Paper Review: "Multi-season evaluation of temperature and wind in the marine boundary layer along the United States northeast coast in the High-Resolution Rapid Refresh model."

This paper examines the National Oceanic and Atmospheric Administration's (NOAA) High-Resolution Rapid Refresh (HRRR) model in the context of its performance within the marine atmospheric boundary layer (MABL) over the coastal and offshore waters south of New England. The investigators leverage data provided by the deployment of a variety of land-based and offshore instruments during the 3rd Wind Forecast Improvement Project (WFIP3) to quantify the HRRR's forecast performance over a period of one year. The paper provides a thorough examination of HRRR's forecasts on diurnal and monthly timescales, focusing on how well the model predicts the wind and thermodynamic profiles at several sites within the WFIP3 domain. The paper provides critical information regarding the performance of HRRR over offshore waters, historically a region suffering from a dearth of observational data. However, the narrative suffers from some lapses in analyses, including an incomplete literature review, use of adjectives (e.g., "excellent") without quantitative support, arbitrary use of certain threshold criteria (such as what constitutes a low-level jet [LLJ]), and lack of more descriptive figures confirming hypotheses put forth in the manuscript (e.g., field plots showing SST and/or wind speed distribution at critical levels). More specific comments follow below.

Page 2, lines 44-45: "Using 9 years of buoy-based lidar data in the New York Bight, Colle and Novak (2010)…." This is untrue. No such data was available or was used by Colle and Novak (2010). They used 10 years (April 1997 - August 2006) of data from a C-MAN station (Ambrose Light, or ALSN6), with wind observations obtained from an anemometer height of 29 m. (Note Ambrose Light was removed and in 2008 after a ship hit it, replaced by buoy 44065.)

Page 2-3, lines 47 - 59: There is no reference to recent papers by McCabe and Freedman (2023, 2025a) which describe detailed observational analysis and modeling sensitivity studies focused on the Low Level Jet (LLJ) characteristic of the New York Bight, coastal and offshore waters that overlap with the WFIP3 study region. This is quite surprising given the topic and geographic location of this paper. Note also that the New York State Energy Research and Development Authority (NYSERDA) deployed two buoys over a three-year period (2019 - 2022; see https://www.nyserda.ny.gov/All-Programs/Offshore-Wind/Focus-Areas/Regional-Collaboration/Siting-Offshore-Wind) in the coastal waters south of Long Island and east of New Jersey.

Page 3, lines 84-85: A deeper mixed layer over land (here, Block Island) is not necessarily an "internal boundary layer."

Page 5, section 2.1.1: No information provided regarding the manufacturer or specifications of the radar wind profilers referenced here and used in this study.

Page 5, section 2.1.2: Same comment as above.

Page 6, section 2.1.5: Same comment as above.

Page 7, Figure 2: This figure is very difficult to read, especially as the legends in (a) and (b) overlap with the data availability (a) and uncertainty profiles (b).

Page 8, line 162: The use of the term "excellent accuracy", lacks quantitative context. (The authors in Letizia et al. (2025[b]) appear to use the term "satisfactory.")

Page 8, line 168: Why the LWP threshold of > 8 g m^{-2}?

Page 8, lines 189-191: "…the number of operational radiosondes at Upton, New York, was not sufficient…." This is confusing. Given the high-resolution radiosonde data (~ 5 m in the vertical) available over the last several years, at ~365 days x twice daily launchings x ~10 years, that works out to approximately 7300 soundings. HRRR profiles at such coarse vertical resolution (10 levels below 1500 m) would seem to be a much inferior data set. Perhaps I am missing something here?

Page 9, equation 1: Based upon Archer (2025), isn't this still an approximation (there, should use the ≈ symbol)?

Page 9, line 225: Given the HRRR's limited number of levels below 1500 m, doesn't this make an accurate determination of the LLJ nose and structure problematic?

Page 10, lines 230 - 235: See McCabe and Freedman (2025a) regarding OSTIA.

Page 10, section 2.4: See McCabe and Freedman (2023, 2025a).

Page 11, lines 244 - 245 and Table 1: Why not use wind speed shear values to constrain LLJ identification? The above-nose wind speed decrease constraint seems rather arbitrary (although any filter would have some subjective characteristics). A more thorough explanation of the classes regarding LLJ identification would seem prudent.

Page 11, lines 257 - 259: Using two different layers for static stability (100 m - 300 m) and wind shear (50 m - 200 m) is inconsistent. The explanation is also arbitrary, and more hand-waving than science. And wouldn't this make comparison with the HRRR problematic (e.g., Bulk Richardson analysis in section 4.3)?

Page 13, Figure 4: It would be interesting to see the wind shear *exponent* (common in wind energy applications) as a variable in the analysis, although I understand this would be redundant.

Page 16, lines 316 -318: "Another reason [for the overestimation of 50 m wind speeds] could be the vicinity of Block Island to Long Island…." This assertion could be tested by examining HRRR model field plots (x-y) to see if any channeling is in fact present. Although eastern Long

Island and Block Island are quite flat (maximum elevations are ~ 50 m), frictional effects could possibly induce some "channeling."

Page 21, lines 366 - 367: Same comment as above.

Page 21, lines 363- 371: SST overestimation may be a result of cold water upwelling given the wind trajectory (southwesterly). See McCabe and Freedman (2025b). Model grids that extend simultaneously over land and water can be problematic.

Page 22, lines 389 - 390: Wouldn't a secondary peak in LLJ frequency (late summer) be inconsistent with the LLJ mechanism being driven by SST-land temperature differences (given SSTs peak in late August early September in this region)? Southwesterly LLJs can also be associated with pre-frontal systems in the northeast US.

Page 27, line 455: Use of "Unprecedented" here seems overly hyped, given previous deployments by NYSERDA and offshore wind developers in the region.

Page 27, lines 464-465: "The HRRR model did an excellent job in capturing these characteristics on average." What does this mean?

Page 28, line 488: "…overall great performance…." Hyperbole?

References

Colle, B. A. and Novak, D. R., 2010: The New York Bight jet: climatology and dynamical evolution, *Mon. Wea. Rev*., **138**, 2385–2404, https://doi.org/10.1175/2009MWR3231.1.

Letizia, S., Turner, D. D., Abraham, A., Rochette, L., and Moriarty, P. J.: Temperature profiling at the American WAKE ExperimeNt (AWAKEN), 2025: methodology and uncertainty quantification, *Wind Energy Science Discussions*, 1–36, https://doi.org/10.5194/wes-2025-198, 2025b.

McCabe, E., and **J. M. Freedman**, 2023: Development of an Objective Methodology for Identifying the Sea Breeze Circulation and Associated Low-Level Jet in the New York Bight. *Wea. Forecasting, 38, 571–589, https://doi.org/10.1175/WAF-D-22-0119.1*.

McCabe, E. J., and **J. M. Freedman**, 2025a: Quantifying the Uncertainty in the Weather Research and Forecasting Model under Sea Breeze and Low-Level Jet Conditions in the New York Bight: Importance to Offshore Wind Energy. *Wea. Forecasting*, 40, 425–450, https://doi.org/10.1175/WAF-D-24-0086.1.

McCabe, E., and **J. M. Freedman**, 2025b: Assessing the Impact of Cold-Water Coastal Upwelling along the New Jersey Coastline: Amplification of the Sea Breeze and Low-Level Jet. *J. of Atmos. and Oceanic Tech.*; DOI 10.1175/JAMC-D-25-0069.1.