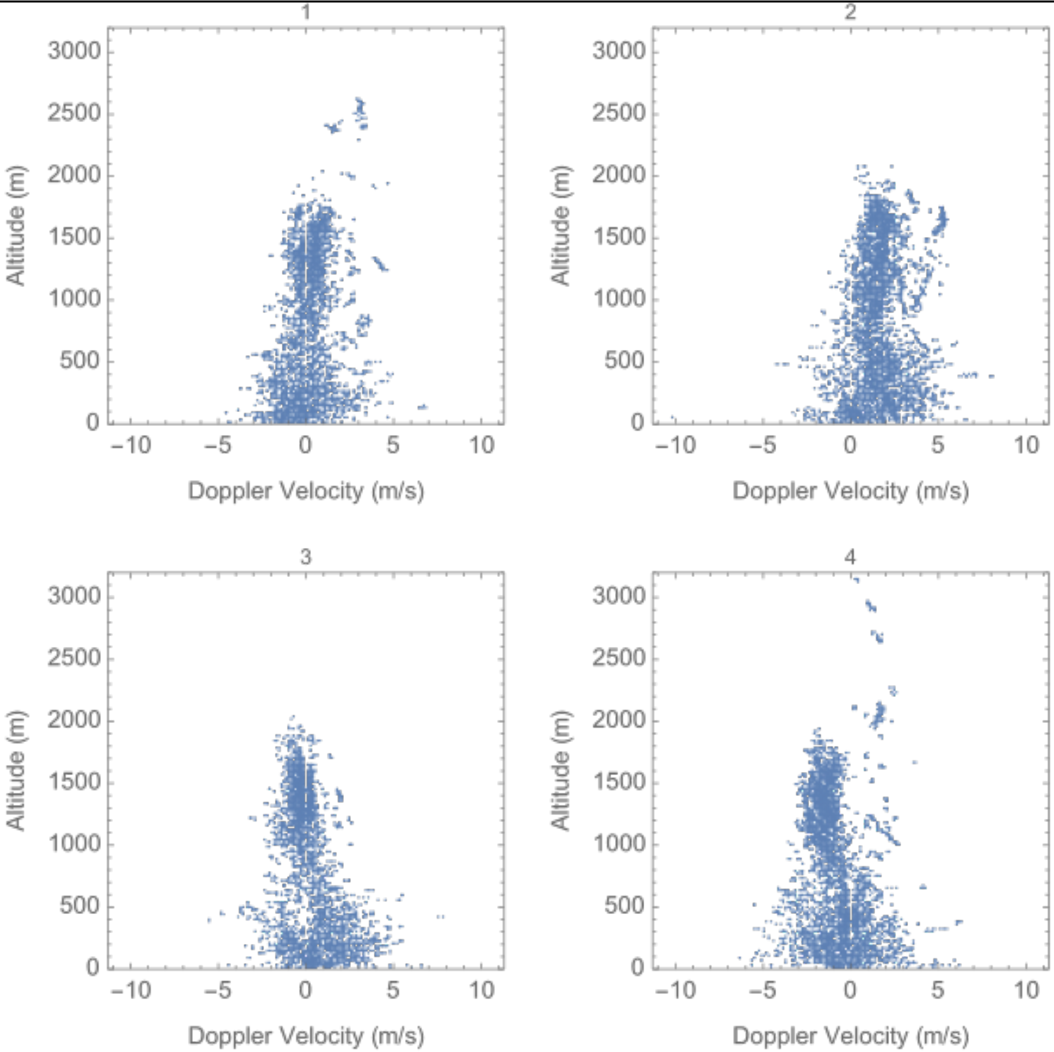
First 6 columns are the inputs to Low Contact Rate Wind Engine. LCRWE computes the spline estimate of the wind velocity profile. LCRWE can be used for both AWiPPR and WiPPR wind velocity profile estimation.

Classifier inputs and output of classifier. Inputs can drive classifiers other than the one used.

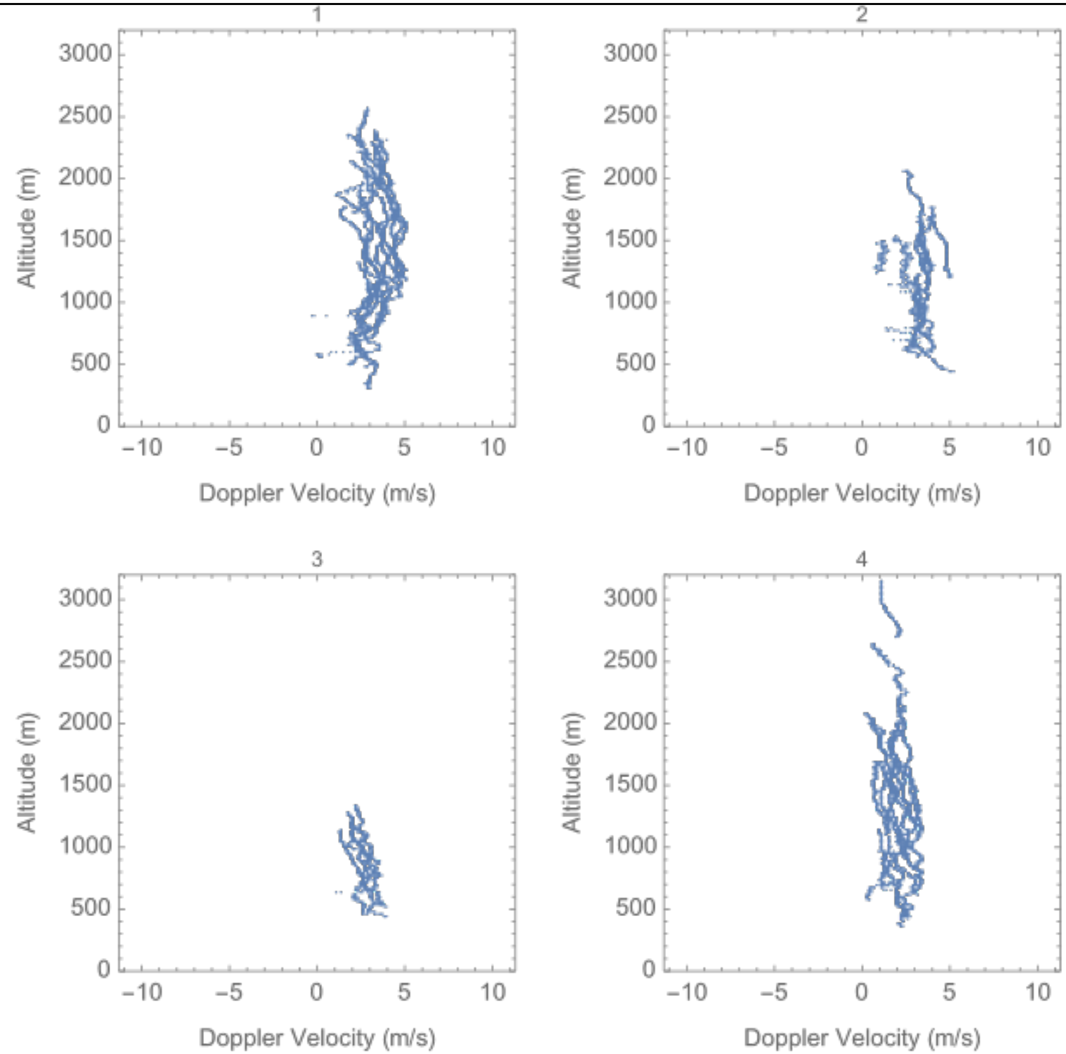
Classifier application: all contacts



Classifier application: clear air only



Classifier application: virga only



Key points

- The classifier described here enables the production of a contact catalog (inputs to LCRWE) that can be used to automatically determine the altitude at which wind sticks should be terminated.
- At present we terminate wind sticks at the altitude of the highest contact. This is a poor policy. Wind sticks should be terminated at the altitude where the system has contacts on at least three beams from the same type of object. This insures that the wind velocity estimate is not biased by the fall velocity of the object.
- The work described here documents a key component in a machine learning software suite that would allow for the sensible operation of WiPPR in an automated fashion.
- Next steps would include training a classifier that recognizes additional contact types such as high-energy noise transients, rain and snow.
- The extensive nation-wide, multi-season measurements that we made in 2103 can be used to support the development of classifiers. It should be strait forward to train a classifier to recognize interference from the raydome