Author's response

Dear Referee #2,

Thank you for your very positive review of our manuscript. Replies to your specific comments are presented below.

Summary:

This paper investigates the development of machine learning (ML) models to predict coastal wind speed profiles and LLJ occurrence from single-level meteorological variables Data from three locations of high relevance to offshore wind energy deployment (the U.S. Northeastern Atlantic Coastal Zone, the North Sea, and the Baltic Sea) were used. The ML models are trained on multiple years of lidar profiles and utilize single-level ERA5 variables as input. The models provide output spatial predictions of coastal wind speed profiles and LLJ occurrence.

General Comment:

The study is interesting and valuable for the offshore wind energy community. The article is well-written. The authors have used a variety of locations with different wind characteristics to apply the methods, which increases the applicability of the study.

Specific comments:

ERA5 data is quite coarse for wind applications in coastal areas. Could the authors comment/discuss if there is a potential advantage of using wind data of higher resolution for both LLJs and coastal wind speed profiles?

This is a fair point. In a paper by Hallgren et al. (2020) the performance of ERA5 was compared to other reanalyses/wind atlas of both higher and coarser resolution at four sites in the coastal waters of the Baltic Sea where lidar observations were available (Utö was one of the sites in the study). It was concluded that, in terms of the average wind profile, ERA5 (approximately 17 km $\times$ 31 km grid resolution in the Baltic Sea) demonstrated similar error metrics as the regional reanalyses UERRA (Uncertainties in Ensembles of Regional Reanalyses, 11 km $\times$ 11 km grid resolution) and as the New European Wind Atlas (NEWA, 3 km $\times$ 3 km grid resolution). However, the other global reanalysis that was investigated, MERRA2 (Second Modern-Era Retrospective analysis for Research and Applications) with coarser resolution (40 km $\times$ 55 km) did not achieve comparable results, heavily underestimating the average wind speed in the profile at all sites.

In terms of LLJs, the same study concluded that UERRA was the model that best captured the frequency of LLJ occurrence and its seasonal variation. All reanalyses struggled resolving LLJs at the correct point in time, with hit rates of 12–18% for ERA5, 28–41% for UERRA, and 15–25% for NEWA for the four sites. False alarm rates were on the other hand relatively low, indicating a general underestimation of LLJs; 1–3% for ERA5, 4–7% for UERRA, and 1–4% for NEWA. Recalling from Eq. 5 in the manuscript, the hit rate and the false alarm rate combines to give the SEDI.

In similar work by Kalverla et al. (2020), ERA5 was compared to both to NEWA and to the Dutch Offshore Wind Atlas (DOWA) with 2.5 km horizontal resolution. The study focused on lidar measurements performed at MMIJ in the North Sea. The authors conclude that DOWA describes the average wind profile best in general, however, with ERA5 outperforming the other models in strongly stable conditions. The relative occurrence of LLJs is too low in all models compared to observations and the hit rate of perfectly timing the occurrence of LLJs was 28% in ERA5, 52% in DOWA, and 33% in NEWA. False alarm rates were low for all models; 0.2% for ERA5, 1% for DOWA, and 0.6% for NEWA. Note that a slightly different definition of the LLJ was used by Kalverla et al. (2020) as compared to Hallgren et al. (2020).

In a recent preprint, Sheridan et al. (2023) analysed model performance comparing model output with lidar measurements performed at two sites off the California coast and focusing on LLJs. ERA5 was compared to two regional models with different planetary boundary layer (PBL) schemes, the National Renewable Energy Laboratory (NREL) data set for the Outer Continental Shelf off the coast of California (CA20-Ext) and the

2023 National Offshore Wind data set (NOW-23), both with 2 km horizontal resolution. Using a high threshold for LLJ identification, ERA5 failed to accurately resolve any LLJs at the correct point in time, and also false alarms rates were very low, 0.05–0.08%. CA20-Ext had a hit rate (false alarm rate) of 47–52% (2–4%) and for NOW-23 the scores were 13–23% (9–10%). Unfortunately for our work, there is not yet any similar work for comparison off the US Atlantic coast.

Concluding from these three studies, it seems that – in terms of the average wind conditions – ERA5 performs reasonably well, also when comparing with higher resolution models. However, in terms of resolving the LLJs, ERA5 struggles in getting the relative frequency of occurrence correct, and as a consequence of this also with accurately predicting the presence of an LLJ in time.

In the revised version of the manuscript, one of the paragraphs in the Introduction now reads (lines 83–91):

Although LLJs have been observed and simulated frequently offshore at heights relevant to wind energy, numerical weather prediction (NWP) models exhibit difficulty in resolving LLJ characteristics with high accuracy, i.e., in terms of timing and morphology (jet core height and speed) of LLJs. Kalverla et al. (2020), Hallgren et al. (2020), and Sheridan et al. (2023) all showed that regional models, optimised for a specific region and with higher horizontal resolution than the global models, are better in resolving coastal LLJs in the North Sea, the Baltic Sea, and off the California coast, respectively. However, not only the horizontal resolution is crucial, but also how calculations in the boundary-layer are treated by the models, i.e., the PBL scheme. When it comes to the average wind conditions in the profile, it seems to be less of a difference between state-of-the-art models, as long as the horizontal and vertical resolutions are good enough (Hallgren et al., 2020), even if the model performance varies with atmospheric stability (Kalverla et al., 2020).

It would be great that the authors provide some numbers regarding the computational time/power used by the different methods.

Unfortunately, as one of the main authors of the manuscript (J.A. Aird), who was in charge of the coding and implementation of the neural network (NN), has moved on to other work outside of research, our possibilities to re-run the NN models are very limited. Here we only present results for the random forest (RF) for the different tasks, but run-time for the NN should be of comparable numbers. We hope you will have indulgence with this.

The time it takes to train the RF, i.e., finding the optimal set of predictors using the forward and backward selection methods, is presented in the Table below. How long it takes to reach the optimal set of predictor mainly depends on how many predictors are chosen by the algorithm, and since this varies between height levels the run-time is presented as a min-max interval for each site. The run-time also depends on the amount of training data, which mainly differs between the sites, but also the data availability differs between different height levels for each site. The task is perfectly parallelizable as the search for the best predictors on one height level is completely independent of the results on other height levels.

For the binary task of predicting LLJs, the training time also depends on the time to find the optimal cost matrix, on top of the time it takes to find the best set of predictors using the forward and backward selection methods. For the LLJ prediction task, the ML models only need to be run once. All run-times presented in the Table below are representative for when the RF is run on a standard laptop (MacBook from 2022, 8 GB RAM, 8 cores).

	ASIT	MMIJ	Utö
Wind speed, one height level	286 – 308 s	243 – 338 s	207 – 486 s
LLJ prediction	164 s	145 s	388 s

The time to calculate predictor importance is not included in the numbers presented in the Table.

In the revised manuscript, the following has been added (lines 352–354) for the WSPE task:

The run-time to find the optimal set of predictors varies between height levels and among the sites, but generally ranges from 4 to 8 minutes per height level on a standard laptop.

For the LLJ prediction, the following was added on lines 396–397:

Finding the optimal predictors for the LLJ classification tasks takes, depending on the site, between 2.5 and 6.5 minutes on a standard laptop.

Once again, thank you for your comments.

Sincerely,
C. Hallgren, J.A. Aird and co-authors

References

Christoffer Hallgren, Johan Arnqvist, Stefan Ivanell, Heiner Körnich, Ville Vakkari, and Erik Sahlée. Looking for an Offshore Low-Level Jet Champion among Recent Reanalyses: A Tight Race over the Baltic Sea. *Energies*, 13(14):3670, 2020. doi: 10.3390/en13143670.

Peter C Kalverla, Albert A M Holtslag, Reinder J Ronda, and Gert-Jan Steeneveld. Quality of wind characteristics in recent wind atlases over the North Sea. *Quarterly Journal of the Royal Meteorological Society*, 146(728):1498–1515, 2020. doi: 10.1002/qj.3748.

Lindsay M Sheridan, Raghavendra Krishnamurthy, William I Gustafson Jr, Ye Liu, Brian J Gaudet, Nicola Bodini, Rob K Newsom, and Mikhail Pekour. Offshore low-level jet observations and model representation using lidar buoy data off the California coast. *Wind Energy Science Discussions*, 2023:1–28, 2023. doi: 10.5194/wes-2023-152.