1   **Response to Reviewers**

We would like to extend our gratitude to the reviewers for their time, examination, and constructive feedback on our manuscript. Their insights and suggestions will significantly improve the quality and clarity of our paper and future work. We are thankful for their dedication and commitment to enhancing the rigor and impact of manuscript. This review process has been instrumental in identifying areas for refinement of analysis and strengthening the conclusions of our study. We hope that the responses that are provided below will suffice in answering your questions and addressing your concerns. Thank you.

2   **See our bulleted responses below:**

3   _____

4   <u>**RC1 (Referee #2):**</u>

5   General comments:

6   The document addresses scientific issues that are relevant to AMT, provided that some elements
7   are clarified. The scientific methods and assumptions seems valid but not clearly outlined. The
8   results are not sufficient to support the interpretations and conclusions. The description of
9   experiments and calculations is not sufficiently complete and precise to allow their reproduction
10   by fellow scientists. The authors give not proper credit to related work but clearly indicate their
11   own contribution. The title clearly reflect the contents of the paper. The abstract provides a concise
12   and complete summary. The overall presentation is structured and clear. The language is fluent and
13   precise. Any parts of the paper (figures) should be clarified. The number and quality of references
14   are not appropriate.

15   • *Thank you for pointing out the need for clarification on our methods and assumptions for*
16   *reproducibility by other researchers. A further discussion of related works is indeed*
17   *relevant and pertinent, and we have increased our literature review in the introduction.*
18   *Our refined results yield sufficient interpretations and conclusions.*

19   This article deals with a technique for identifying NLLJs using machine learning. Has machine
20   learning already been used in this type of identification or is it an innovation? This should be
21   specified and references introduced if not.

22   • *To our knowledge at time of submission and exploration of this work. We did not find any*
23   *NLLJ isolation method that utilized supervised machine learning in wind profiles. Thus,*
24   *this work is an innovation to explore the use of machine learning for this task. To further*
25   *provide context for this effort, please see the revised introduction and section 3.*

26   Automatic identification of NLLJs has already been developed using physics-based algorithms, so
27   why use machine learning when these algorithms have already proved their worth ? A paragraph
28   could be added to discuss these aspects.

29   • *Several previous works have been published regarding the identification of Low-Level Jets*
30   *in wind profiles. These methods have employed peak detection of wind speed maximums in*
31   *single profiles with threshold criteria on coherent height, speed, direction, and duration.*
32   *These methods are robust in their objectives of identifying continuous low-level wind*
33   *maxima. However, for our objective of detecting solely a specific type of low-level wind*
34   *maxima for the Mid-Atlantic region we have explored the use of supervised machine*
35   *learning. Our study region encompasses complex terrain, with mountainous terrain*
36   *towards the west and coastal plains towards the east. Our overall goal is to construct a*
37   *fully automated machine learning algorithm for the network of wind profiles that is adept*
38   *at isolating, classifying, and characterizing, low-level wind regimes and thus we report our*
39   *exploration of supervised machine learning for this task of isolating NLLJs. We see this*
40   *manuscripts as the first logical step on publishing our fundamental methods and we expect*
41   *to entrain other data resources in later versions of this process in subsequent manuscripts.*
42
43   *Contrary to work published by other researchers, the conceptual model of our detection*
44   *method presented here relies on single measured points in vertical and temporal space that*
45   *with the multiple dimensions of the dataset [wind speed (SPD), wind direction (DIR), radial*
46   *velocity (RAD 1-5), and signal-to-noise ratio (SNR 1-5)]. This was attempted in order to*
47   *mask the profiles for solely NLLJ activity, regardless of issues in data gaps in the profile. *

48   Many references do not appear in the bibliography (e.g. Delgado 2013, Weldegauber 2009, Caroll
49   2021, ...) while others present in the bibliography are not cited in the text (e.g. Bonner 1968,
50   Dejong 2024, Doubler 2015, Hu 2013a & b, Karipot 2009, Liang 2018, Lima, 2018 & 2019, ...).
51   Please, review the entire bibliography. As it stands, it is impossible to verify the veracity of
52   everything written.

- *Thank you for pointing this out. This is an oversight on our part from multiple versions of the manuscript. We present an updated references list in the revised manuscript.*

The context is well presented but is repeated many times in the text (for example : NLLJs are noctural events).

- *This too is an oversight on our part resulting from multiple versions of the manuscript. We have removed these unnecessary repetitions in the revised manuscript.*

As far as the figures are concerned, each of the panels should be identified by a letter and the captions should be more detailed. The legends are not complete.

- *We appreciate this comment to improve the clarity of our discussion. All of our panel figures have been updated to include lettering in the figure itself and where mentioned in the text.*

There is not enough detail on how the algorithm works to reproduce it.

- *We have extended the discussion of how the algorithm is developed. We hope that a longer discussion along with the mentioned packages used will make it easier for other researchers to follow along and be able to reproduce it.*

Specific comments:

L.27 "noctural events" already introduced in L. 26.

- *Removed*

L.37-39 Wind speed does not decrease as far as the free troposphere in Figure 2.

- *Clarified – Wind speed decrease with increasing altitude*

L.54 LLJs not defined

- *Corrected – Low-Level Jets introduced in the introduction*

L.61 NLLJ already defined

- *Removed*

L66. MD not defined

- *Corrected – MD introduced in the introduction section as Maryland (MD)*

L.75 This paper is not based on a radar network but on a single radar forming part of a network.

80
81   ● *Clarified – This work uses Howard University – Beltsville Campus site which is apart of the Maryland Department of Environment Network of Radar Wind Profilers*

82 L.78 identified by which radar?

83
84   ● *Clarified – In the Observations section we discuss the measurement site (i.e., Howard University Beltsville – Campus site)*

85 Section 2 contains a single sub-section 2.1 and an excessively long introduction. Please, fix this
86 section.

87
88   ● *Clarified – We have converted Section 2 in a discussion of the Observations and Measurement Site. Section 3 then is solely about the NLLJ isolation Algorithm*

89 L.84 - 85 This paper is not based on a network of radars but on a single radar forming part of a
90 network.

91
92   ● *Clarified – This work uses Howard University – Beltsville Campus site which is a part of the Maryland Department of Environment Network of Radar Wind Profilers*

93 L.85 Is there a reference with the full characteristics of this RWP?

94   ● *Clarified – We have included the full name and model of the radar wind profiler.*

95 L 86. MD not defined

96   ● *Corrected – MD introduced in the introduction section as Maryland (MD)*

97 Figure 1: Is it necessary to indicate the location of the Cambrigde and Cumberland RWPs? If so,
98 highlight the location of the Beltsville RWP.

99   ● *Corrected - It is not required the other site be listed.*

100 L.94 What is the value of the wind speed identified by sonde?

101
102   ● *Clarified – We initial used the location of the wind speed maximum by the sonde to choose the pressure level need for plotting the reanalysis wind contour*

103 L.95 NOAA NCEP not defined.

104   ● *Removed – We will use ERA5 Reanalysis instead and include a definition*

105 L.95 Unlikely date and please add time.

106   ● *Corrected – This is a type and should say May 20, 2021*

107 L.96 Wind speed measured by sonde not known. The measured wind profile could be added to
108 figure 1 in order to clearly see the jet.

109   ● *Removed – We now use ERA5 Reanalysis*

110 L.99-110 This part should be included in the previous one, some things are repeated. It is not about
111 data or method.

112   ● *Corrected – We have converted Section 2 in a discussion of the Observations and*
113   *Measurement Site. Section 3 then is solely about the NLLJ isolation Algorithm*

114 L.112 Figure 5 is too far from this paragraph.

115   ● *Corrected – We have converted Section 2 in a discussion of the Observations and*
116   *Measurement Site. Figure 5 will now be Figure 2 located in Section 2 to supplement the*
117   *discussion of the wind profile observations dataset, and just before the discussion of the*
118   *training dataset. Section 3 then is solely about the NLLJ isolation Algorithm (e.g. training,*
119   *development, etc.)*

120 L.116 Why are some data not available?

121   ● *Clarified - The grey lines indicate the areas where the BELT daily file was available from*
122   *the MDE record, while the red lines indicate days that are unavailable because of*
123   *instrument failure or scheduled maintenance.*

124 L.117 Are we to understand that these are daily files? This is not specified.

125   ● *Clarified - The grey lines indicate the areas where the BELT daily file was available from*
126   *the MDE record, while the red lines indicate days that are unavailable because of*
127   *instrument failure or scheduled maintenance.*

128 L.120 Only one sub-section follows, not several.

129   ● *Removed*

130 Figure 2(B) SPD not defined, present the curves in the legend to Figure 2. The figure should be
131 centred on the NLLJ event in order to show its development clearly (from 20:00 UTC on 19 May,
132 for example).

133   ● *Clarified - [wind speed (SPD), wind direction (DIR), radial velocity (RAD 1-5), and signal-*
134   *to-noise ratio (SNR 1- 5)].*

135 L.123-124 Insert a reference.

136   ● *Clarified – To gather a suitable dataset for machine learning we have compiled scenarios*
137   *expected in operation (i.e. incomplete daily files, missing data, large scale weather*
138   *systems, etc.).*

139   L.130 What time does night begin?

140   ●   *Clarified – Sunrise and Sunset time have been added to our figures as vertical dashed lines.*

141   L.133 -134 repetition of encountered

142   ●   *Removed*

143   L.137 what does a file represent? The number of columns and rows is irrelevant.

144   ●   *Clarified - These RWP instruments measure the radial velocity of wind from one zenith and*
145   *four azimuthal beams at 915 MHz. These are used to calculate the horizontal speed and*
146   *direction with sub-100-meter vertical resolution (100 m – 3000 m AGL) at a sub-30-minute*
147   *temporal resolution. Each file contains a continuous daily profile of [Wind Speed (SPD),*
148   *Wind Direction (DIR), Radial Velocity (RAD 1-5), Number of Profiles to Consensus (CNT*
149   *1-5), Signal-to-Noise Ratio (SNR 1- 5)].*

150   L 138. Section 2.2 is missing.

151   ●   *Removed*

152   Section 3 should be expanded to provide a better understanding of how the algorithm works by
153   focusing on its more detailed relevant phases. For example, section 3.1 could be the subject of a
154   section in its own right, giving step-by-step details of how the algorithm works. Presented like this,
155   it is difficult to understand clearly how it works. Note that there are no references in this key
156   paragraph. At the end of the paragraph, the test results are not clear enough to be properly
157   understood.

158   ●   *We will extend the discussion of how the algorithm is developed. We hope that a longer*
159   *discussion along with the mentioned packages used will make it easier for other*
160   *researchers to follow along and be able to reproduce it.*

161   L.147 Detail this analysis.

162   ●   *Clarified – The analysis description and definitions are expounded*

163   L.148 Why use radial speeds when wind strength and direction are already taken into account?
164   Why use the signal-to-noise ratio, what does it provide?

165   ●   *Clarified – We used the radial speed as well as the speed and direction to supply higher*
166   *dimensional analysis to the algorithm. When studying the covariance matrix, it appears to*
167   *be relevant to the detection but does not create collinearity. Previous researchers in our*
168   *group have used the signal-to-noise ratio from radar wind profilers to demonstrate*

169   *boundary layer height detection and thus we supply this as well. However, the signal-to-*
170   *noise ratio of each beam does create collinearity as it is dependent on the thermodynamic*
171   *profile of the atmosphere which does not change much with direction in the field of view of*
172   *the instrument. Thus, we take only the average signal-to-noise ratio of all five beams.*

173   Figure 3, some elements are illegible and some acronyms are not defined.

174   ● *Clarified – The figure is adjust for height resolution and structural clarity*

175   Section 3.2 already contains results and could be introduced in section 4. In addition, this section
176   only focuses on 2 cases, which does not seem sufficient to properly qualify the algorithm's
177   performance. A more detailed study would enable us to test it more thoroughly by comparing the
178   NLLJs identified by the algorithm with all the NLLJs identified by the manual method in a year
179   other than the one used for training (the test set already identified, for example). This would make
180   it possible to better characterise the algorithm's shortcomings, by quantifying the number of false
181   events not taken into account, the % of missing data on average per event, etc. Without this kind
182   of statistic, no conclusions can be drawn.

183   • *Clarified – These is no absolute truth value for where an NLLJ is present in the wind*
184   *profiles. This is why we rely on the algorithm's performance testing to be qualitative by*
185   *previously reported and depicted by Delgado et al. (2013), Weldegauber (2009), and*
186   *Sullivan et al. (2017) that were captured by the same instrument and station (i.e., BELT*
187   *RWP). We also include a figure of the confusion matrix of the algorithm training and testing*
188   *instead of simply listing its testing results like previous.*

189   Simple post-processing could be used directly to eliminate outliers if no neighbours are present
190   and to include all the data between the ground and the jet.

191   • *Clarified – This was mentioned in the manuscript as an option for actual use of the*
192   *algorithm. However, we decided to leave these errors in so we may discuss its shortcoming*
193   *and show them visually in isolation and how this will propagate into the later analysis if*
194   *not removed.*

195   Figure 4: Each sub-panel should be indicated by a letter. Perhaps this figure should be split into
196   two separate figures focusing on each event. As with figure 2, the events should be centred to better
197   see their development.

198   • *We appreciate this comment to improve the clarity of our discussion. All of our panel*
199   *figures have been lettered in the figure itself and where mentioned in the text.*

200   L.218-220 Without seeing the winter months, such a conclusion cannot be drawn.

201   • *Clarified – We only seek to search for the warm-season Mid-Atlantic NLLJ as we are*
202   *primarily interested on its implications towards air quality. As we develop this approach*
203   *further, we plan to expand into the full year, but many of those low-level wind maxima are*
204   *of a different classification.*

205   Please show the missing months in Figure 5. In addition, in Figure 5, it can be seen that May
206   contains the most events and not June, July or August.

207   • *Clarified – See above response*

208   L.222-224 include references.

209   • *Corrected*

210   L.226 The year 2017 contains more events than 2019.

211   • *Corrected*

212   L.238-239 Give examples of synergy and cite references.

213   • *Clarified – The synergy we seek is from all available observational datasets in our study*
214   *region (e.g., Aerosol and Ozone Lidars, Sondes, AERONET, Pandora) along with*
215   *Reanalysis Datasets.*

216   L 246 Core time not defined

217   • *Clarified – The analysis description and definitions are expounded*

218   L.247 Replace all « m/s » with m.s-1

219   • *Corrected*

220   L.257-259 What percentage of events does this represent? Using simple post-processing, why not
221   exclude this erroneous data?

222   • *Clarified – This was mentioned in the manuscript as an option for actual use of the*
223   *algorithm. However, we decided to leave these errors in so we may discuss its shortcomings*
224   *and show them visually in the isolation and how this will propagate into the later analysis*
225   *if not removed.*

226   Section 4.3 As NLLJs are nocturnal events and sunset times vary according to the season, the data
227   should be standardised according to these times. Otherwise, the morphology of NLLJs could not
228   be correctly presented.

229 ● *Corrected – Thank you for this insight. We have adjusted the plotting to be centered around*
230   *scaled around sunset and sunrise.*

231 L.276 EDT is local time? Mention this earlier in the article and add sunset and sunrise hours on all
232 figures.

233 ● *Corrected – EDT stands for Eastern Daylight Time and will be mentioned before the*
234   *Figures.*

235 Section 5: Some conclusions may need to be modified in the light of the above changes.

236 ● *Noted – The conclusions are update with discussion from the adjustments to the figures*

237 L.333 I-95 corridor not defined

238 ● *Clarified – I-95 refers to the U.S. Interstate 95 highway which spans the Eastern Coast of*
239   *the United States. The I-95 corridor refers to the portion of this highway that spans the*
240   *most populated portion of the United States – Washington, DC to New York City, New York.*

241 L.346 delete « also »

242 ● *Corrected*

243

---

244 **RC2 (Referee #1)**

245 The study described in the manuscript is the development of an ML-based algorithm for the
246 identification of nocturnal low level jet events in Beltsville, MD from radar wind profiler data. The
247 wind profiler data is used to identify the events and to characterize the wind characteristics and
248 seasonality of events at this location. There are some major issues with the manuscript that I
249 explain below related to the articulation of the need for the ML-based algorithm, the methods used
250 for evaluation of the algorithm and finally with the claims made in the summary that are not based
251 on findings of the study. Based on these issues I recommend publication of this manuscript only
252 after the major revisions detailed below.

253

254 Major Issues:

255

256 1) Motivation of the need for a new NLLJ identification algorithm - The manuscript quite clearly
257 cites the previous studies that have examined NLLJs in the Mid-Atlantic and explains how the

258   events are identified. These events are used as part of the evaluation of the ML-based algorithm
259   developed in the study. If there is a robust enough method to identify these events, robust enough
260   to be used to evaluate the ML model(s), why is there a need for an ML-based identification method
261   at all? Please articulate this, that is, what the benefit of an ML based algorithm is and why the
262   present method is inadequate.

263   • *Thank you for drawing attention to our oversight in the manuscript. We have addressed a*
264   *similar concern from Referee #2 (see lines 22 - 25 of this document) and have elaborated*
265   *on this in the revision, largely in the introduction in the introduction and section 3.*

266   2) The method for the identification of the events for training and then for evaluation is not well
267   explained in the manuscript as written. Based on what is written, it appears that a year's worth of
268   RWP data and a pre-defined set of NLLJ events were used as training, and then the evaluation was
269   done based on a subjective "by eye" examination of a selection of events found in the literature.
270   The algorithm is then put to use for the 2017-2021 period and the events' wind speed characterized.
271   This is not a robust training and evaluation method and should be improved before the study is
272   published.

273   • *This is an important concern that you and Referee #2 both shared (see lines 185 – 190 of*
274   *this document). To clarify: the training dataset is used for both training and testing,*
275   *however, our oversight was to not include the results of the testing against the gradient*
276   *method for concerns that it may be out of the scope of the manuscript and journal. The*
277   *testing "by-eye" with previously reported and depicted NLLJs is a performance test with*
278   *the trained algorithm using qualitative inspection by experts on the Mid-Atlantic NLLJ. To*
279   *which there is no "absolute truth" validation set.*

280   3) The manuscript's introduction contains a description of LLJ events and their characteristics from
281   the literature that include intertial oscillations, temperature profiles and wind characteristics, as
282   well as the influence of these events on the local atmosphere. Both the introduction and the
283   summary refer to the study as characterizing NLLJ events and helping to understand them, but the
284   characterization here is limited to wind characteristics. I recommend the use of some auxilliary
285   dataset (perhaps a reanalysis) to characterize the events properly once they are identified by the
286   algorithm.

287   • *Our contribution to the study of the Mid-Atlantic NLLJ is in two-phases: First – to establish*
288   *the groundwork for automated NLLJ isolation and preliminary statistics and morphology;*
289   *Second – To establish a long-term (~20 year) climatology, formation, and impact analysis*
290   *of NLLJ events in the Mid-Atlantic that combines all available datasets in the study area*

291 *(see line 215 – 217 of this document). This endeavor is too large to conduct "by-eye" and*
292 *thus we sought to develop and automated system that will classify wind regimes, and the*
293 *exploration of machine learning for this task of NLLJ isolation is described in the*
294 *manuscript.*

295

296 Line by line:

297 Line 46 - "It is believed that the mid-Atlantic NLLJ is akin to the SGP NLLJ..."
298 Need a reference here or say that you will show this here.

299 ● *Corrected*

300 Line 96 - "...clear disagreement..." - what clear disagreement is being referred to? between what?

301 ● *Clarified – Disagreement in the Wind Speeds (Observations vs Reanalysis).*

302 Line 98 - NARR is not an operational model - its an analysis (reanalysis).

303 ● *Removed – We now use ERA5 Reanalysis*

304 {Fig 1 - what is the shading? 900 mb wind speed from NARR? Also - manuscript says for the case
305 of may 20, 2024 and the figure says "composite" - what is plotted?}

306 ● *Removed and Clarified – We now use ERA5 Reanalysis with more described figures*

307 Line 114 - This section talks about Figure 5 (before any manuscript mention of figs 2-4) - please
308 reorder the figures to be consistent with the order they are referred to in the manuscript

309 ● *Corrected – We have converted Section 2 in a discussion of the Observations and*
310 *Measurement Site. Figure 5 will now be Figure 2 located in Section 2 to supplement the*
311 *discussion of the wind profile observations dataset, and just before the discussion of the*
312 *training dataset. Section 3 then is solely about the NLLJ isolation Algorithm (e.g. training,*
313 *development, etc.)*

314 Lines 111-121 - should move to inside section 2.1.

315 ● *Corrected – We have converted Section 2 in a discussion of the Observations and*
316 *Measurement Site. Section 3 then is solely about the NLLJ isolation Algorithm*

317 Lines 127-129 - There is not enough explanation here about the "inflection points", what they are
318 and why they are important. I assume this detail is included in Zhang et al 2006, but some more of
319 the detail is needed here.

320   ● *Clarified - The training dataset for this attempt was hand-selected from NLLJ events during*
321   *2021, while the validation dataset was selected from previously reported and depicted by*
322   *Delgado et al. (2013), Weldegauber (2009), and Sullivan et al. (2017) that were captured*
323   *by the same instrument and station (i.e. BELT RWP). To gather a suitable dataset for*
324   *machine learning we have compiled scenarios expected in operation (i.e. incomplete daily*
325   *files, missing data, large scale weather systems, etc.). A manual and rudimentary isolation*
326   *method was applied using gradient detection solely on the southerly winds (180 – 270*
327   *degrees from North) with maximums greater than 5 m s-1 in both time and altitude to*
328   *capture the evolution and vertical extent of the NLLJ. This approach is demonstrated in*
329   *figure 3, where (A) depicts the final isolated NLLJ events from the speed and direction*
330   *profile (C and D), and (B) represents the visual representation of the gradient detection in*
331   *the temporal evolution. This method takes the wind speed evolution averaged from 0 - 2000*
332   *m and then interpolated and smoothed. The resulting time-series is then used to find the*
333   *first positive gradient and the last negative gradient, which are taken as the start and end*
334   *of the NLLJ event. This process is then repeated for the vertical extent using each profile*
335   *to find the top and bottom at each time step. We found that the manual tuning needed for*
336   *thresholds on time constrain, continuity, and direction evolution were important for*
337   *isolating NLLJs, but require attention in many different cases and thus we used the well*
338   *isolated cases from this method as a training set for the supervised machine learning*
339   *ensemble. The training set is comprised of 50 NLLJ events that were sufficiently isolated*
340   *and 50 events that contained no low-level wind maxima that contain low-level wind*
341   *maxima that we do not consider as LLJ relevant to this study for reasons of direction, or*
342   *evolution.*

343 Line 154 - "...visually conceptualized in Figure 3." How/why is the data pre-processing included
344 in the algorithm execution loop? is this done more than once?

345   ● *Clarified - The pre-processing step is merely as set of transforms that format the data for*
346   *analysis by the model. It is always necessary so that model has the data in the expected*
347   *format. Since the model works on single measurement points it is independent of both*
348   *vertical and temporal resolution but needs to be a properly formatted matrix for analysis.*

349 Line 176 - Please explain what an f1 macro test is, what the scores mean, and how this was
350 evaluated. Alternatively, remove this statement.

351   ● *Clarified - We will extend the discussion of how the algorithm is developed. We hope that*
352   *a longer discussion along with the mentioned packages used will make it easier for other*

353   *researchers to follow along and be able to reproduce it. We elect to keep this statement in*
354   *the manuscript so that our methods may be reproduced by other researchers. It is important*
355   *to note the algorithm training scoring method. F1-macro is a common supervised machine*
356   *learning testing framework that we will explain in the revised text with the included*
357   *discussion of the training and preliminary testing results.*

358 Line 186 - "...more than satisfactory..." is not quantifiable, particularly when the algorithm testing
359 is by visual inspection (of 50 cases used for training or for all the cases identified in fig 5?).

360   ● *Clarified (see lines 277 – 290 of this document)*

361 Line 211 - Why does the present study not include the "ongoing model refinement"?

362   ● *Clarified – We feel that the exploration of supervised machine learning for this task is*
363   *complete, and we seek to continue its use to further our overall goal (see lines 289 – 296*
364   *of this document). We feel that only this portion of the overall goal is sufficient to pursue*
365   *publication.*

366 line 315 - The connection to synoptic situation not established in study - connection to season, yes.

367   ● *Removed – We explore synoptic situation in our project.*

368 line 317 - "..understanding the atmospheric at play..." was not part of the study, and a connection
369 to predictive capability was not established.

370   ● *Clarified – This is a pursuit of our next project.*

371 322 - "critical characteristics"... also not established.

372   ● *Clarified – We have included further discussion of our motives and nomenclature in the*
373   *introduction section. "Critical Characteristics" - maximum wind speed, height of*
374   *maximum, duration, wind direction, etc.*