Mid-Atlantic Nocturnal Low-Level Jet Characteristics: A machine learning analysis of radar wind profiles

Maurice Roots[1,2], John T. Sullivan[3], Belay Demoz[1,2]

[1]Department Physics, University of Maryland, Baltimore County (UMBC), Baltimore, 21250, USA
[2]Goddard Earth Sciences Technology and Research (GESTAR) II, Baltimore, 20771, USA
[3]NASA Goddard Space Flight Center (GSFC), Greenbelt, 20771, USA

Correspondence to: Maurice Roots (mroots1@umbc.edu)

Abstract. This paper introduces a machine-learning-driven approach for automated Nocturnal Low-Level Jet (NLLJ) identification using observations of wind profiles from a Radar Wind Profiler (RWP). The work discussed here is an effort to lay the groundwork for a systematic study of the Mid-Atlantic NLLJ's formation mechanisms and their influence on nocturnal and diurnal air quality in major urban regions by establishing a general framework of NLLJ features and characteristics with an identification algorithm. Leveraging a comprehensive wind profile dataset maintained by the Maryland Department of Environment's RWP network, our methodology employs supervised machine learning techniques to isolate the features of the south-westerly NLLJ, because of its association with pollution transport in the Mid-Atlantic states. This methodology was developed to illuminate spatiotemporal patterns and nuanced characteristics of NLLJ events, unveiling their significant role in shaping the planetary boundary layer. This paper discusses the construction of this methodology, its performance against known NLLJs in the current literature, intended usage, and a preliminary statistical analysis. First light results from this analysis have identified a total of 90 south-westerly NLLJs from May - September of 2017 - 2021 as captured by the RWP stationed in Beltsville, MD (39.05° N, 76.87° W, 135 m ASL). A composite of these 90 jets is presented to better illustrate many of the bulk parameters, such as core height, duration, and maximum wind speed, associated with the onset and decay of the Mid-Atlantic NLLJ. We hope our study equips researchers and policymakers with further means to monitor, predict, and address these nocturnal dynamics phenomena that frequently influence boundary layer composition and air quality in the U.S. Mid-Atlantic and Northeastern regions.

1 Introduction

Low-Level jets (LLJs) are broadly defined as localized wind speed maxima that occur within the lower troposphere accompanied by decreasing wind speed above the maximum (Stensrud, 1996). LLJs have been reported all over the world under a wide range of formation mechanisms with varying characteristics, and subsequently different impacts to the lower troposphere (De Jong et al., 2024; Ortiz-Amezcua et al., 2022; Lima et al., 2019, 2018; Tuononen et al., 2017; Ranjha et al., 2015; Karipot et al., 2009; Baas et al., 2009; Zhang et al., 2006; Corsmeier et al., 1997; Blackadar, 1957). Under this broad definition, there are many different types of LLJs, however, in this study, we focus on long-lived nocturnal LLJs (NLLJs) as a focus for future study of impacts to boundary layer chemistry. These NLLJs are important in moisture transport and air pollutant transport (Wei et al., 2023; Roots et al., 2023; Sullivan, 2017; Delgado et al., 2015; Weldegaber, 2009; Tollerud et al., 2008; Weaver and Nigam, 2008; Ryan, 2004; Corsmeier et al., 1997; Stensrud, 1996). We focus on furthering the study of the NLLJs reported by researchers in the East Coast Mid-Atlantic region of the United States. We do this by developing a framework for our future work in developing a climatology and systematic study of LLJs by characterizing their "critical characteristics" (maximum wind speed, height of maximum, duration, wind direction, etc.) and formation mechanisms (synoptic influence, temperature gradients, inertial oscillation, diurnal cycle, etc.). Herein we describe an algorithm developed to detect South-Westerly NLLJs in the study area of Maryland (MD) in the United States using the Maryland Department of Environment's (MDE) 915 MHz DeTect RAPTOR DBS-BL/LAP-3000 Radar Wind

Profiler (RWP) stationed in Beltsville, MD. We use these systems because we hope to adapt our methods for the network of wind profilers in the area, and our region lacks sufficient decadal measurements of wind profilers from other more commonly used systems like Doppler wind lidar.

The Mid-Atlantic nocturnal low-level jet (NLLJ), similar to the Southern Great Plains (SGP) NLLJ, arises from the cooling of the low-level air mass relative to the air above it, resulting in a stratified nocturnal boundary layer and subsequent decoupling (Rabenhorst et al., 2014; Zhang et al., 2006). This decoupling facilitates the development of a low-friction residual layer where a super-geostrophic wind maximum emerges near the surface due to inertial oscillation, as described by Blackadar (1957) and later refined by Holton (1967). The genesis of NLLJs is influenced by a confluence of geographic-specific atmospheric dynamics, including the formation of a pronounced temperature inversion within the stratified nocturnal boundary layer, diurnal pressure shifts, and the influence of terrain (Shapiro and Fedorovich, 2010; Holton, 1967; Blackadar, 1957). These jets typically exhibit wind speed maxima at altitudes between 200 to 800 meters above ground level (AGL), with directional variability governed by geographical and meteorological conditions, though they generally flow northward as per Blackadar's theory. The unique synoptic and diurnal physical conditions that define NLLJs make them more prevalent during spring and summer when formation conditions are more favourable (Bonner, 1968; Shapiro et al., 2016; Shapiro and Fedorovich, 2010; Zhang et al., 2006). The implications of NLLJs extend significantly into weather, climate, and air quality, as they play a crucial role in the transport and mixing of atmospheric constituents such as pollutants, moisture, and heat, thereby influencing air quality and promoting cloud formation (Baas et al., 2009; Banta, 2008; Mahrt, 1998). In the Great Plains, NLLJs have been extensively documented since the 1950s, where they contribute to moisture transport and regional weather, including convective storm development (Banta et al., 2003; Carroll et al., 2019, 2021; Lundquist, 2003; Stensrud, 1996; Tollerud et al., 2008; Whiteman et al., 1997). The Mid-Atlantic NLLJ, while analogous to the SGP NLLJ in its reliance on inertial oscillation theory combined with the influence of temperature gradients induced by sloping terrain (Shapiro et al., 2016); however, with lower wind-speed maximums and vastly different topographic influences, with the Appalachian Mountains to the East and North, the Chesapeake Bay and Atlantic Ocean to the West, and the Coastal Plains and Piedmont region in between.

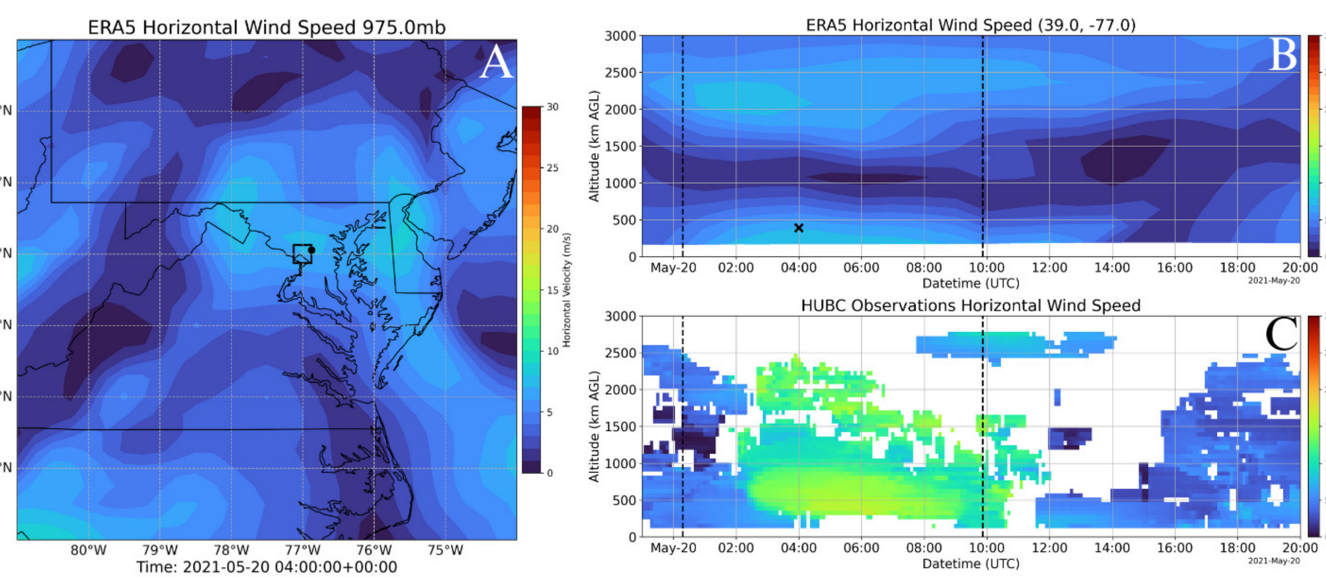

Figure 1: Example depiction of the nocturnal low-level jet in the Mid-Atlantic US on May 20, 2021: (A) ERA5 Horizontal Wind Speed at 975 mb ("x"); (B) shows the vertical profile evolution of the horizontal wind speed taken from a vertical slice (black square); (C) shows the horizontal wind speed from (black circle). Dashed vertical lines indicate the sunset and sunrise times, respectively.

We define the Mid-Atlantic NLLJ closely following the results found by Zhang et al. (2006) and Ryan, (2004) in their Fort Meade, MD RWP (decommissioned in 2006) observations. These studies provided detailed observations and analysis of NLLJ events, mainly focusing on their occurrence, structure, and dynamics within the Mid-Atlantic region of the United States. According to Zhang et al. (2006) and Ryan, (2004), the Mid-Atlantic NLLJ is characterized by a robust and low-level wind speed maximum that typically occurs during the nighttime hours. These jets are predominantly observed during the warm season (late spring through early fall). The studies noted that winds from the south and southwest directions dominate most of the Mid-Atlantic NLLJs. Figure 1 provides an illustrative example of the spatial and temporal extent of the Mid-Atlantic nocturnal low-level jet (NLLJ) as depicted by an event reported on May 20, 2021, by Roots et al., 2023: Panel (A) displays the 0.25° spatial and hourly temporal resolution European Centre for Medium-Range Weather Forecasts Reanalysis 5th Generation (ERA5, Hersbach et al., 2020) reanalysis data showing horizontal wind speeds at 975 mb (chosen to be near the core of the NLLJ in both the model and observations), highlighting the spatial distribution of the NLLJ across the region. Panel (B) presents the vertical evolution of the horizontal wind speed along a vertical slice (indicated by the black square in panel A), capturing the temporal progression of the jet's strength and altitude through the night. Panel (C) provides a visual comparison between this reanalysis data and the observations from our site in Maryland, showing much less magnitude in the horizontal wind speed evolution than the observations from the black circle marked in panel B. The impact of NLLJs on air pollution in the Mid-Atlantic is particularly significant during warm seasons, as these jets contribute to the transport of pollutants across the East Coast, elevating surface ozone and particulate matter concentrations (Delgado et al., 2015; Roots et al., 2023; Sullivan, 2017; Weldegaber, 2009; Ryan, 2004). Zhang et al. (2006) reported that approximately 60% of the Mid-Atlantic NLLJs observed during their study period (warm seasons in 2001 and 2002) exhibited this southerly and south-westerly wind direction. Ryan (2004) contributed to our definition by providing insights into the frequency and timing of NLLJ occurrences, noting that these events were common during the study period from 1998 to 2002, noting 80 warm season cases in total. Together, the work of Zhang et al. (2006) and Ryan, (2004) defines the Mid-Atlantic NLLJ as a nocturnal atmospheric phenomenon characterized by a significant increase in wind speed (~15 m s-1) at low altitudes (400 – 600 m AGL), typically showing a preferential direction from the south or southwest.

This work presents the culmination of an investigation into the NLLJ phenomena within the Mid-Atlantic region, leveraging a supervised machine-learning model tested against a comprehensive dataset including previously reported NLLJ events. The model, designed with a focus on advancing our capability to detect and analyze NLLJs, was evaluated using cases from notable studies by Sullivan et al. (2017), Delgado et al. (2015), and Weldegaber (2009) based on data from the RWP stationed in Beltsville, Maryland, US (39.05° N, 76.87° W, 135 m ASL). Without established benchmarks for NLLJ detection accuracy, our analysis adopts a qualitative approach, emphasizing visual inspection to assess the model's performance in accurately capturing NLLJ characteristics, particularly wind speed and direction. The primary objective of this research is to transition from episodic, qualitative analyses to a systematic, quantitative understanding of NLLJ physics and its impacts, utilizing observational data to explore the temporal distribution, morphology, and statistical properties of Mid-Atlantic NLLJs. Furthermore, developing a generalized representation of Mid-Atlantic NLLJs based on observational data marks a significant step forward in our ability to identify and analyze these phenomena.

The rest of the paper is structured as follows: Section 2 outlines the dataset and study area. Section 3 describes the methods, alongside the development and application of machine learning algorithms for detecting NLLJ features. Section 4 evaluates the performance of these algorithms in isolating NLLJ characteristics within the wind profile data, addressing the efficacy and limitations encountered. Section 4 also presents a brief analysis of the NLLJs identified by the RWP from May to September 2017 to 2021, revealing insights into their morphological characteristics. Finally, Section 5 synthesizes the study's key findings and discusses their implications for the understanding of the Mid-Atlantic NLLJ and similar atmospheric phenomena, proposing directions for future research to enhance model accuracy and expand the scope of study within the field of atmospheric science.

2 Observations

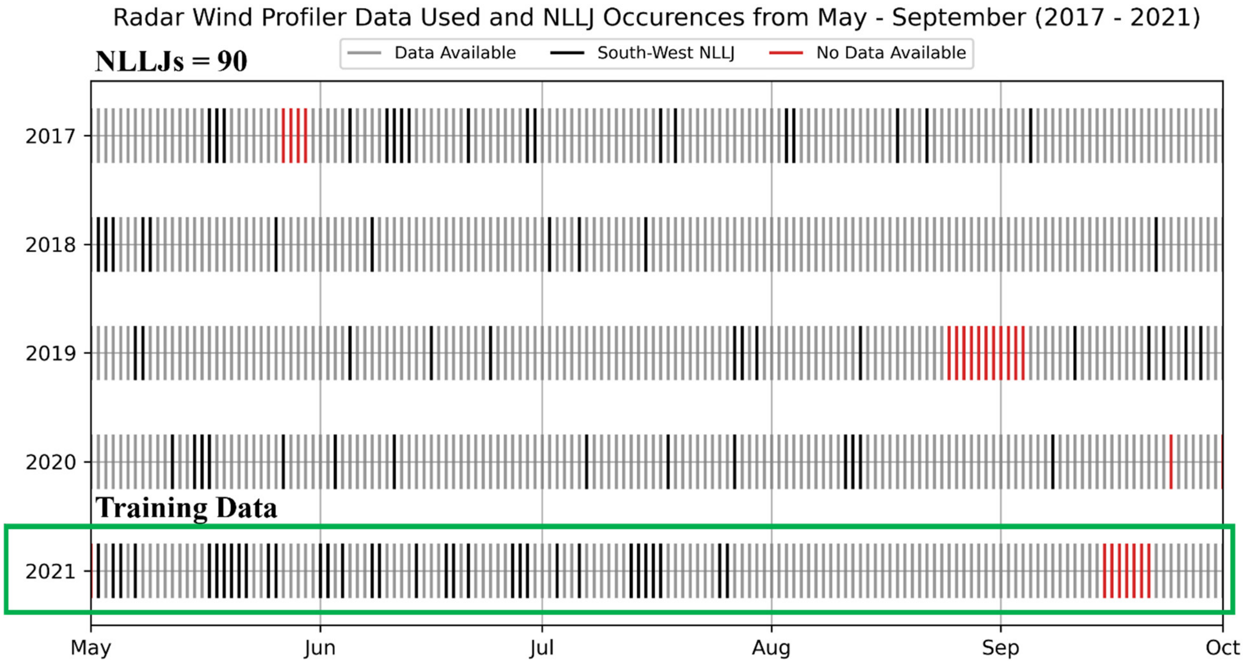

Figure 2: Occurrences of NLLJs (black lines) captured by algorithm from the Beltsville, MD RWP. (red lines) Data that were not available in this study while (grey lines) were available. The green box denotes the year where the training dataset originates.

This study solely employs the 2017 – 2021 dataset of continuous daily wind profiles from the Howard University – Beltsville Campus (HUBC) RWP stationed in Beltsville, MD (referred to as BELT); see Figure 1 for location reference. These RWP instruments measure the radial velocity of wind from one zenith and four azimuthal beams at 915 MHz. These are used to calculate the horizontal speed and direction with sub-100-meter vertical resolution (100 m – 3000 m AGL) at a sub-30-minute temporal resolution. Thus, the resolution of the dataset is sufficient to capture the temporal and vertical extent of the NLLJ events occurring at the HUBC site. The dataset from BELT is visually depicted in Figure 2, where an events plot depicts the temporal distribution of data availability. The grey lines indicate the areas where the BELT daily file was available from the MDE record, while the red lines indicate days that are unavailable because of instrument failure or scheduled maintenance. Only 25 files were unavailable during the study period of May – September of 2017 – 20121, making for an optimal observational dataset to analyze the Summertime Mid-Atlantic NLLJ. The HUBC site is located in between the U.S. Appalachian Mountains and the Chesapeake Bay and then the Atlantic Ocean. The mountainous region is about 200 km east with a peak elevation of about 2 km. HUBC resides in the Piedmont geographic region which spans the space between the mountainous terrain to the west and the coastal plains to the east. This zone creates two boundaries for a latitudinal flow regime with orographic to the west and thermal to the east.

This strategic selection of the training period ensures that the model is exposed to a wide array of conditions typical of the NLLJ season, thereby improving its ability to generalize and detect NLLJs accurately. However, the presence of data gaps, as indicated by the red vertical bars, may present challenges to the comprehensive characterization of NLLJs, as such gaps could potentially coincide with periods of NLLJ events. While unfortunate, the marked seasons and days without data serve as an important reminder of the limitations inherent in observational datasets. Future work should also consider incorporating additional datasets, potentially using data synergy techniques to fill the gaps and provide a more continuous and comprehensive picture of the NLLJ occurrences and extent using all available observational datasets in our study region (e.g., aerosol and ozone lidars, sondes, ground-based spectrometers, and radiometers). Moreover, based on the surrounding temporal and environmental context, the algorithm could be trained to predict the likelihood of NLLJ events on days when data are missing. This predictive ability would be invaluable for atmospheric research and could significantly enhance our capacity to anticipate and respond to the implications of NLLJ events on weather patterns and climate dynamics in the Mid-Atlantic region.

3 Nocturnal Low-Level Jet Isolation

Several previous works have been published regarding identifying low-level jets in wind profiles. These methods have employed peak detection of wind speed maximums in single profiles with threshold criteria on coherent height, speed, direction, and duration. These methods are robust in their objective of identifying continuous low-level wind maxima (De Jong et al., 2024; Tuononen et al., 2017; Baas et al., 2009). Our overall goal is the complete a fully automated system to be used on the network of wind profiles that is adept at identifying, classifying, and characterizing, low-level wind maxima and thus we report our exploration of supervised machine learning for this task. The conceptual model of the detection method presented here relies on single measured points in vertical and temporal space that with the multiple dimensions of the dataset [wind speed (SPD), wind direction (DIR), radial velocity (RAD 1-5), and signal-to-noise ratio (SNR 1- 5)].

3.1 Training Dataset

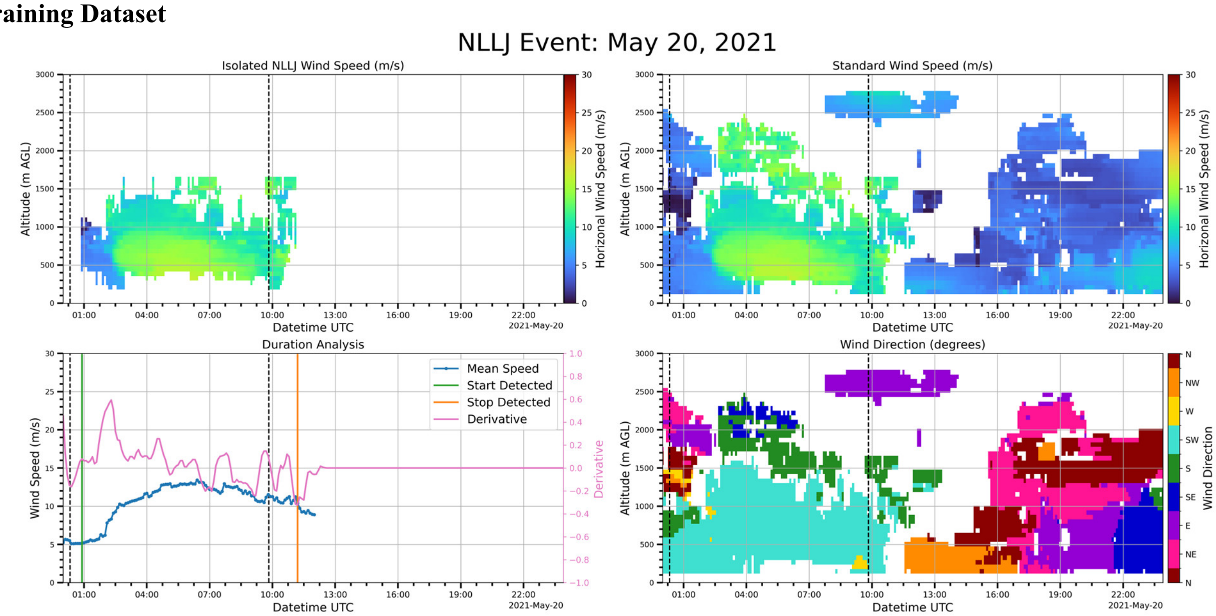

Figure 3: Occurrences of NLLJs (black lines) captured by algorithm from the Beltsville, MD RWP. (red lines) Data that were not available in this study while (grey lines) were available. The green box denotes the year where the training dataset originates. Dashed vertical lines indicate the sunset and sunrise time, respectively.

The training dataset for this experiment was hand-selected from NLLJ events during 2021, while the validation dataset was selected from previously reported and depicted by Sullivan et al. (2017), Delgado et al. (2015), and Weldegaber (2009) that were captured by the same instrument and station (i.e. BELT RWP). To gather a suitable dataset for machine learning we have compiled scenarios expected in operation (e.g. incomplete daily files, missing data, large-scale weather systems, etc.). A manual and rudimentary isolation method was applied using gradient detection solely on the southerly winds (180 – 270 degrees from North) with maximums greater than 5 m s-1 in both time and altitude to capture the evolution and vertical extent of the NLLJ. This approach is demonstrated in Figure 3, where (A) depicts the final isolated NLLJ events from the speed and direction profile (C and D), and (B) represents the visual representation of the gradient detection in the temporal evolution. This method takes the wind speed evolution averaged from 0 - 2000 m and then interpolated and smoothed. The resulting time series is then used to find the first positive gradient and the last negative gradient, which are taken as the start and end of the NLLJ event. This process is then repeated for the vertical extent using each profile to find the top and bottom at each time step. We found that the manual tuning needed for thresholds on time constrain, continuity, and direction evolution was important for isolating NLLJs, but required attention in many different cases and thus we used the well-isolated cases from this method as a training set for the supervised machine learning ensemble. The training set is comprised of 50 NLLJ events that were sufficiently isolated and 50 events that contained no low-level wind maxima that contain low-level wind maxima that we do not consider as LLJ relevant to this study for reasons of direction, or evolution.

3.2 Algorithm Development

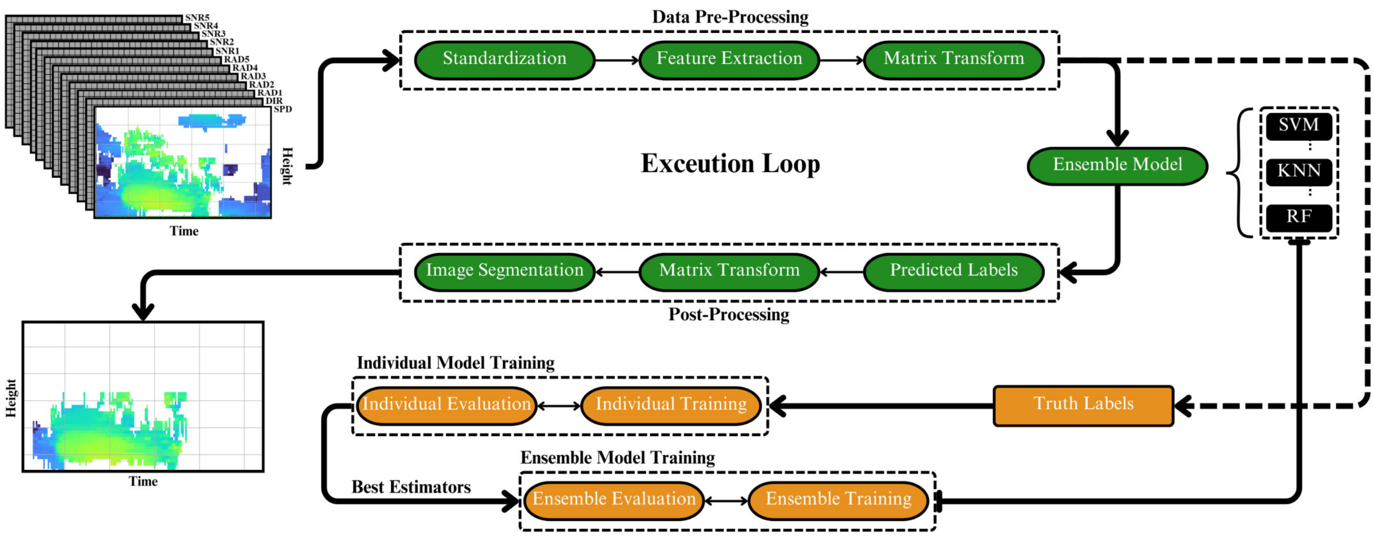

Figure 4: Schematic of the supervised machine learning algorithm execution (Execution Loop, left) and training (right).

The flowchart schema shown in Figure 4 illustrates the process used to execute (green) and train (orange) the supervised machine learning algorithm for detecting NLLJ events in vertically resolved wind profiles. The process begins with data pre-processing, where daily files from the RWP instrument are submitted and then standardized in both height and vertical resolution to ensure the uniformity of profiles. This was accomplished by re-gridding for time and altitude, coupled with the strategic filling of missing data points with 'not-a-number' (NaN) placeholders, thereby ensuring that the dataset maintained uniform dimensions across all the datasets. Following this, feature extraction is performed, extracting the critical variables (i.e., wind speed, wind direction, radial velocities, the averaged signal-to-noise ratios, height, and time). These were determined These parameters (or features) are then transformed into a single matrix where the columns indicate the features and rows indicate the indexes of each variable at a given time and height, in turn, creating a structured dataset ready for input into the machine learning model. The output of the model is then matched to the input matrix as predicted labels, which then undergo the reverse matrix transform from the data pre-processing and an image segmentation process is applied to return the largest cluster of identified points as an NLLJ event. The image segmentation process will not be employed in the results of this work to show its importance and demonstrate the shortcomings of this approach and how it can propagate through to the analysis.

Central to the detection process (execution loop) is the ensemble model, which integrates multiple supervised machine learning algorithms – specifically Support Vector Machine (SVM), k-nearest Neighbours (KNN), and Radom Forest (RF) that are available and open-source in Python from Sci-kit Learn package (Pedregosa et al., 2011). Each model in this ensemble contributes uniquely to the overall predictive capability by leveraging different mathematical principles. The SVM works by identifying a hyperplane in high-dimensional space that best separates the data points of different classes to maximize the distance between the hyperplane and the nearest data points from each class, known as support vectors (Cortes and Vapnik, 1995). The KNN operates on a different principle, classifying a data point based on the majority class among the “k” nearest neighbours in the feature space, with the Euclidean distance used as the metric for determining proximity, where the algorithm then assigns the class most common among the nearest neighbours (Cover and Hart, 1967). The RF model is itself an ensemble of decision trees, each trained on a randomly selected subset of the dataset, in terms of samples and features. Each decision tree in the random forest individually classifies the data by making splits based on criteria such as information gain, entropy, and minimization of Gini impurity. The final classification by the RF is determined by the majority vote of the decision trees, ensuring that the model captures a broad array of patterns in the data (Breiman, 2001). By integrating the insights gained from each model’s approach into a two-thirds majority voting system we have found that this approach yields a suitable method of isolating NLLJ features in wind profiles.

4 Results & Discussion

The results of our testing and subsequent analysis discussion, detail the algorithm's efficacy and limitations. We illustrate its proficient detection of south-westerly NLLJ events while also acknowledging deficiencies in capturing the complete structural nuances of these atmospheric features. The findings underscore the necessity for future model refinement, with a specific focus on training set optimization and the importance of implementing image segmentation techniques to improve representation of model outputs. Furthermore, we present a preliminary analysis of the morphology and statistical attributes of the Mid-Atlantic NLLJs isolated by our methods. We utilize the Beltsville, MD RWP data to introduce a general representation of the Mid-Atlantic NLLJ, derived from a composite analysis of wind profiles, to serve as a foundational tool for our future research and that of others studying this phenomenon.

As previously shown in Figure 2, we have identified 90 warm-season (May- September) NLLJ events using the Beltsville, MD RWP dataset from 2017 – 2021. Zhang et al. (2006) and Ryan, (2004) established much of the Mid-Atlantic NLLJ frequency analysis which they based on wind profile observations with an older version of the RWP instrument used in this study stationed at Fort Meade, MD (~10 km from Beltsville, MD). They established statistics that defined the south-westerly Mid-Atlantic NLLJ as a predominantly summertime nocturnal boundary layer phenomenon. As previously shown in Figure 2, we have identified 90 warm-season (May – September) NLLJ events using the Beltsville, MD RWP datasets over a 5-year period (2017 - 2021), where Ryan (2004) reported 80 summer-season events over 5 years (1998-2002).

4.1 Algorithm Evaluation

The process of evaluating the performance of the algorithm is complex due to the absence of an absolute ground truth for NLLJ detection. The training dataset, or truth labels, represents our best attempt at programmatically isolating NLLJs, yet this process is challenging because it lacks a definitive standard for what constitutes a true NLLJ activity located to every measurement point of the wind profile datasets. To address this complexity, the evaluation approach involves two critical stages following the training phase. The first stage entails comparing the algorithm's results against the gradient method (see section 3.1), which serves as a quantitative benchmark. The second stage involves a qualitative visual inspection by a trained observer, providing an additional layer of evaluation that helps mitigate the challenges posed by the absence of a standardized metric for NLLJ detection. The flowchart from Figure 4 (orange) provides a visual representation of the training process, beginning with the selection of the best estimators. This selection process involves a search routine for hyperparameters that score the highest in each model, see Figure 5 for results of the best estimators. Once the best estimators are identified, they are combined into an ensemble model, which is then fully trained on the remaining unseen portion of the training dataset and evaluated against the supplied truth labels. The final stage involves applying this trained model to previously reported and depicted NLLJs from previous research studies, such as those by conducted by Sullivan et al. (2017), Delgado et al. (2015), and Weldegaber (2009), all of which used the same instrumentation in the same study area.

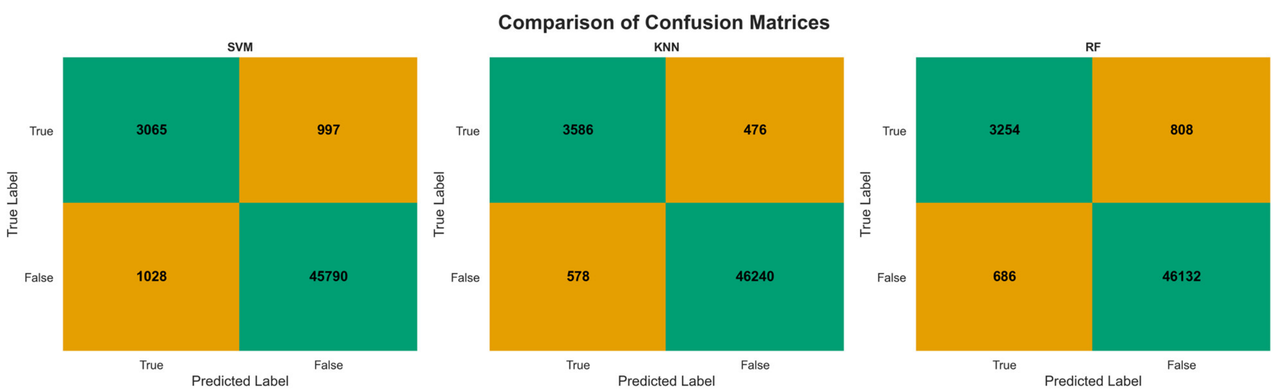

Figure 5: Model training confusion matrices: (green) shows correctly labelled points by the algorithms while (orange) shows incorrectly labelled. Top-left indicates true-positive labels, bottom-right indicates true-negative; while bottom-left indicated false-positive, and top-right indicates false-negative.

Figure 5 illustrates a comparison of confusion matrices from training the machine learning algorithms, where the individual algorithm matrices show results from ~16 daily files from the training dataset, and the ensemble matrix shows results from ~64 daily files. Each confusion matrix provides a breakdown of the model's performance by showing the counts of true negatives (top left quadrant: green), true positives (bottom right quadrant; green), false positives (top right quadrant; orange), and false negatives (bottom left quadrant: orange). Overall, each shows a strong ability to correctly determine non-NLLJ activity, relatively balanced performance on false-negatives and false-positives, and relatively consistent abilities in correct prediction of true-positives. We attribute these results to the implementation of rigorous cross-validation and tuning of hyperparameters in a two-stage process (see Figure 4 individual and ensemble model training). This was done to address the spareness of NLLJ features in the training dataset. Note that these scores are based on the truth-labels supplied, which themselves are imperfect isolations of NLLJ features (see Figure 2). With this in mind, we reserved our final judgment for visual inspection of performance with Mid-Atlantic NLLJs depicted in previous literature for validation of testing results.

Figure 6 illustrates the results of this inspection from the observations of BELT on June 12, 2015 (Sullivan et al., 2017), August 03, 2007, and June 12, 2008 (Delgado et al., 2015; Weldegaber, 2009) The figure is organized into three key panels for each event: isolated NLLJ activity (panels A1, B1, C1), horizontal wind speed (panels A2, B2, C2), and wind direction (panels A3, B3, C3). These panels plot the data against altitude and time, with sunrise and sunset times indicated by dashed vertical lines. In panels A1, B1, and C1, the algorithm effectively demonstrates its capability to detect the characteristic wind speed maximums and corresponding wind direction shifts indicative of NLLJ events. These panels highlight the algorithm's proficiency in identifying the vertical structure and temporal evolution of these jets, capturing key phases such as the onset, peak, and dissipation of the NLLJ events. However, as noted by the circles and dashed boxes in Figure 1, the algorithm does have certain limitations.

On June 12, 2015, the wind speed shows a well-defined NLLJ structure, with wind speeds peaking in the early morning. The corresponding wind direction data for the same date transitions from southerly in the early hours to westerly later in the day, a directional shift that the model effectively captures. The capability of the model to detect these patterns is crucial. It suggests that the model can identify the presence of an NLLJ and its evolution over time. On June 14, 2008, the initial wind speed data might suggest that the NLLJ event concluded by 15 UTC, however, the wind direction gradients reveal the transition to a westerly-dominated regime, thus indicating the end of the NLLJ event at 12 UTC. The wind direction data from this date indicate a significant shift, with winds starting from a southerly direction and transitioning to a westerly direction as the day progresses. This shift to westerly-dominated high wind speeds indicates the kind of directional change the model can discern, echoing the findings of Rabenhorst et al. (2014) regarding the transitional phases of Appalachian Downslope Winds.

The circles indicate instances where the algorithm may have falsely identified NLLJ activity, suggesting potential issues of overestimation. This overestimation could be attributed to imperfections in the training dataset, which may cause the model to be overly sensitive to features that do not necessarily correspond to genuine NLLJ events. On the other hand, the dashed boxes highlight regions where the algorithm struggles to accurately identify NLLJ activity. For example, in the events observed on June 12, 2015, and August 03, 2007, the algorithm appears to have difficulty accurately representing the lower boundary of the NLLJ. This difficulty is likely due to a less pronounced NLLJ signal in the vertical profile, making it challenging for the model to distinguish the NLLJ from surrounding atmospheric conditions. The consistent issue of missed lower-level structures across different test cases suggests that modifications to the training set could significantly enhance the model's performance. The presence of outlier points in the visualized data further underscores the necessity for additional post-processing steps, such as image segmentation (Figure 4), to refine the model output and reduce the occurrence of false positives. The dashed box in the June 14, 2008, observation marks an area where the core of the NLLJ was not identified by the algorithm. This oversight seems to result from a sharp increase in wind speed coupled with only a slight directional shift, leading to a case of false negatives. That withstanding, the isolated NLLJ activity aligns well with the analysis discussions from the literature, validating the potential for supervised-machine learning to be proficient in this task of discerning NLLJs from the broader atmospheric flow. The level of accuracy achieved is sufficient to support robust long-term statistical analysis, which is critical for advancing our understanding of NLLJs and their impact on regional atmospheric composition.

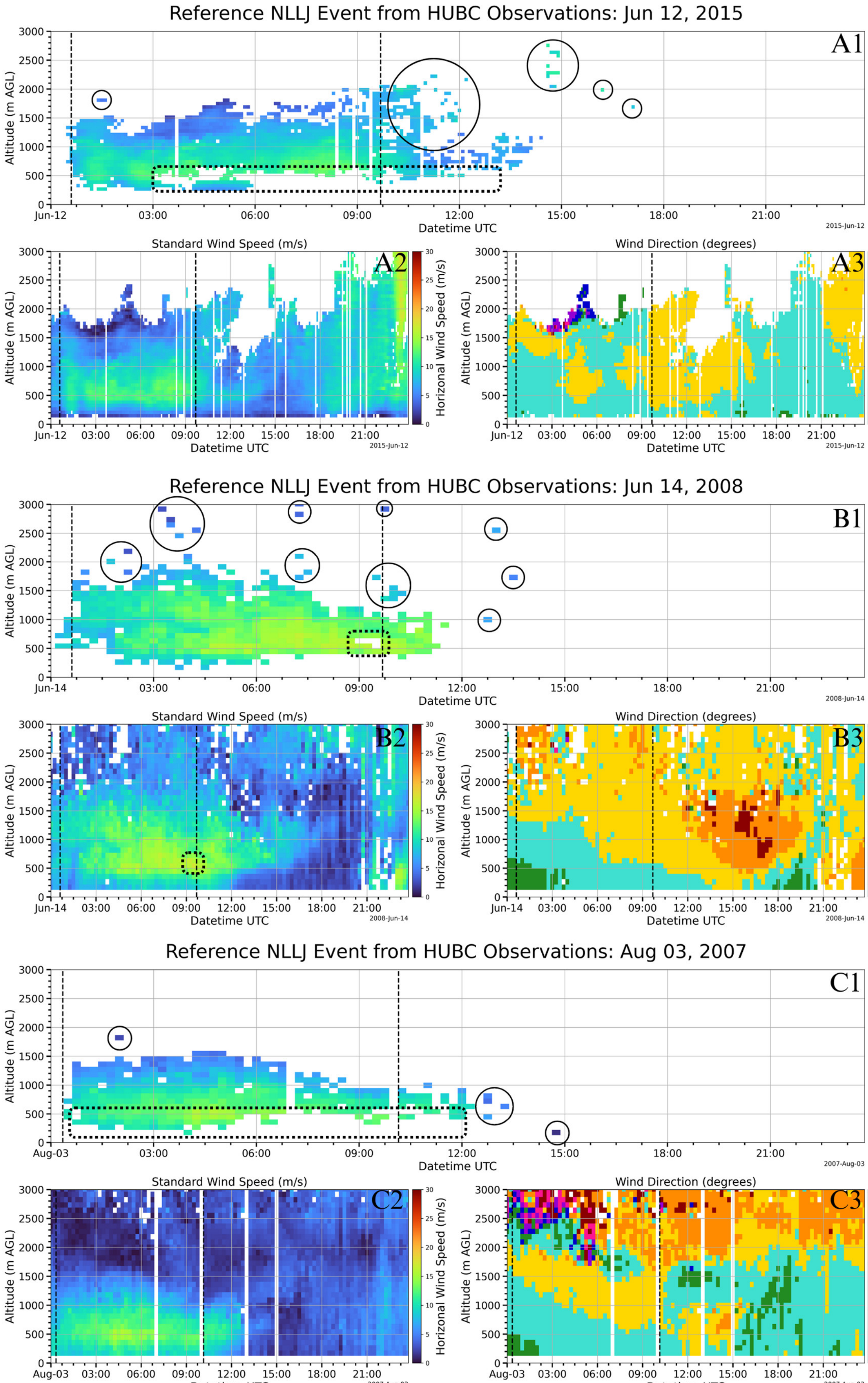

Figure 6: Evaluation of NLLJ isolation algorithm with reference events from literature illustrating the evolution of the NLLJ event reported on (A) June 12, 2015 (Sullivan et al., 2017); August 03, 2007 and Jun 12, 2008 (Delgado et al., 2015; Weldegauber, 2009), where panel 1 shows the isolated NLLJ, panel 2 shows the horizontal wind speed and panel 3 shows the wind direction.

4.2 Statistical Analysis

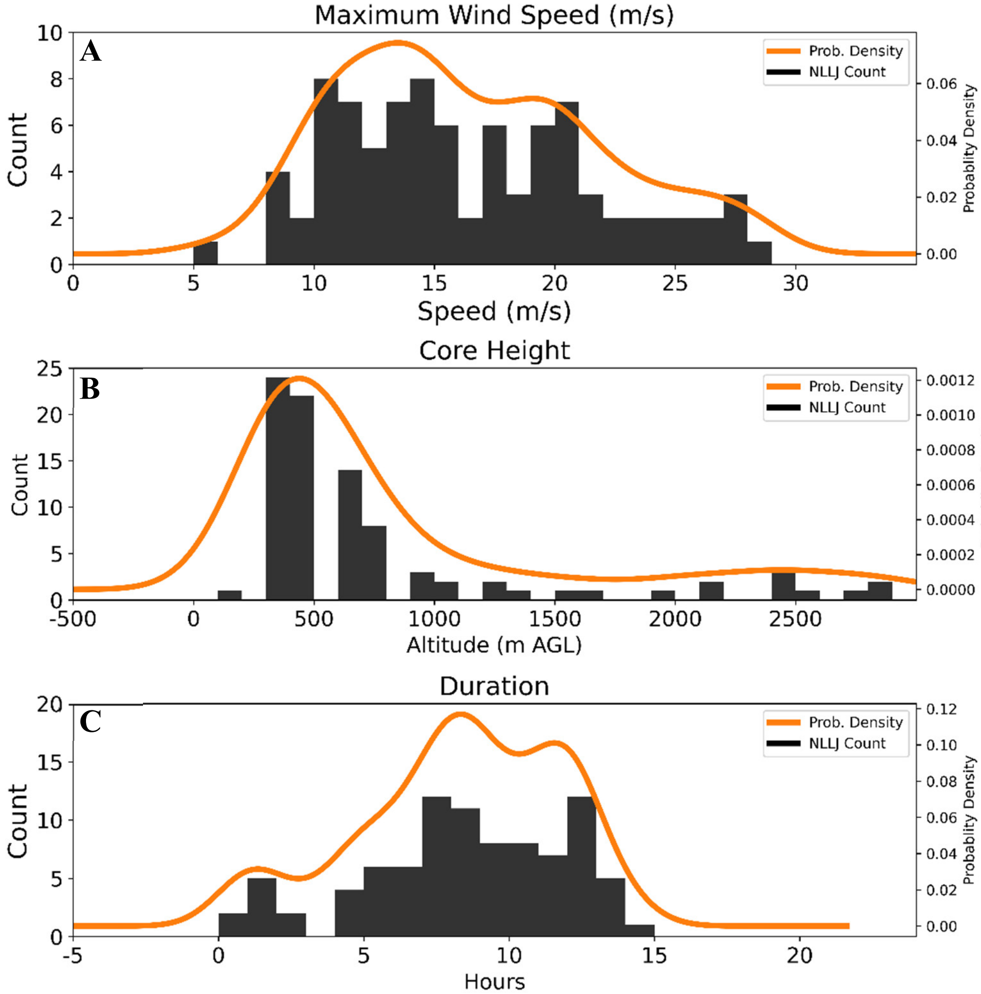

Figure 7: Histogram NLLJ characteristics from the 90 events noted in Figure 2, where: (A) is the distribution of maximum wind speeds, (B) is the height of the wind speed maximum and (C) is the duration of the event at the core height.

To investigate the critical characteristics of the Mid-Atlantic South-Westerly NLLJ, we performed a preliminary statistical analysis of the 90 NLLJ events we have identified. The histograms, shown in Figure 7, provide a statistical representation of maximum wind speeds, core heights, and core time for each NLLJ event, as derived from the dataset of 90 NLLJ events shown in Figure 7. Figure 7 (A) shows the distribution of maximum wind speed of each jet event, spanning 5 m s-1 to nearly 30 m s-1. The wind speed maximum probability density curve (orange) suggests that the most probable core speeds are between 10 m s-1 and 15 m s-1, with decreasing probability between 15 and 20 m s-1 and the least probable being between 20 and 30 m s-1. This would suggest that in the most probable range exists a similar formation mechanism.

The core height (Figure7: B) illustrates the height at which the maximum wind speed was measured for each NLLJ event capture. We notice that most of the maximum wind speeds occur around 500 m AGL, showing a sharp peak at this altitude range, which can be interpreted as the typical altitude for the core. The narrowness of this peak in the probability density curve implies a strong consensus for this characteristic height, aligning with the notion that the jets are confined to the edge of the stable nocturnal boundary layer. Note the low probability and frequency of core height being above 1000 m AGL; we attribute the presence of these altitudes as removable noise from the isolation due to the imperfect nature of our training dataset (see section 3.1).

When considering the histogram for the duration of the NLLJ (Figure7: C), we encounter a more complex distribution. The duration of the NLLJ event is calculated by finding the elapsed time from each jets core height. The histogram and probability distribution suggest a multi-modal distribution, with two apparent peaks around 8 and 12 hours. This multi-modal nature may hint at the additional influencing factors, such as the baroclinicity of the region, which could induce variations in the timing of the jet's maximum wind speeds.

4.3 Mid-Atlantic NLLJ Morphology

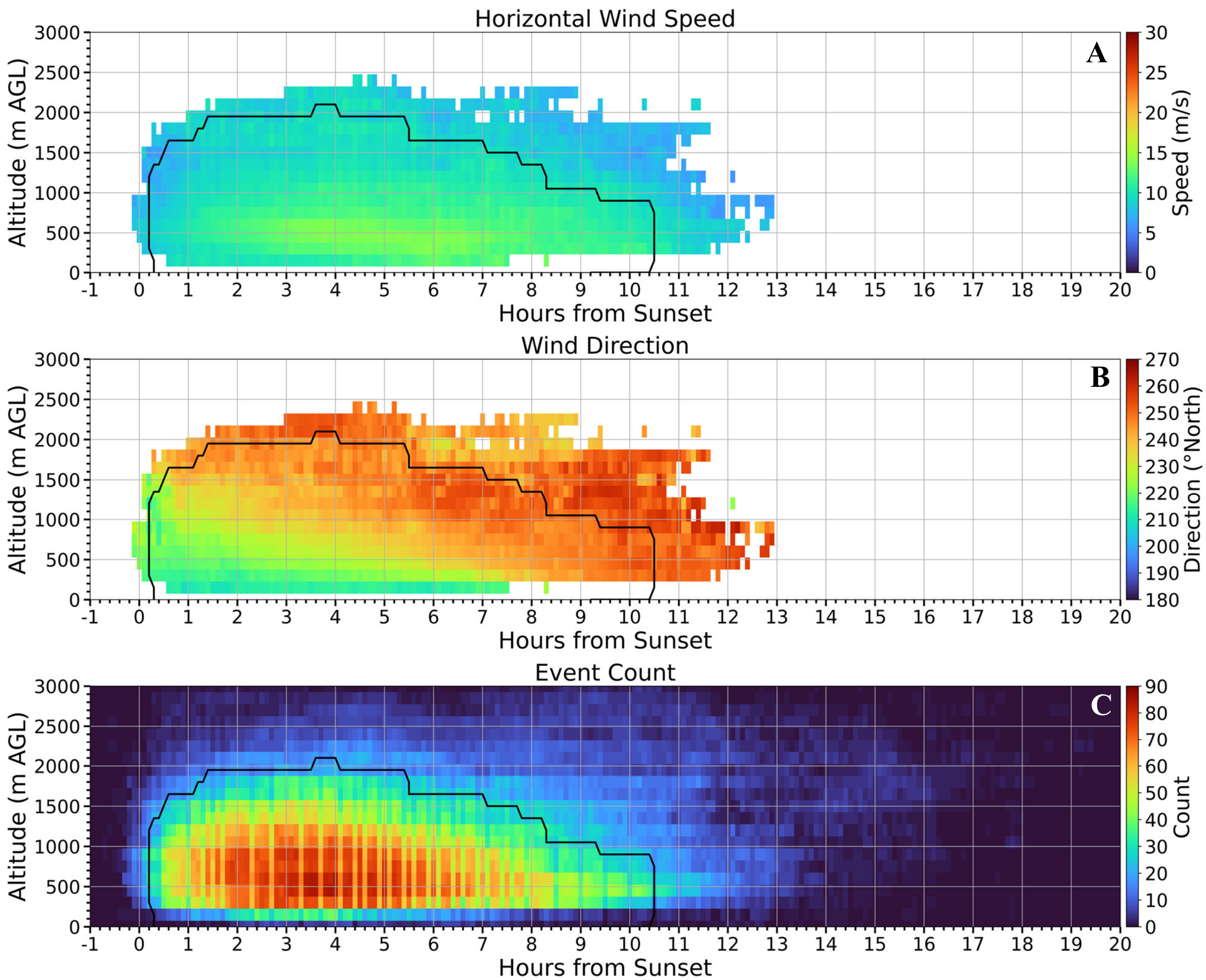

Figure 8: ***Composite Vertical Profiles of Nocturnal Low-Level Jet (NLLJ) Characteristics: (A) Average Wind Speed (B) Average Wind Direction, and (C) Event Count.***

As a result of this isolation of NLLJ occurrence in wind profiles, we have created a general representation of the Mid-Atlantic NLLJ using observations. Figure 8 shows a composite plot of the NLLJ structure using the median of the 90 NLLJ datasets to visualize the temporal evolution as seen from the observations at Beltsville, MD. This offers a basis for future identification and analysis of the general south-westerly NLLJ. The region enclosed by the black line indicates the region in which more than 50% of the cases were present (see Figure 8: C), thus serving as the general structure of the Mid-Atlantic NLLJ. Outside the enclosed region is the variability of NLLJs found in our 90 events datasets. Figure 8 (A) describes the horizontal wind speed, which exhibits a shallow layer of high wind speed (~ 15 m s-1) concentrated around 500 m AGL; this is noted as the NLLJ's core. This general NLLJ structure lasts from just after sunset (0 hours) to almost 11 hours after sunset. The vertical extent is shown to be persistent around 1500 m AGL until the arrival of the horizontal wind speed maximum (NLLJ core) around 4 hours after sunset, at which point the vertical structure of the NLLJ begins to decay.

The panel B of Figure 8, which illustrates wind direction, shows a clear transition between the dominance of the meridional (South to North; Southerly) and zonal (West to East; Westerly) winds and is indicative of the dynamic atmospheric processes that govern the behavior of the Mid-Atlantic NLLJ. The progression to more westerly winds across the night reflects the diurnal wind shift and underscores the influence of large-scale atmospheric circulation patterns on the NLLJ. This shift in wind direction is often related to the Coriolis force acting on the regional air mass over the night. As the land cools after sunset, the pressure gradients adjust, and the NLLJ develops, initially following the temperature gradient. As the night progresses, the Coriolis force begins to turn the flow toward the right in the northern hemisphere, resulting in the NLLJ acquiring a more westerly component.

The vertical dependence of the oscillation between wind vectors, as observed in panel B of Figure 8, indeed underscores the manifestation of inertial oscillation theory in the behaviour of the Mid-Atlantic NLLJ. This oscillation between wind vectors at different altitudes signifies the vertical shear, which is characteristic of the NLLJ structure. The presence of wind shear is significant for various atmospheric processes, such as the development of turbulence, the dispersion of aerosols, and the vertical transport of momentum and heat within the atmosphere; notably, strong wind shear associated with NLLJs can induce turbulent downbursts, thereby affecting aviation safety, efficiency of wind energy generation, and surface level air quality. The works of Roots et al. (2023) and Sullivan et al. (2017) both noted the increase of surface-level ozone from a polluted ozone reservoir in the residual layer during the arrival of the NLLJ core, which, as shown in Figure 8, is the maximum point of the horizontal speed and balance between the zonal and meridional wind vectors.

Collectively, these panels deliver a cohesive understanding of the NLLJ's vertical and temporal structure. They demonstrate a pronounced nocturnal intensification in wind speed at low-level altitudes, accompanied by a veering wind direction, which indicates the inertial oscillation's influence on the jet's formation. Furthermore, the event count substantiates the observed morphology, confirming that the algorithm effectively captures the climatological presence of the NLLJ in the dataset. The variability outside the core zone may be attributed to synoptic-scale influences that modulate NLLJ behavior. Understanding this variability is essential for improving weather prediction models, particularly for events sensitive to low-level jet dynamics.

5 Conclusions

The study successfully applied machine learning (ML) algorithms to detect and characterize Nocturnal Low-Level Jets (NLLJs) using Radar Wind Profiler (RWP) data.

- This study has investigated the usage of supervised machine-learning for detection of NLLJ events in the Mid-Atlantic region which was quantitatively and qualitatively evaluated and shown its potential for useful atmospheric analysis.
- Statistical analysis of 90 NLLJ events show key patterns in wind speed, core height, and event duration that agree with previous literature on the theory and case studies of NLLJs in the Mid-Atlantic.
- The morphology composite (Figure 8) provides a clear visualization of the vertical and temporal structure of NLLJs in the Mid-Atlantic region, offering a useful reference for understanding the typical characteristics and evolution.

Further work is needed to attribute whether the interannual variability observed, within specific years with more frequency of events, underscores the connection between NLLJ occurrences and broader synoptic patterns that drive regional dynamics. Identifying these patterns has profound implications for the future study of NLLJs, necessitating an extended analysis over a longer period and across a broader observational network. This would be critical for comprehensively understanding the atmospheric forces at play in NLLJ formation and thus when implemented into atmospheric models, enhancing the accuracy.

The presence of data gaps within the observational period presents a challenge yet serves as a crucial reminder of the inherent limitations of empirical datasets. Addressing these gaps by incorporating additional datasets and employing data synergy techniques will be imperative in future studies to create a more continuous and complete picture of NLLJ occurrences. As presented in Figure 7, the statistical analysis has yielded a quantitative representation of the critical characteristics defining the Mid-Atlantic NLLJ. The distribution of maximum wind speeds and core heights clearly indicates a similar atmospheric mechanism driving these jets, with most NLLJ cores residing around 500 m AGL. The bi-modal nature of the core timing distribution suggests additional factors, such as regional baroclinicity, may influence the timing of NLLJ maxima, indicating deviations from classical inertial oscillation predictions. These findings are pivotal for enhancing our understanding of NLLJ behaviors and refining their representation in atmospheric models. The morphological analysis of the NLLJ, informed by the composite plots in Figure 7, has established a general representation of the Mid-Atlantic NLLJ's structure. This representation, highlighting a pronounced nocturnal intensification of wind speed and a distinct oscillation pattern in wind direction, is instrumental in defining the typical NLLJ profile. The observed variability outside the core zone highlights the influence of large-scale atmospheric processes on NLLJ behavior, emphasizing the need for further investigation into these external factors.

The continued study of these low-level wind phenomena is crucial for weather prediction and the strategic management of air quality, particularly in understanding the transport and dispersion of pollutants along the Mid-Atlantic region of the US. This research serves as a stepping stone for future investigations into the complex dynamics of low-level mesoscale phenomena and their broader climatic and environmental implications. Accurately identifying and characterizing NLLJs is vital for validating and refining regional climate models, thus enhancing our predictive capabilities regarding future climate scenarios. This work contributes significantly to boundary layer studies, offering a detailed examination of NLLJ phenomena and providing valuable insights for future research. This study's integration of machine learning within the field of atmospheric science marks a promising step toward more advanced meteorological analyses and the future of climate prediction. Future studies should integrate similar techniques to understand the genesis mechanisms of low-level wind phenomena from large observational datasets to enhance our understanding of regional pollutant distribution and mesoscale transport of moisture, momentum, and mass. Furthermore, there is a clear opportunity to expand this research to encompass other geographic regions and atmospheric phenomena, thereby testing the adaptability and versatility of the methodology.

Acknowledgments

The authors would like to acknowledge the continued support, guidance, and supply of datasets from the Maryland Department of the Environment (MDE), namely James Boyle. The authors gratefully acknowledge the support provided by the NASA Tropospheric Composition Program, including the NASA Tropospheric Ozone Lidar Network (TOLNet). The work of Mr. Maurice Roots was supported by the NASA FINESST Program (Future Investigators in National Aeronautics & Space Administration (NASA) Earth and Space Science and Technology) as a 2022 recipient **[Award #: 80NSSC22K1453]** and NASA

Goddard Space Flight Center (GSFC) and the GEM consortium for partly funding this work and coordinating this collaboration as a 2021 Fellow. *Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the view of NASA.*

Code & Data Availability

Radar wind profiler datasets are available from the Maryland Department of the Environment upon request. The algorithm development, data processing, and analysis codes are currently available upon request, as they are in their beginning stages. However, version 1.0 will be easily usable, installable, and open-source through the corresponding authors' Git Hub.

Author Contribution

Maurice Roots: Conceptualization, Methodology, Software, Validation, Formal Analysis, Investigation, Resources, Data curation, Writing – original draft. **John T. Sullivan:** Writing – review & editing, Supervision. **Belay Demoz:** Project Administration, Funding Acquisition.

Competing Interests

Some authors are members of the editorial board of Atmospheric Measurement Techniques – Gases.

References

[1] Baas, P., Bosveld, F. C., Klein Baltink, H., and Holtslag, A. A. M.: A Climatology of Nocturnal Low-Level Jets at Cabauw, J. Appl. Meteorol. Climatol., 48, 1627–1642, https://doi.org/10.1175/2009JAMC1965.1, 2009.

[2] Banta, R. M.: Stable-boundary-layer regimes from the perspective of the low-level jet, Acta Geophys., 56, 58–87, https://doi.org/10.2478/s11600-007-0049-8, 2008.

[3] Banta, R. M., Pichugina, Y. L., and Newsom, R. K.: Relationship between Low-Level Jet Properties and Turbulence Kinetic Energy in the Nocturnal Stable Boundary Layer, J. Atmospheric Sci., 60, 2549–2555, https://doi.org/10.1175/1520-0469(2003)060<2549:RBLJPA>2.0.CO;2, 2003.

[4] Blackadar, A. K.: Boundary Layer Wind Maxima and Their Significance for the Growth of Nocturnal Inversions, Bull. Am. Meteorol. Soc., 38, 283–290, https://doi.org/10.1175/1520-0477-38.5.283, 1957.

[5] Bonner, W. D.: CLIMATOLOGY OF THE LOW LEVEL JET, Mon. Weather Rev., 96, 833–850, https://doi.org/10.1175/1520-0493(1968)096<0833:COTLLJ>2.0.CO;2, 1968.

[6] Breiman, L.: Random Forests, Mach. Learn., 45, 5–32, https://doi.org/10.1023/A:1010933404324, 2001.

[7] Carroll, B. J., Demoz, B. B., and Delgado, R.: An Overview of Low-Level Jet Winds and Corresponding Mixed Layer Depths During PECAN, J. Geophys. Res. Atmospheres, 124, 9141–9160, https://doi.org/10.1029/2019JD030658, 2019.

[8] Carroll, B. J., Demoz, B. B., Turner, D. D., and Delgado, R.: Lidar Observations of a Mesoscale Moisture Transport Event Impacting Convection and Comparison to Rapid Refresh Model Analysis, Mon. Weather Rev., 149, 463–477, https://doi.org/10.1175/MWR-D-20-0151.1, 2021.

[9] Corsmeier, U., Kalthoff, N., Kolle, O., Kotzian, M., and Fiedler, F.: Ozone concentration jump in the stable nocturnal boundary layer during a LLJ-event, Atmos. Environ., 31, 1977–1989, https://doi.org/10.1016/S1352-2310(96)00358-5, 1997.

[10] Cortes, C. and Vapnik, V.: Support-vector networks, Mach. Learn., 20, 273–297, https://doi.org/10.1007/BF00994018, 1995.

[11] Cover, T. and Hart, P.: Nearest neighbor pattern classification, IEEE Trans. Inf. Theory, 13, 21–27, https://doi.org/10.1109/TIT.1967.1053964, 1967.

[12] De Jong, E., Quon, E., and Yellapantula, S.: Mechanisms of Low-Level Jet Formation in the U.S. Mid-Atlantic Offshore, J. Atmospheric Sci., 81, 31–52, https://doi.org/10.1175/JAS-D-23-0079.1, 2024.

[13] Delgado, R., Rabenhorst, S. D., Demoz, B. B., and Hoff, Raymond. M.: Elastic lidar measurements of summer nocturnal low level jet events over Baltimore, Maryland, J. Atmospheric Chem., 72, 311–333, https://doi.org/10.1007/s10874-013-9277-2, 2015.

[14] Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.: The ERA5 global reanalysis, Q. J. R. Meteorol. Soc., 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

[15] Holton, J. R.: The diurnal boundary layer wind oscillation above sloping terrain, Tellus, 19, 199–205, https://doi.org/10.1111/j.2153-3490.1967.tb01473.x, 1967.

[16] Karipot, A., Leclerc, M. Y., and Zhang, G.: Characteristics of Nocturnal Low-Level Jets Observed in the North Florida Area, Mon. Weather Rev., 137, 2605–2621, https://doi.org/10.1175/2009MWR2705.1, 2009.

[17] Lima, D. C. A., Soares, P. M. M., Semedo, A., and Cardoso, R. M.: A Global View of Coastal Low-Level Wind Jets Using an Ensemble of Reanalyses, J. Clim., 31, 1525–1546, https://doi.org/10.1175/JCLI-D-17-0395.1, 2018.

[18] Lima, D. C. A., Soares, P. M. M., Semedo, A., Cardoso, R. M., Cabos, W., and Sein, D. V.: A Climatological Analysis of the Benguela Coastal Low-Level Jet, J. Geophys. Res. Atmospheres, 124, 3960–3978, https://doi.org/10.1029/2018JD028944, 2019.

[19] Lundquist, J. K.: Intermittent and Elliptical Inertial Oscillations in the Atmospheric Boundary Layer, J. Atmospheric Sci., 60, 2661–2673, https://doi.org/10.1175/1520-0469(2003)060<2661:IAEIOI>2.0.CO;2, 2003.

[20] Mahrt, L.: Stratified Atmospheric Boundary Layers and Breakdown of Models, Theor. Comput. Fluid Dyn., 11, 263–279, https://doi.org/10.1007/s001620050093, 1998.

[21] Ortiz-Amezcua, P., Martínez-Herrera, A., Manninen, A. J., Pentikäinen, P. P., O'Connor, E. J., Guerrero-Rascado, J. L., and Alados-Arboledas, L.: Wind and Turbulence Statistics in the Urban Boundary Layer over a Mountain–Valley System in Granada, Spain, Remote Sens., 14, 2321, https://doi.org/10.3390/rs14102321, 2022.

[22] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., and Cournapeau, D.: Scikit-learn: Machine Learning in Python, Mach. Learn. PYTHON, n.d.

[23] Rabenhorst, S., Whiteman, D. N., Zhang, D.-L., and Demoz, B.: A Case Study of Mid-Atlantic Nocturnal Boundary Layer Events during WAVES 2006, J. Appl. Meteorol. Climatol., 53, 2627–2648, https://doi.org/10.1175/JAMC-D-13-0350.1, 2014.

[24] Ranjha, R., Tjernström, M., Semedo, A., Svensson, G., and Cardoso, R. M.: Structure and variability of the Oman coastal low-level jet, Tellus Dyn. Meteorol. Oceanogr., 67, 25285, https://doi.org/10.3402/tellusa.v67.25285, 2015.

[25] Roots, M., Sullivan, J. T., Delgado, R., Twigg, L., and Demoz, B.: An integrated monitoring system (IMS) for air quality: Observations of a unique ozone-exceedance event in Maryland, Atmos. Environ., 313, 120028, https://doi.org/10.1016/j.atmosenv.2023.120028, 2023.

[26] Ryan, W. F.: The Low Level Jet in Maryland: Profiler Observations and Preliminary Climatology, n.d.

[27] Shapiro, A. and Fedorovich, E.: Analytical description of a nocturnal low-level jet, Q. J. R. Meteorol. Soc., 136, 1255–1262, https://doi.org/10.1002/qj.628, 2010.

[28] Shapiro, A., Fedorovich, E., and Rahimi, S.: A Unified Theory for the Great Plains Nocturnal Low-Level Jet, J. Atmospheric Sci., 73, 3037–3057, https://doi.org/10.1175/JAS-D-15-0307.1, 2016.

[29] Stensrud, D. J.: Importance of Low-Level Jets to Climate: A Review, J. Clim., 9, 1698–1711, https://doi.org/10.1175/1520-0442(1996)009<1698:IOLLJT>2.0.CO;2, 1996.

[30] Sullivan, J. T.: Lidar observations revealing transport of O3 in the presence of a nocturnal low-level jet: Regional implications for ``next-day'' pollution, Atmos. Environ., 12, 2017.

[31] Sullivan, J. T., Rabenhorst, S. D., Dreessen, J., McGee, T. J., Delgado, R., Twigg, L., and Sumnicht, G.: Lidar observations revealing transport of O3 in the presence of a nocturnal low-level jet: Regional implications for "next-day" pollution, Atmos. Environ., 158, 160–171, https://doi.org/10.1016/j.atmosenv.2017.03.039, 2017.

[32] Tollerud, E. I., Caracena, F., Koch, S. E., Jamison, B. D., Hardesty, R. M., McCarty, B. J., Kiemle, C., Collander, R. S., Bartels, D. L., Albers, S., Shaw, B., Birkenheuer, D. L., and Brewer, W. A.: Mesoscale Moisture Transport by the Low-Level Jet during the IHOP Field Experiment, Mon. Weather Rev., 136, 3781–3795, https://doi.org/10.1175/2008MWR2421.1, 2008.

[33] Tuononen, M., O'Connor, E. J., Sinclair, V. A., and Vakkari, V.: Low-Level Jets over Utö, Finland, Based on Doppler Lidar Observations, J. Appl. Meteorol. Climatol., 56, 2577–2594, https://doi.org/10.1175/JAMC-D-16-0411.1, 2017.

[34] Weaver, S. J. and Nigam, S.: Variability of the Great Plains Low-Level Jet: Large-Scale Circulation Context and Hydroclimate Impacts, J. Clim., 21, 1532–1551, https://doi.org/10.1175/2007JCLI1586.1, 2008.

[35] Wei, W., Zhang, H., Zhang, X., and Che, H.: Low-level jets and their implications on air pollution: A review, Front. Environ. Sci., 10, 1082623, https://doi.org/10.3389/fenvs.2022.1082623, 2023.

[36] Weldegaber, M. H.: Investigation of stable and unstable boundary layer phenomena using observations and a numerical weather prediction model, Thesis (Ph. D.)--University of Maryland, Baltimore County, 2009., 2009.

[37] Whiteman, C. D., Bian, X., and Zhong, S.: Low-Level Jet Climatology from Enhanced Rawinsonde Observations at a Site in the Southern Great Plains, J. Appl. Meteorol., 36, 1363–1376, https://doi.org/10.1175/1520-0450(1997)036<1363:LLJCFE>2.0.CO;2, 1997.

[38] Zhang, D.-L., Zhang, S., and Weaver, S. J.: Low-Level Jets over the Mid-Atlantic States: Warm-Season Climatology and a Case Study, J. Appl. Meteorol. Climatol., 45, 194–209, https://doi.org/10.1175/JAM2313.1, 2006.