https://doi.org/10.5194/amt-2023-69
Preprint. Discussion started: 11 April 2023
© Author(s) 2023. CC BY 4.0 License.

Atmospheric Measurement Techniques Discussions

Classification of flying insects in polarimetric weather radar using machine learning and aphid trap data

Samuel Kwakye[1], Heike Kalesse-Los[1], Maximilian Maahn[1], Patric Seifert[2], Roel van Klink[3,4], Christian Wirth[3], and Johannes Quaas[1,3,5]

[1]Leipzig Institute for Meteorology, Leipzig University, Leipzig, Germany.
[2]Leibniz Institute for Tropospheric Research, Leipzig, Germany
[3]German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany.
[4]Department of Computer Science, Martin Luther University-Halle Wittenberg, 06099, Halle (Saale), Germany.
[5]ScaDS.AI - Center for Scalable Data Analytics and Artificial Intelligence, Leipzig University, Humboldtstraße 25, 04105 Leipzig, Germany

Correspondence: samuel.kwakye@uni-leipzig.de

Abstract. Over the past decades, studies have observed strong declines in biomass and the abundance of flying insects. However, there are many locations where no surveys of insect biomass or abundance are available. Weather radars are known to provide quantitative estimates of flying insect biomass and abundance, and can therefore be used to fill knowledge gaps in space and time. In this study, we go beyond previous studies by combining a machine-learning approach with ground-truth observations from an aphid trap network. In this study, radar echoes from Level-II (Base) data of three Next Generation Weather Radar (NEXRAD) stations in the U.S. are classified using machine learning approaches. Weekly aphid counts from suction traps at Manhattan (Kansas), Morris (Illinois), and Rosemount (Minnesota) are used as validation data. Variability and distribution of the radar signals of four scatterer classes (insects, light rain, heavy rain, and plankton) are assessed. Probability density functions (PDF) of reflectivities of insects and plankton were found to be distinct from those of light- and heavy rain. Furthermore, the PDF of radar variables of the insect scatter class was also characterized by a broad distribution of spectrum width, cross-correlation ratio, and a broad range of differential reflectivity values. Decision trees, random forests, and support vector machine models were generated to distinguish three combinations of scatterers. A random forest classifier is found to be the best-performing model.

1 Introduction

The atmosphere can be considered as a living habitat for flying insect fauna and is then specifically termed the aerosphere. The abundance (concentration) and species richness of flying insects in the aerosphere is a key indicator of the quality of the Earth's biodiversity (Chilson et al., 2017). There is mounting evidence that insects, and in particular flying insects, are declining in biomass and density in many parts of the world (Van Klink et al., 2020; Hallmann et al., 2017), which could have

https://doi.org/10.5194/amt-2023-69
Preprint. Discussion started: 11 April 2023
© Author(s) 2023. CC BY 4.0 License.

Atmospheric Measurement Techniques Discussions

grave consequences for the ecosystem services insects deliver and the integrity of the ecosystems they are part of. However, the local differences in changes in insect abundance and biomass are large, and vary across the globe, across the terrestrial and freshwater realm, and across vegetation strata (Van Klink et al., 2020; Pilotto et al., 2020; Crossley et al., 2020) and only local data from a limited number of locations are available. Therefore larger-scale data is needed to assess global changes in insect abundance and biomass. Information on high-flying insects in the aerosphere can provide such large-scale information (Noskov et al., 2021), but is rarely available (Hu et al., 2016). As such, there is an urgency for quantitative knowledge of the changes in insect concentration, in its spatial extent and temporal scales. Weather radars are known to give quantitative information about flying insects to a considerable extent of the aerosphere. Networks of weather radar are spatially extensive and operate diurnally at all times. Insects have been assessed at such broad spatial and temporal scales in a number of studies before (Tielens et al., 2021; Stepanian et al., 2020).

Weather radars are generally designed to observe meteorological elements, therefore the derivation of insect echoes requires special techniques (Drake and Reynolds, 2012). Zrnic and Ryzhkov (1998) used polarimetric radar signatures of scatterers to differentiate insects, birds, and meteorological scatterers from an alternating polarimetric weather radar operating in the S-band frequency using numerical analysis. Insect signals in clear air were distinguished from Bragg scatterers using statistical analysis in Oklahoma from the frequency-modulated, continuous-wave (FMCW) S-band radar observation (Contreras and Frasier, 2008). Jatau and Melnikov (2019) used a fuzzy logic algorithm to separate insect echoes from precipitation and bird signals. With the increasingly widespread use of machine learning methods, the automation of insect identification methods from weather radars, calls for using artificial intelligence approaches (Chilson et al., 2019). Luke et al. (2008) used an automatized neural network algorithm to mask insect clutter from the U.S. Department of Energy Atmospheric Radiation Measurement (ARM) millimeter cloud radar echoes to distinguish them from cloud radar returns. Using a fully convolutional network classification method, Hu et al. (2019) classified weather echoes and biological scatterers. Evidently, artificial intelligence methods have contributed to the focus of research on the interpretation of more biologically, ecologically, and physically relevant information (phenotype of insects swarm, migration dynamics, and also number density of insects). One recent study of notice is the identification of different biological scatterer types and rain from the network of polarimetric weather surveillance radar (WSR-88D) of the United States using a random forest algorithm (Gauthreaux and Diehl, 2020). However, the study was not based on an in-situ validation dataset of insect occurrence.

As such, the present study seeks to i) find the distribution and variability of radar polarimetric variables for insect echoes from ground-truth observations from aphid suction trap data and precipitation scatterers (light rain, heavy rain) and plankton (plankton, as considered here, consists of leaves or other plant debris lifted to the atmosphere) ii) from a machine learning perspective, distinguish the scatterers by decision tree, random forest, and support vector machine algorithms.

2 Data and Methods

The S-band weather surveillance radar network (NEXRAD) of the United States of America (USA) is considered for this study. The radar network consists of 151 Doppler weather radars initialized in 1957 and established in 1959. After a series of

https://doi.org/10.5194/amt-2023-69
Preprint. Discussion started: 11 April 2023
© Author(s) 2023. CC BY 4.0 License.

Atmospheric Measurement Techniques Discussions

upgrades since 2013 is providing new insights due to its polarimetric capabilities (Stepanian et al., 2016). The radars with an antenna diameter of 8.5 m operate at a wavelength of 10 cm, beam-width of 0.95°, and a pulse width of 1.57 μs. The radars operate in two modes based on the atmospheric conditions i.e. the super and standard resolution. The super-resolution data products obtained in the lowest elevation scans have a resolution of 0.5° azimuthal by 0.25 km range resolution and the upper scans have the standard resolution of 1° azimuthal by 0.25 km range resolution. The radar scans azimuthally in a series of elevation angles ranging from 0.5° to 19.5°. Each elevation scan has a temporal resolution of 5 to 10 min with three legacy products (reflectivity, mean doppler velocity, and spectrum width) and three dual polarization products (differential reflectivity, correlation coefficient, and differential phase) (Handbook, 2005).

In this study, the radar moments differential reflectivity, cross-correlation ratio, reflectivity, spectrum width, and the range derivative of the differential phase which is the specific differential phase at 0.5° elevation angle were used. Reflectivity measures the strength of the returned horizontal signal. The abundance and size of scatterers can be inferred from the variable. Spectrum width is the standard deviation of the radial velocity of targets in the scattering volume, it provides the variability of the velocity of all scatterers in a radar observation volume. The differential reflectivity is derived as the logarithmic ratio of the reflectivity in the horizontally polarized and vertically polarized. The signal shapes of scatterers in a sampled volume radar influence this parameter. The correlation of the power and phase signals in the horizontal and vertical polarized is the cross-correlation coefficient. The diversity of scatterers can be inferred from this parameter (Kumjian, 2013). The differential phase is the delay in the phase of the horizontal and vertical radar signals. The specific differential phase which is derived from the range derivative of the differential phase is proportional to the concentration of scatterers. Specific differential phase was used due to its high variable importance in the differentiation of meteorological and non-meteorological radar echoes (Lakshmanan et al., 2015).

Four classes of radar scatterers were considered due to the seasonality of the in-situ validation data (described below) of aphids: insects, light rain, heavy rain, and plankton in this study. The scans of the radar moments were used in combination with the hydrometeor classification of the processed level III product. The fuzzy logic procedure of the hydrometeor classification is described in Park et al. (2009). To derive the corresponding classification of the level II radar moments scans of each scatterer class; the mask of the hydrometeor classification scatterer class was extracted and applied to the level II radar moments scans. In addition to the radar observations, weekly aphids count data from the suction trap at a height of 5.8 m above ground of the US Midwest suction trap network were used as in situ data for the occurrence of flying insects. The traps operate from mid-May to October and are functional during daylight hours. The suction traps are set up to capture aphids with 34 traps operating in 8 states. Aphids are small (1-5mm) long sap-sucking insects with complex life cycles (Van Emden and Harrington, 2017). Many species are important pests to cultivated plants, as they weaken the plant directly through suction damage, and indirectly as vectors of plant viruses. Three suction trap stations were considered in this study due to their proximity within 50 km of the nearest NEXRAD WSR-88D radar (Fig. 1). The stations were Manhattan (Kansas) (39.21°, -96.60°), Morris (Illinois) (41.35°, -88.38°), and Rosemount (Minnesota) (44.71°, -93.10°) and the radars are KTWX (38.99°, °-96.23), KLOT (41.60°, -88.08°), and KMPX (44.85°, -93.57°) respectively. Land cover types between the radar and the suction trap were mostly dominated by croplands and natural vegetation (Fig. 1).

https://doi.org/10.5194/amt-2023-69
Preprint. Discussion started: 11 April 2023
© Author(s) 2023. CC BY 4.0 License.

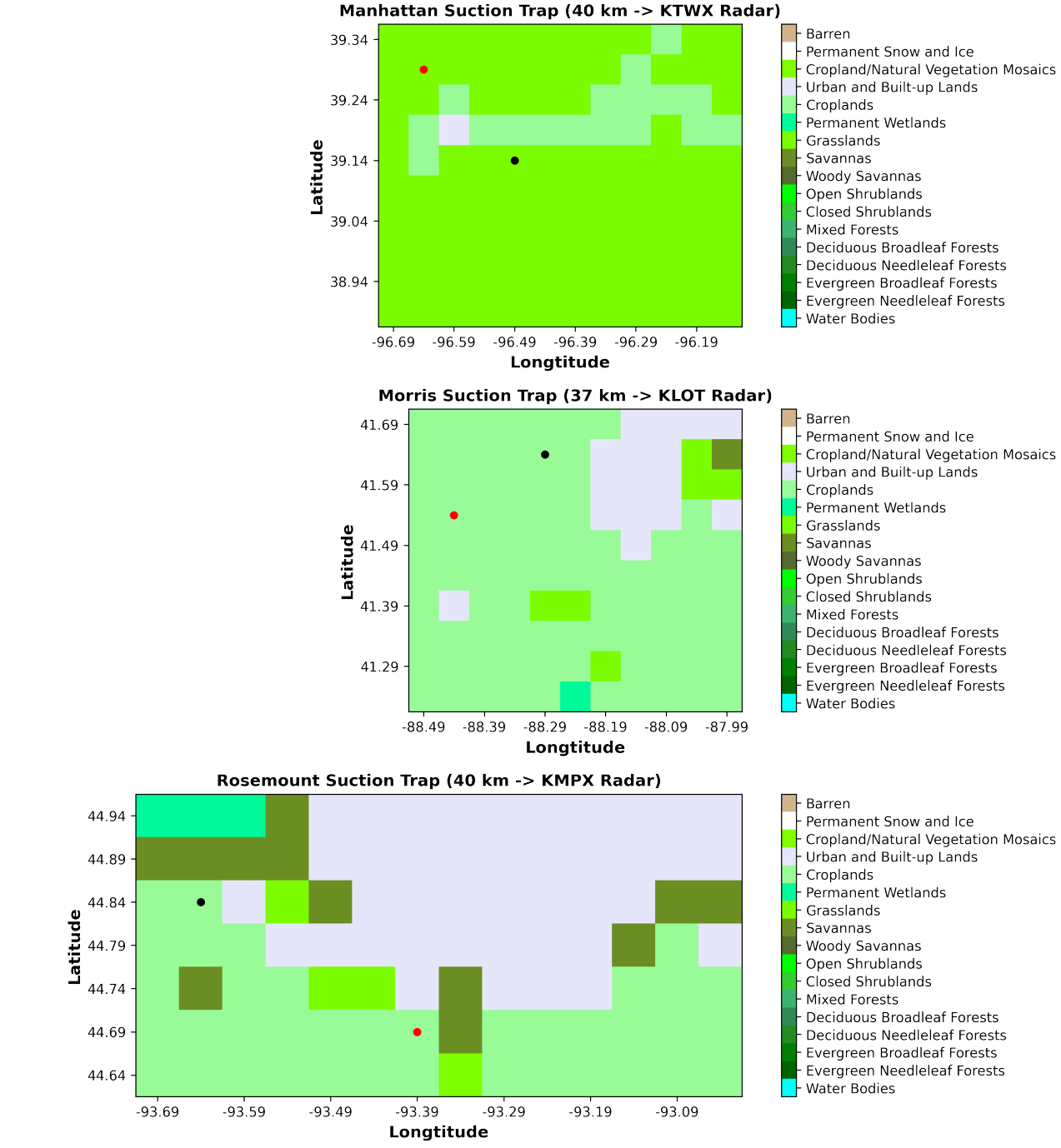


Figure 1. Map of land cover types at the suction trap stations in 2020. The red and black spots indicate the position of the suction trap and the radar respectively.

https://doi.org/10.5194/amt-2023-69
Preprint. Discussion started: 11 April 2023
© Author(s) 2023. CC BY 4.0 License.

Atmospheric Measurement Techniques Discussions

Atmospheric conditions favorable for aphids' flight initiation and concentration are calm weather with no precipitation (optimally, temperature > 18 °C and wind speeds < 2 ms^{-1}). Meteorological variables were inferred from the Local Climatological Data (LCD) of the National Centers for Environmental Information's Integrated Surface Data (ISD) data set. Taking into consideration the best weekly data was selected (Fig. 2).

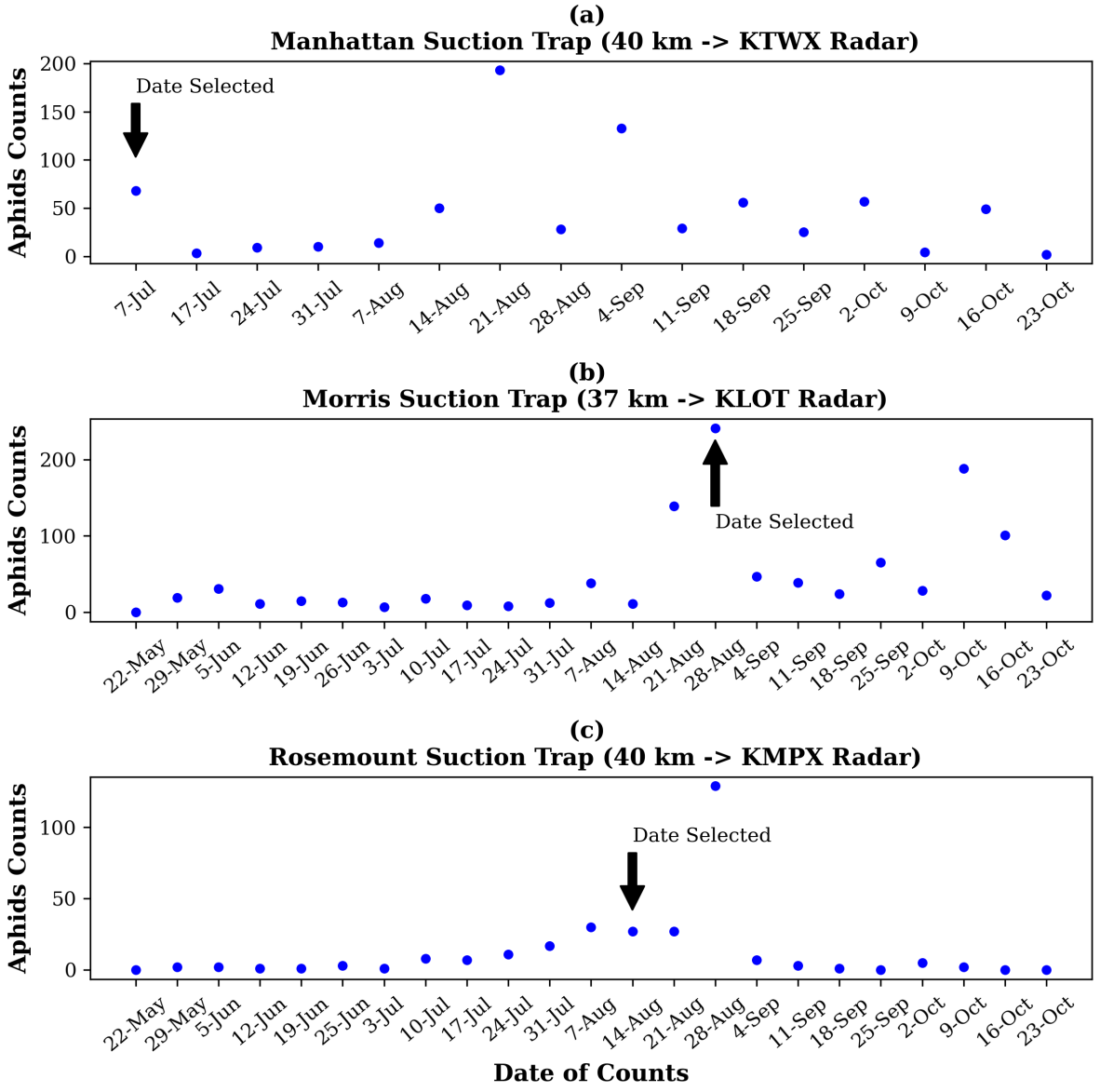


Figure 2. Time series plot of weekly aphid counts (date of collection and counting) at Manhattan (a), Morris (b), and Rosemount (c) suction trap station of the US Midwest suction trap network in 2020. The dates of the best weekly data are indicated.

https://doi.org/10.5194/amt-2023-69
Preprint. Discussion started: 11 April 2023
© Author(s) 2023. CC BY 4.0 License.

Atmospheric Measurement Techniques Discussions

Aphid concentration in the aerosphere usually peaks at 1100 h and 1700 h (Johnson, 1954). Occurrences of light rain (defined as rain rates measured at the meteorological station of $R < 2.5$ mmh^{-1}), heavy rain (defined as rain rates measured at the meteorological station of $R > 7.6$ mmh^{-1}) and plankton (clear air (non-rainy) days with high surface winds (> 22 knots) were inferred from the weather data. Due to discrepancies in the temporal resolution of aphids count data and the radar scans (weekly and 4 to 10 min respectively); the following procedure is followed in order to select individual radar scans which are then used to identify the cases of insect occurence at Manhattan, Morris, and Rosemount suction traps and assess the variability of the radar variables of the insect echoes:

1. The weeks where there are integral aphid counts are indicated in figures 1 to 4 as the selected week data.
2. As discussed above, 1100 h is identified as a time of day when aphids tend to fly preferably, so the scan at around 1100 h for each of the seven days is selected.
3. As discussed above, it is preferable for insect flight that temperatures are above 18°C and wind speeds below 2 m s^{-1}. Using LCD data, the optimal days of the week at the peak time that meet these criteria are selected.
4. Scans were selected for light rain and heavy rain, snow, and plankton based on the times of occurrence of each meteorological phenomenon. The times of occurrence were inferred from the LCD.

Scatterer Class	**Date and Time**		
	Manhattan (Kansas)	Morris (Illinois)	Rosemount (Minnesota)
Insects	06 July 2020 (17:01 UTC)	22 August 2020 (17:05 UTC)	10 August 2020 (17:00 UTC)
Light rain	19 June 2020 (18:03 UTC)	10 July 2020 (05:23 UTC)	02 June 2020 (22:41 UTC)
Heavy rain	07 April 2019 (9:15 UTC)	18 May 2019 (14:03 UTC)	11 June 2017 (05:04 UTC)
Plankton	01 June 2020 (16:02 UTC)	09 June 2020 (17:41 UTC)	01 June 2020 (14:02 UTC)

Table 1. Dates and times of selected radar scans for each scatterer class to determine the distribution and variability of the scatterers.

The extracted radar variables of each scatterer class at the four stations were aggregated and the variability and distribution of the radar variables of the scatterer classes were assessed by histograms.

Machine learning deduces patterns from a given dataset, the learning is based on statistical, mathematical, and computational approaches (Zewdie et al., 2019). Supervised learning as used in this study finds a functional relation that maps the input dataset to the classification label. To obtain the dataset for machine learning procedure these steps are followed at each station.

1. 50 scans of each scatterer were obtained within the selected aphid peak times. The preceding hours (1000 UTC and 1600 UTC) and following hours (1200 UTC and 1800 UTC) in addition to the peak hours of aphids (1100 UTC and 1700 UTC) at the selected times of insect occurence (Table 1).

https://doi.org/10.5194/amt-2023-69
Preprint. Discussion started: 11 April 2023
© Author(s) 2023. CC BY 4.0 License.

Atmospheric Measurement Techniques Discussions

2. 50 scans were selected for light rain and heavy rain, snow, and plankton based on the times of occurrence of each meteorological phenomenon. The times of occurrence were inferred from the LCD.

3. As stated the radar variables of each resolution volume were retrieved. The data was filtered to pixel points that echoed the five radar variables.

4. In combination, 160000 data points were retrieved for the four scatterers. The total radar data set was 480000 and was divided into a training (70%) and a testing dataset (30%).

Three supervised machine learning techniques are used in this study (Decision tree, random forest, and support vector machine) as they are suitable for pixel-level classification (Jatau et al., 2021). Decision tree classifier models are non-parametric and classify data by learning logical rules or threshold values (Kotsiantis, 2013). The module used in the study (Scikit-learn (Pedregosa et al., 2011)) distinguishes between classes based on threshold values and features at nodes with the highest information gain. A supervised classification algorithm consisting of an ensemble of decision trees is known as random forest (Pal, 2005). These trees are built from a subset of the training vectors at each tree node with equal distribution. Generally, each tree votes for a classification label as the forest selects the label with the highest votes (Breiman, 2001). However, the mean probabilistic prediction of the classification is used to determine the best label in the random forest as applied in this study (Pedregosa et al., 2011). The support vector machine classifies data by finding the best hyperplane that separates the classes at a decision boundary. The best hyperplane is obtained from a set of given functions evaluated by statistical learning theory (Shmilovici, 2009). The multi-class strategy used in this study was the one vs the rest to classify the scatterers. Decision tree, random forest, and support vector machine models were generated to classify and distinguish the scatterer classes. Optimally, the algorithm should be able to unambiguously identify each of these four classes insect, light rain, heavy rain, and plankton. The performance of the models was evaluated from the metrics: balanced accuracy, geometric mean score, and precision score. Balanced accuracy measures the mean of the ratio of positive classes determined by the model which is the recall defined as

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{1}$$

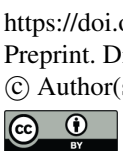
where TP is the number of predicted true positives samples, and FN is the number of predicted false negatives samples.
The ability of the models to correctly label positive and negative samples (true positives and true negatives) was measured by the precision score.

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2}$$

where TP is the number of predicted true positive samples, and FP is the number of predicted false positive samples.
The geometric mean score metric is selected as in addition to measuring the accuracy also determines the specificity (precision) and sensitivity score (recall) which are the ratio of the positive recall and negative recall of the data classes respectively. The geometric mean is given by the square root of the product of the specificity and the sensitivity.

$$\text{geometricmean} = \sqrt{\text{precision} * \text{recall}} \tag{3}$$

https://doi.org/10.5194/amt-2023-69
Preprint. Discussion started: 11 April 2023
© Author(s) 2023. CC BY 4.0 License.



Atmospheric Measurement Techniques Discussions

The performance of each scatterer class was assessed by the confusion matrix which gives values of the true positives and false negatives of each scatterer for the combinations.

2.1 Results

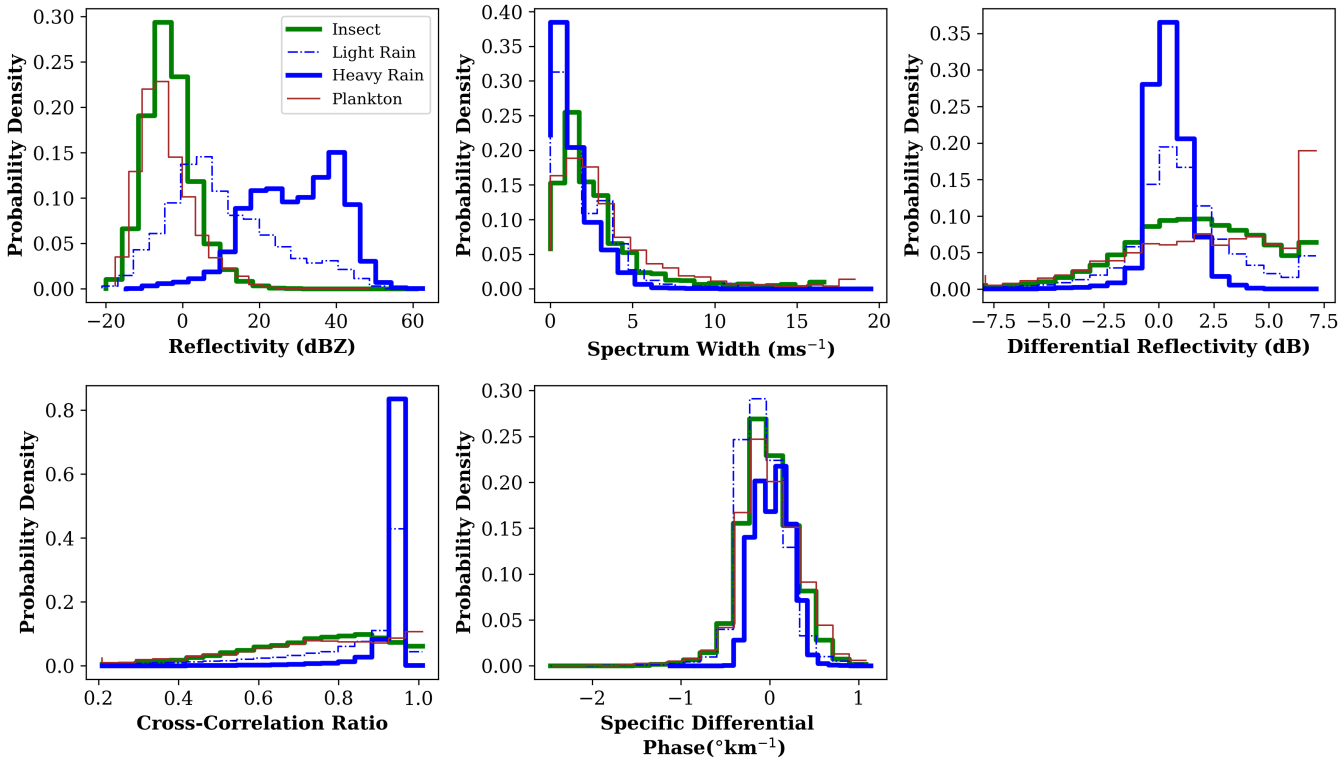


Figure 3. Normalized histograms of radar variables for insects as well as light rain, heavy rain, and plankton. Reflectivity, spectrum width, differential reflectivity, cross-correlation ratio (top left to right), differential reflectivity, and specific differential phase (bottom left to right). (bottom left to right). All scans for the dates listed in Table 1 are combined in the histograms

To assess the variability and distribution of the classes of scatterers, the extracted variables for the selected times of each scatterer are presented as histograms (Fig. 3). Compared to the other radar variables, reflectivity differentiated the distribution of the precipitation scatterers well with distinctive peaks. Plankton shows lower values and variability than insects. Insect's value of spectrum width is the highest at 2.5 ms^{-1} as the peak of the three other scatterers was near 0 ms^{-1}. Plankton is shown to have higher values than heavy and light rain.

The distribution of the precipitation scatterers shows differential reflectivity with peaks centered around 0 dB. The peak of heavy rain is almost two magnitudes higher than light rain with lower variability. The distribution of the insect's differential reflectivity has a broad peak that is maximum at 2 dB. Apart from a distinct peak at 7 dB, plankton has a high variability with

https://doi.org/10.5194/amt-2023-69
Preprint. Discussion started: 11 April 2023
© Author(s) 2023. CC BY 4.0 License.

Atmospheric Measurement Techniques Discussions

a peak at about -1 dB for differential reflectivity. Most of the cross-correlation ratio values of the precipitation scatterers are greater than 0.9 with the same peaks and cross-correlation ratios of plankton very highly variable and ranged between 0.2-1.0. As observed with differential reflectivity, the peak of heavy rain is two magnitudes higher than light rain. The majority of the insect scatterers have cross-correlation ratios between 0.70 to 0.85 and a slight peak at 0.85. Plankton has the highest peak of cross correlation ratio but with much more variability than the precipitation scatterers as the insect scatterers showed the highest variability for the variable. The specific differential phase distribution of the scatterer classes has minute differences. Precipitation scatterers showed less variability than insects and plankton. The peaks of the precipitation approach zero with the distribution skewed to the left.

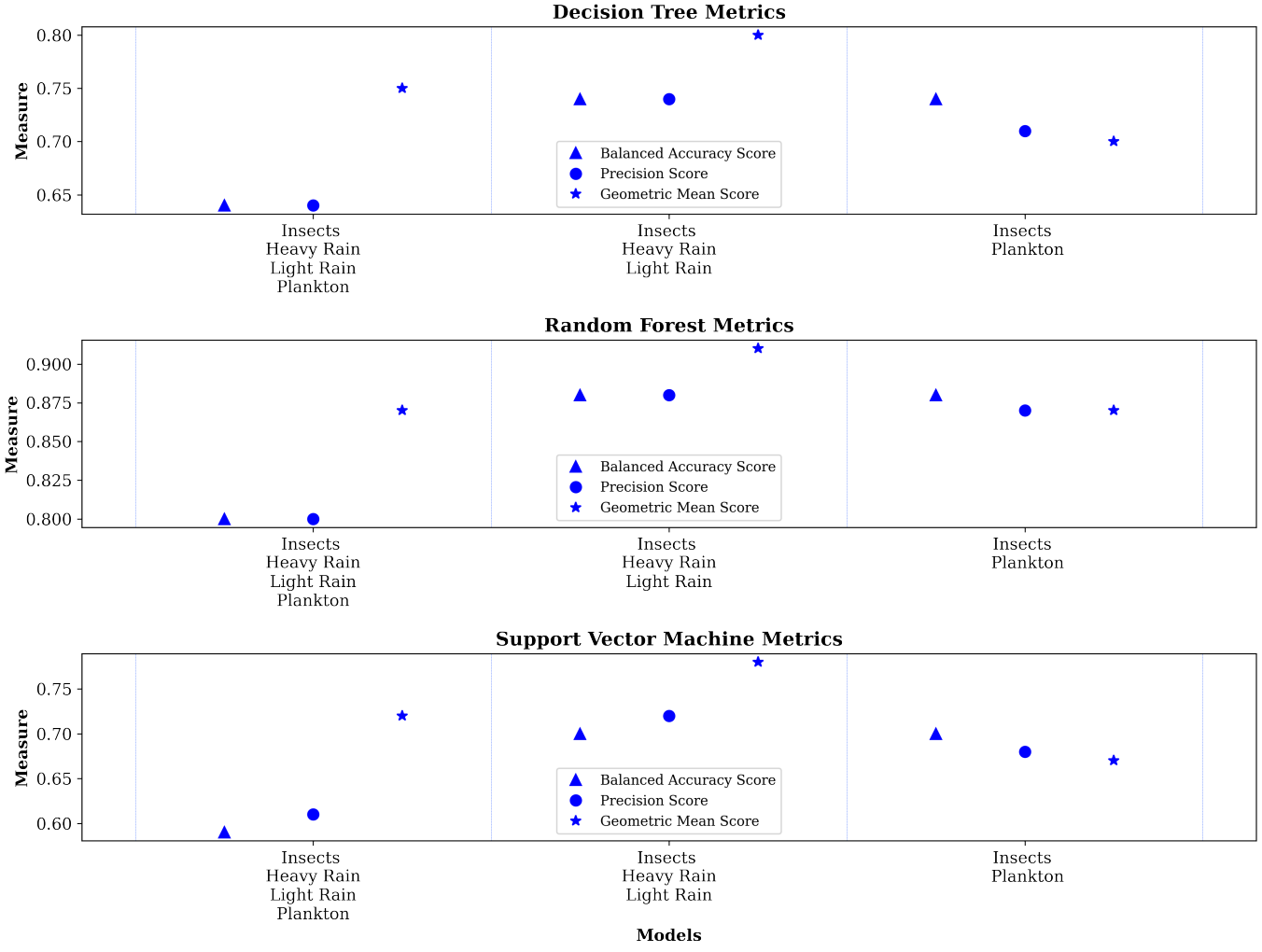


Figure 4. Metrics (balanced accuracy score, precision score, and geometric mean score) of the decision tree, random forest, and support vector machines models.

https://doi.org/10.5194/amt-2023-69
Preprint. Discussion started: 11 April 2023
© Author(s) 2023. CC BY 4.0 License.

Atmospheric Measurement Techniques Discussions

Each combination of the machine learning models was trained on a size of 70% of the dataset with a test size of 30%. The performance and evaluation of the three machine learning methods on the combination of scatterers are presented in Figure 4. In this study, the echoes of interest ie. insects, and the three other scatterers (I, HR, LR, P) precipitation scatterers were classified best by the random forest model. The prediction balanced accuracy and geometric mean score were 0.80 and 0.87 respectively as the precision score was 0.80. The two models for the decision tree and support vector machine classified this group of scatterers with a balanced accuracy score of 0.64 and 0.59 respectively. The values for the two other metrics were also higher for the decision tree model.

The random forest model classified best when omitting the plankton class and using only the insects' and precipitation scatterers (I, HR, LR) across the three metrics; balanced accuracy score (0.88), precision score (0.88), and geometric score (0.91). The decision tree was seen to perform slightly better than the support vector machine. As stated, plankton in this study are scatterers observed during conditions with high surface winds ie. >22 knots (11.32 ms^{-1}) which uplifts leaves and other debris in the atmosphere. This is distinguished from the insects' scatterer classes. Across the three machine learning models the classification metrics score was greater for the random forest model. The model had a prediction accuracy of 0.88, a precision score of 0.87, and a geometric mean of 0.87. The decision tree classifier was distinguished with a balanced accuracy score of 0.74 and the support vector machine with a score of 0.70.

https://doi.org/10.5194/amt-2023-69
Preprint. Discussion started: 11 April 2023
© Author(s) 2023. CC BY 4.0 License.

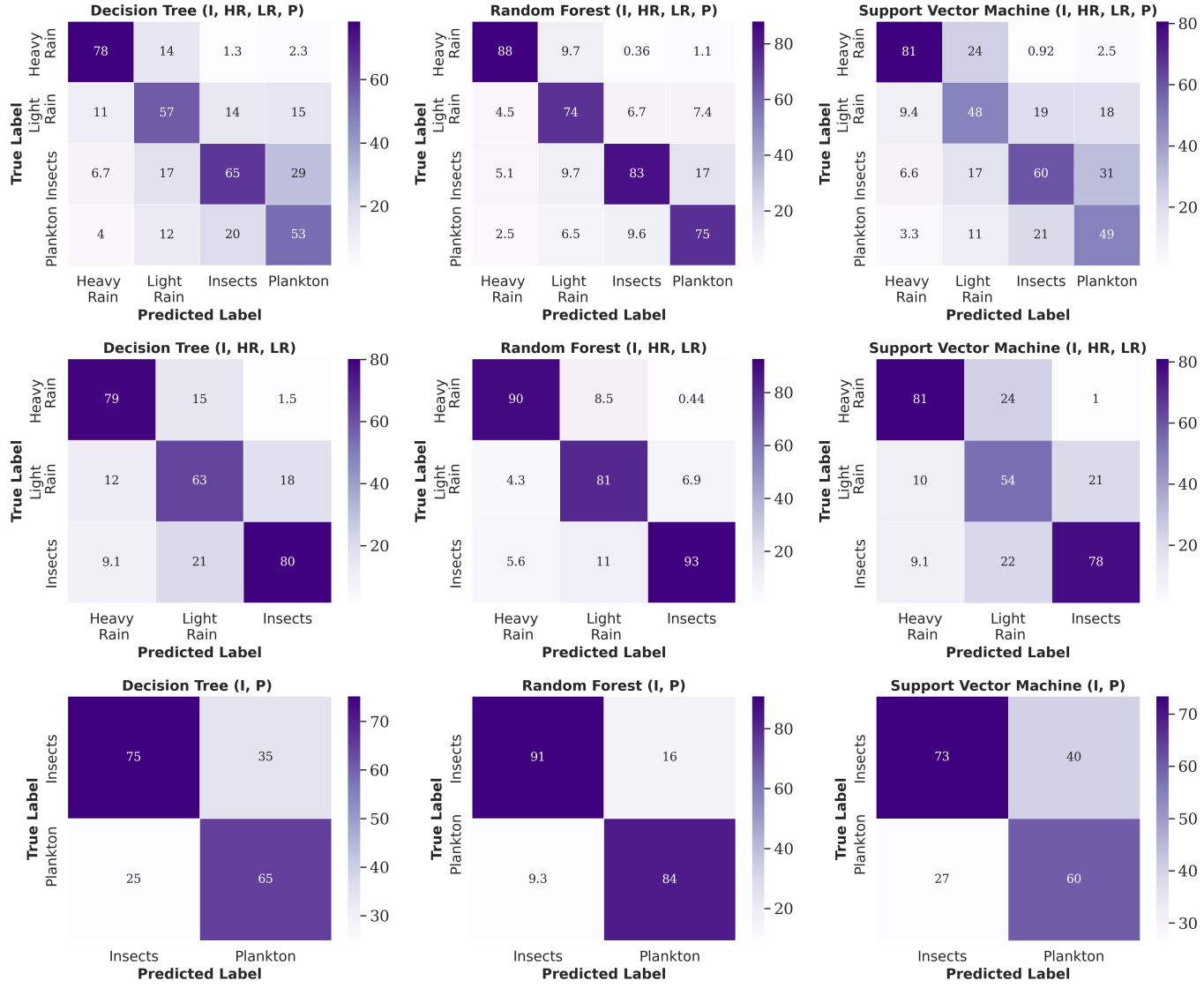


Figure 5. Normalized Confusion matrix in the percentage of insect, heavy rain, light rain, and plankton (upper panel); insect, heavy rain, and light rain (center panel); and insect and light rain (lower panel). Decision tree, random forest, and support vector machine (right to left).

The confusion matrix delved into the performance of each scatterer class for the combinations of the classification models (figure 5). Heavy rain was distinguished best from the other classes for the machine learning models. However, the insect scatterer was the most distinct apart from heavy rain with random forest performing best. The insect scatterer was misclassified most as plankton, the same was observed for the precipitation scatterer although the classification rate was lowest for light rain. This trend was observed for the other two models. In figure 5 (I, HR, LR), the insect scatterer was classified best with the random forest model, and the highest misclassification was light rain as light rain was also misclassified as insects the most.

https://doi.org/10.5194/amt-2023-69
Preprint. Discussion started: 11 April 2023
© Author(s) 2023. CC BY 4.0 License.

Atmospheric Measurement Techniques Discussions

This misclassification rate was observed for the decision tree and support vector machine models with heavy rain having higher true positives for the support vector machines and being equivalent with insects for the decision tree models. The highest true positive of the combination of insects and plankton was observed for the random forest model with the highest false-negative observed for the support vector machine model.

2.1.1 Discussion

Even though the size range of raindrops and insects is similar, the number concentration of raindrops can be much higher than that of insects in a radar measurement volume leading to higher reflecitivities.(Kumjian, 2013). Plankton conditions are almost clear air with minute interaction with radar signals. The reflectivity observed shows that the power of the scatterers returned signals ascended from plankton, insects, light rain, and heavy rain. The spectrum width showed large variability in the velocity of the insects' flight pattern as precipitation showed much more uniform velocity in their drop velocity.

Precipitation signals tend to have differential reflectivity values close to 0 dB due to their relatively spherical shape in polarimetric measurements and variates around this median value (Kumjian, 2013) (Gauthreaux and Diehl, 2020). In this study, this is clearly observed without exceptional deviations. Differential reflectivity is sensitive to the shape of echo signals of objects encountered. The disproportionate shape of insects returns high values of differential reflectivity (Melnikov et al., 2015). However, the value of the insect signal is relatively lower. This can be expected due to the small size and plump shape of aphids used as a proxy for insect occurrence. Insects' signals are characteristically inhomogeneous and usually yield low correlation coefficient values and vice versa for precipitation scatterers (Kumjian, 2013). This is evidenced in the distribution obtained. The high peak of plankton scatterer can be attested to its low diversity due to the clear air nature of this atmospheric condition. Although the specific differential phase showed differences in their distributional variability and peaks for all the scatterers; the high concentration of heavy rain showed much higher peaks than the other scatterer classes.

Distinguishing and classifying echoes of weather scans is ideal to evaluate insect echoes. Machine learning methods are effective in the classification of radar echoes (Jatau et al., 2021). The results of the three combinations of scatterers showed that the random forest algorithm distinguished the scatterers best ahead of the decision tree and support vector machine. Throughout, insects and the precipitation scatterer classes was seen to be distinguished best by the classifiers. As seen from the confusion matrices, the most distinguishable scatterer was heavy rain from the other scatterers. This might be due to its strong echo signals as observed in the distribution of the radar moments. As seen also, insect signals were the most distinguishable class amongst the other scatterer class.

3 Conclusions

Insect decline has been noticeable with limited evidence for aphids of the Hemiptera order (Sánchez-Bayo and Wyckhuys, 2019). The dynamics of aphid flight initiation and migration intensity are dependent on weather conditions. Weather radars are spatially extensive, operate at all times, and are able to observe insects' activity. However, insects as targets of weather radar are secondary to meteorologists and mainly interesting because they bias weather phenomena echoes. Weekly aphid counts from

https://doi.org/10.5194/amt-2023-69
Preprint. Discussion started: 11 April 2023
© Author(s) 2023. CC BY 4.0 License.

Atmospheric Measurement Techniques Discussions

suction traps obtained from Kanawha, Manhattan, Morris, and Rosemount are used as validation data for insect occurrence. Weather conditions optimal for aphids and peak times were used to obtain radar scans for the insect's signals. Light rain, heavy rain, and plankton radar scans are obtained from climatological data times of occurrence. Radar measurements from four S-band WSR-88D NEXRAD radars nearest to the suction traps are analyzed (KDMX, KTWX, KLOT, and KMPX). The distribution and variability of the radar variables of the scatterers were assessed. Generally, the insect situations are different from the three other scatterer classes in particular in the following:

- Differential reflectivity is larger for the insects.
- Cross-correlation ratio is lower for insects.
- Spectrum width values of insects were high and distinct.
- Reflectivity tends to be in between the ranges of -3 and 5.25 dBZ.

Machine learning classifiers have been applied to weather radar echoes to distinguish insects' signals from other scatterers. Random forest classifiers are seen to perform reasonably better in distinguishing the combination of scatterers compared to the other algorithms.

Data availability. Next generation weather radar level-II data, level-III hydrometeor classification, and the Local Climatological Data (LCD) of the National Centers for Environmental Information's Integrated Surface Data (ISD) data set of US were obtained and openly available at (https://www.ncdc.noaa.gov/). Aphids count data was obtained and openly available at (https://suctiontrapnetwork.org/data/). MODIS Land cover type data was obtained and openly available at (https://www.earthdata.nasa.gov/)

.

Author contributions. SK and JQ designed the experiments and SK carried them out. SK and JQ prepared the manuscript with contributions from all co-authors.

Competing interests. One of the (co-)authors (Maximilian Maahn) is a member of the editorial board of Atmospheric Measurement Techniques

https://doi.org/10.5194/amt-2023-69
Preprint. Discussion started: 11 April 2023
© Author(s) 2023. CC BY 4.0 License.

Atmospheric Measurement Techniques Discussions

Acknowledgements. Thanks to Doris Lagos-Kutz of the Department of Crop Sciences, the University of Illinois Glen Hartman of USDA-ARS as data producers, and Joseph LaForest as database coordinator for the aphids count data of Morris. The suction trap data is part of the North Central Soybean Research Program of the Midwest suction trap network.

https://doi.org/10.5194/amt-2023-69
Preprint. Discussion started: 11 April 2023
© Author(s) 2023. CC BY 4.0 License.

Atmospheric Measurement Techniques Discussions

References

Breiman, L.: Random forests, Machine learning, 45, 5–32, https://doi.org/10.1023/A:1010950718922, 2001.

Chilson, C., Avery, K., McGovern, A., Bridge, E., Sheldon, D., and Kelly, J.: Automated detection of bird roosts using NEXRAD radar data and Convolutional Neural Networks, Remote Sensing in Ecology and Conservation, 5, 20–32, https://doi.org/doi.org/10.1002/rse2.92, 2019.

Chilson, P. B., Frick, W. F., Kelly, J. F., and Liechti, F.: Aeroecology: An integrative view of the atmosphere, in: Aeroecology, pp. 3–11, Springer, https://doi.org/10.1007/978-3-319-68576-2, 2017.

Contreras, R. F. and Frasier, S. J.: High-resolution observations of insects in the atmospheric boundary layer, Journal of Atmospheric and Oceanic Technology, 25, 2176–2187, https://doi.org/10.1175/2008JTECHA1059.1, 2008.

Crossley, M. S., Meier, A. R., Baldwin, E. M., Berry, L. L., Crenshaw, L. C., Hartman, G. L., Lagos-Kutz, D., Nichols, D. H., Patel, K., Varriano, S., et al.: No net insect abundance and diversity declines across US Long Term Ecological Research sites, Nature Ecology & Evolution, 4, 1368–1376, https://doi.org/10.1038/s41559-020-1269-4, 2020.

Drake, V. A. and Reynolds, D. R.: Radar entomology: observing insect flight and migration, Cabi, https://doi.org/10.1079/9781845935566.0000, 2012.

Gauthreaux, S. and Diehl, R.: Discrimination of biological scatterers in polarimetric weather radar data: Opportunities and challenges, Remote Sensing, 12, 545, https://doi.org/10.3390/rs12030545, 2020.

Hallmann, C. A., Sorg, M., Jongejans, E., Siepel, H., Hofland, N., Schwan, H., Stenmans, W., Müller, A., Sumser, H., Hörren, T., et al.: More than 75 percent decline over 27 years in total flying insect biomass in protected areas, PloS one, 12, e0185 809, https://doi.org/10.1371/journal.pone.0185809, 2017.

Handbook, F.: Federal meteorological handbook no. 11, Doppler Radar Meteorological Observations, Part B, 2005.

Hu, C., Li, S., Wang, R., Cui, K., Wu, D., and Ma, S.: Extracting animal migration pattern from weather radar observation based on deep convolutional neural networks, The Journal of Engineering, 2019, 6541–6545, https://doi.org/10.1049/joe.2019.0041, 2019.

Hu, G., Lim, K. S., Horvitz, N., Clark, S. J., Reynolds, D. R., Sapir, N., and Chapman, J. W.: Mass seasonal bioflows of high-flying insect migrants, Science, 354, 1584–1587, https://doi.org/10.1126/science.aah43, 2016.

Jatau, P. and Melnikov, V.: Classifying Bird and Insect Radar Echoes at S Band, in: 99th American Meteorological Society Annual Meeting, AMS, 2019.

Jatau, P., Melnikov, V., and Yu, T.-Y.: A machine learning approach for classifying bird and insect radar echoes with S-band polarimetric weather radar, Journal of Atmospheric and Oceanic Technology, 38, 1797–1812, https://doi.org/10.1175/JTECH-D-20-0180.1, 2021.

Johnson, C.: Aphid migration in relation to weather, Biological Reviews, 29, 87–118, https://doi.org/10.1111/j.1469-185X.1954.tb01398.x, 1954.

Kotsiantis, S. B.: Decision trees: a recent overview, Artificial Intelligence Review, 39, 261–283, https://doi.org/10.1007/s10462-011-9272-4, 2013.

Kumjian, M. R.: Principles and Applications of Dual-Polarization Weather Radar. Part I: Description of the Polarimetric Radar Variables., Journal of Operational Meteorology, 1, https://doi.org/10.15191/nwajom.2013.0119, 2013.

Lakshmanan, V., Karstens, C., Krause, J., Elmore, K., Ryzhkov, A., and Berkseth, S.: Which polarimetric variables are important for weather/no-weather discrimination?, Journal of Atmospheric and Oceanic Technology, 32, 1209–1223, https://doi.org/10.1175/JTECH-D-13-00205.1, 2015.

https://doi.org/10.5194/amt-2023-69
Preprint. Discussion started: 11 April 2023
© Author(s) 2023. CC BY 4.0 License.

Atmospheric Measurement Techniques Discussions

Luke, E. P., Kollias, P., Johnson, K. L., and Clothiaux, E. E.: A technique for the automatic detection of insect clutter in cloud radar returns, Journal of Atmospheric and Oceanic Technology, 25, 1498–1513, https://doi.org/10.1175/2007JTECHA953.1, 2008.

Melnikov, V. M., Istok, M. J., and Westbrook, J. K.: Asymmetric radar echo patterns from insects, Journal of Atmospheric and Oceanic Technology, 32, 659–674, 2015.

Noskov, A., Bendix, J., and Friess, N.: A Review of Insect Monitoring Approaches with Special Reference to Radar Techniques, Sensors, 21, 1474, https://doi.org/10.3390/s21041474, 2021.

Pal, M.: Random forest classifier for remote sensing classification, International journal of remote sensing, 26, 217–222, https://doi.org/10.1080/01431160412331269698, 2005.

Park, H. S., Ryzhkov, A., Zrnić, D., and Kim, K.-E.: The hydrometeor classification algorithm for the polarimetric WSR-88D: Description and application to an MCS, Weather and forecasting, 24, 730–748, https://doi.org/10.1175/2008WAF2222205.1, 2009.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E.: Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research, 12, 2825–2830, https://doi.org/10.48550/arXiv.1201.0490, 2011.

Pilotto, F., Kühn, I., Adrian, R., Alber, R., Alignier, A., Andrews, C., Bäck, J., Barbaro, L., Beaumont, D., Beenaerts, N., et al.: Meta-analysis of multidecadal biodiversity trends in Europe, Nature communications, 11, 1–11, https://doi.org/10.1038/s41467-020-17171-y, 2020.

Sánchez-Bayo, F. and Wyckhuys, K. A.: Worldwide decline of the entomofauna: A review of its drivers, Biological conservation, 232, 8–27, https://doi.org/10.1016/j.biocon.2019.01.020, 2019.

Shmilovici, A.: Support vector machines, in: Data mining and knowledge discovery handbook, pp. 231–247, Springer, https://doi.org/10.1007/978-0-387-09823-4, 2009.

Stepanian, P. M., Horton, K. G., Melnikov, V. M., Zrnić, D. S., and Gauthreaux Jr, S. A.: Dual-polarization radar products for biological applications, Ecosphere, 7, e01 539, https://doi.org/10.1002/ecs2.1539, 2016.

Stepanian, P. M., Entrekin, S. A., Wainwright, C. E., Mirkovic, D., Tank, J. L., and Kelly, J. F.: Declines in an abundant aquatic insect, the burrowing mayfly, across major North American waterways, Proceedings of the National Academy of Sciences, 117, 2987–2992, https://doi.org/10.1073/pnas.191359811, 2020.

Tielens, E. K., Cimprich, P. M., Clark, B. A., DiPilla, A. M., Kelly, J. F., Mirkovic, D., Strand, A. I., Zhai, M., and Stepanian, P. M.: Nocturnal city lighting elicits a macroscale response from an insect outbreak population, Biology Letters, 17, 20200 808, https://doi.org/10.1098/rsbl.2020.0808, 2021.

Van Emden, H. F. and Harrington, R.: Aphids as crop pests, Cabi, https://doi.org/10.1564/v28_oct_09, 2017.

Van Klink, R., Bowler, D. E., Gongalsky, K. B., Swengel, A. B., Gentile, A., and Chase, J. M.: Meta-analysis reveals declines in terrestrial but increases in freshwater insect abundances, Science, 368, 417–420, https://doi.org/10.1126/science.aax9931, 2020.

Zewdie, G. K., Lary, D. J., Liu, X., Wu, D., and Levetin, E.: Estimating the daily pollen concentration in the atmosphere using machine learning and NEXRAD weather radar data, Environmental monitoring and assessment, 191, 1–9, https://doi.org/10.1007/s10661-019-7542-9, 2019.

Zrnic, D. S. and Ryzhkov, A. V.: Observations of insects and birds with a polarimetric radar, IEEE Transactions on Geoscience and Remote Sensing, 36, 661–668, https://doi.org/10.1109/36.662746, 1998.