We thank the reviewers for their valuable feedback and comments. We provide responses to their comments and suggestions below in the Table in blue, point-by-point.

General comments RC1		
1	This manuscript describes combining wind-wave coupled modeling using the COAWST modeling system with an LES modeling approach, using multiple nested grids extending down to less than 10 m horizontal resolution. When available, model output is compared to that from the RUNE offshore dataset. The simulations are ambitious, and it is impressive that reasonably realistic results were generated. However, it seems the sensitivity of the model setup versus coarser scale setups are quite modest, and as the authors note, the largest sensitivity appears to be due to a shortcoming in how the LES reproduces vertical shear near the surface. Though understandable for the large range of scales, the jump from the finest-scale mesoscale domain to the coarsest scale LES (D03-D04) uses an 11:1 nesting ratio, which is quite extreme, and ensures that most of D04 has underdeveloped turbulent eddies. Furthermore, D03 and D04 seem to be the focus on much of the analysis. While the authors note the problem and show how the cell perturbation method can at least partially mitigate against it, the vertical / horizontal aspect ratio of the grid spacing of D04 also being about 10:1 can cause distortions of the LES results, as the authors note on Line 228.	RC1
	Thanks for this comment and we agree with the reviewer. It is a compromise we need to make when performing multiscale simulations in real-time with WRF. The reason we jump from a mesoscale resolution (1100 m) to (100 m) is mostly that we want to avoid the grey zone (1000-300 m) in which neither a PBL or LES are appropriate. Domain 03 will in principle be able to resolve some scales, perhaps a few of them within the 3D turbulence scales. We therefore assume that the 3D turbulent eddies would be able to develop by having a good fetch to the boundary domain (in D04) combined with the cell perturbation method.	Authors
2	Though I'm not sure of the additional computational expense that would be involved, I'm wondering why more analysis wasn't at least done with D05, which is an LES domain that can at least benefit from the eddies generated on D04 through its boundary conditions. Or alternatively, why an additional mesoscale domain wasn't used with resolution intermediate between D03 and D04.	RC1
	Thank you for your comment. The simulations for D05 and D06 were run for a much shorter duration (approximately 4 hours after spin-up) due to the high computational cost, The reason that data in the two domains were not analyzed together with measurements is that we do not have matching high-frequency observational measurements at the site. The purpose of these simulations is primarily to illustrate the limitations in producing resolved turbulence at a grid spacing of 100 m. Even at D05, with a horizontal grid spacing of 20 m, the simulated vertical velocity variances remained very small.	Authors
Specific comments		
1	Line: 129: '9 min/km' is recommended for radiation – but does this mean that all domains update radiation at different times?	RC1
	Thank you for this comment. The radiation update can be at different times for each domain, but it also increases the computational cost twice for this case, and the effect on the results is negligible, therefore, 9 min/km is used here for all of the domains. The corresponding text has been revised to clarify this point.	Authors
2	Line 146: 'The CPM method…introduces controlled perturbations.' Controlled perturbations of what?	RC1
	Thank you for pointing this out. The CPM method introduces controlled perturbations of the potential temperature. The corresponding text has been corrected .	Authors
3	Line 150: I think that SWAN is not a phase-resolved wave model, but for the finest resolution WRF-LES domains, the grid spacings are certainly on the order of the	RC1

	wavelengths of the water waves. Could some discussion of the implications of this be provided?	
	Indeed, SWAN is a phase-averaged wave model and does not resolve wave phases, even if the WRF-LES grid spacing is comparable to the wave wavelength. In the coupled system, waves affect the atmosphere only through phase-averaged quantities such as roughness length and surface stress. Therefore, the LES represents flow over a sea-state-dependent rough surface, not over phase-resolved moving waves. Wave-coherent effects are not captured, which is a limitation of this approach. This is added in the Discussion Section. Wave models are sensitive to the accuracy of the forcing wind, and it is the same with the SWAN model here. The higher-resolution WRF-LES improves the representation of near-surface winds, which can improve the accuracy of wave modeling. Through the two-way coupling, this improved wave state feeds back to the atmosphere and can further improve the wind field.	Authors
4	Line 167: I would be curious to know what were the computation time / expenses of these simulations.	RC1
	Thank you for your interest. For the simulations with 24 hours of integration plus 6 hours of spin-up, the uncoupled simulation required approximately 18 hours using 100 CPUs. For the coupled simulation, the computational time increased by about 10%, with an additional 16 CPUs used for the wave component. When the fifth and sixth domains were included, the total simulation time increased substantially. In this case, the uncoupled simulation required approximately 10 days to complete 6 hours of spin-up and 4 hours of simulation time. Additional CPUs could not be used because the largest domain contained only 100 grid points in each horizontal direction, and WRF does not allow the use of more than 10 CPUs for domains of this size. As this is too technical, we did not include it in the paper.	Authors
5	Line 192: textual issue	RC1
	Thank you, it is corrected	Authors
6	Line 205: Why would excessive vertical wind shear explain the change in wind direction? Shouldn't it affect v as well as u? Does the explanation involve an assumed Ekman-spiral type force balance?	RC1
	Indeed, excessive vertical wind shear affects both u and v, not just one component. In our case the change in wind direction results from a relatively larger change in the v-component compared to the u-component, which shifts the vector slightly toward the east (from ~232° to ~229°). The wind direction change can be understood in terms of the near-surface momentum balance and turbulent mixing rather than a classic Ekman spiral. In a neutral boundary layer, increased shear near the surface enhances turbulent mixing and modifies the wind vector in both components, which can rotate the wind direction. While an Ekman spiral is a geostrophic–frictional balance useful at larger scales, the LES here is primarily driven by shear and turbulence production, and the directional change arises from how the Reynolds stresses redistribute momentum between u and v near the surface.	Authors
7	Line 214: 'mean value' – can you clarify how the mean value was computed? Specifically, was a horizontal average over the LES domain performed, or was just a model time series at one location used? The horizontal average would be more consistent with the standard LES definitions of resolved / SGS, and should be provided if available, (though of course this method usually isn't available from point measurements, necessitating some Taylor's-hypothesis assumptions).	RC1
	Thank you for your comment. The mean values were obtained from a 10-minute time average of sampling rate of 10 s at the grid point closest to the measurement site. Spatial averaging was not applied because the LES domain is non-periodic and not horizontally homogeneous, unlike idealized LES setups. The explanation is added to the manuscript.	Authors
8	Line 241: 'closely matches' – I might say 'more closely matches' instead. The LES with CPM is certainly much closer to the measurements than the other models, but over the time	RC1

	scales where observations exist (down to 20 minutes) LES-CPM is still generally about one-third smaller. For timescales from 20 minutes down to about 10 minutes LES-CPM very well could be much closer to reality, but unfortunately there are no measurements there.	
	Thank you for this comment. The text has been edited as suggested. Indeed it is unfortunate, high-frequency wind data were not available for the study site.	Authors
Major comments RC2		
1	The authors concluded that their findings demonstrate the benefits of high-resolution coupled modeling for capturing offshore boundary layer dynamics and improving turbulence representation. However, the discussion suggests that the apparent improvement in wind may arise for the wrong reason (excessive wind shear induced by SGS). Moreover, no direct turbulence measurements were available to assess the accuracy of the simulated turbulence fields. The analysis is based on a single observation location, and improved representation of vertical wind profiles, particularly at low levels, cannot be assumed to generalize to other cases. In addition, the claim of improved turbulence representation in the coupled system remains speculative without observational validation. While the results indicate that coupling increases turbulence intensity and alters its vertical distribution, it is unclear whether these changes constitute a genuine improvement in physical realism.	RC2
	Thanks for the clarifying comment. We now discuss (in Section 3.3) that the improved agreement of mean wind speeds simulated in domain 04 (d04) for the vertical levels lower than 100 m is due to, although perhaps counterintuitive, a mismatch of the simulated vertical profile in d04. The simulated profile in d04 shows higher wind shear than both the simulations in d03 (a mesoscale domain) and the observations. These are actual observations based on both the measurements and simulations, and a similar behavior is observed for the coupled simulations. The question is what is causing the excessive vertical shear in d04. The principal suspect is the SGS model, which is known to cause excessive vertical shear close to the surface (Sullivan 1994 and Gadde 2021). We clarify that this is a speculative comment in Section 3.3, since as you noticed we only simulated with one SGS model and over a particular event. We also acknowledge that, without turbulence measurements, we cannot verify whether the changes in the turbulence fields represent improvement on the way we simulate the atmosphere. However, the increase in turbulence fluxes in the coupled simulations is consistent with the expected physical response to wave-modified- surface roughness, and this clarification has been added to the turbulence section and conclusions.	Authors
2	Lines 85-86: The authors justify the selection of the 5th of December based on the 'good wind-wave alignment' observed for nearly a full day. While this provides a clean case for investigating coupled dynamics, the physics of air-sea momentum exchange can differ significantly during periods of wind-wave misalignment, which are common during the rapid intensification and frontal passage of ETCs. Specifically, surface drag and the resulting LES wind profiles are known to be sensitive to the angle between wind stress and wave movement. Could the authors discuss how the model performance or the importance of the two-way coupling might change during the periods of misalignment seen before or after the 'gray-marked' window? A brief sensitivity analysis or at least a discussion of this limitation in Section 4 would help strengthen the generalizability of the findings.	RC2
	We agree that wind-wave misalignment can alter the air–sea momentum exchange, and the effect of coupling may differ under such conditions. We added a paragraph in the Discussion Section acknowledging this limitation and stating that misaligned conditions should be examined in future work.	Authors
3	The authors utilize a two-way coupled WRF-SWAN system but do not activate the ocean model (ROMS) within COAWST. In a shallow shelf sea like the North Sea, storm-induced surges and tidal currents can significantly modify wave steepness and, consequently, the surface roughness calculated by the WBLM. Furthermore, storm-driven SST cooling can influence the heat fluxes and stability of the marine atmospheric boundary layer. Could the authors clarify if the influence of surface currents and dynamic SST was assessed, or discuss the potential impact of neglecting this oceanic feedback on the resulting wind profiles?	RC2

	In this study, the only update in SST field is through OSTIA and, neither current driven- changes in wave steepness nor SSTdriven stability changes are represented. We acknowledge that these feedbacks influence air–sea exchange in shallow shelf regions. However, our previous analysis of this storm (Hamzeloo- et al., 2025) showed that the event behaved as a deep-water wave case as the wave height follows the Forristall distribution, which suggests a limited impact of surge induced- steepening or tidal currents. This clarification has been added to the Discussion. Incorporating ROMS to include current and SST feedbacks will be considered in future work, as this study is the first step toward fully coupled WRF–SWAN-ROMS multiscale simulations.	Authors
4	The domain configuration (Table 1) utilizes a nest ratio of 11 between D03 (1.1 km) and D04 (100 m). Standard practice typically limits nest ratios to 3:1 or 5:1 to maintain numerical stability and ensure a smooth energy cascade across the 'grey zone.' Could the authors justify the use of such a large jump? Specifically, have they assessed whether this ratio leads to numerical artifacts at the domain boundaries or a delay in the spin-up of resolved turbulence in the LES domains?	RC2
	We agree that large nesting ratios (larger than typically recommended 1:3) could result in instability and numerical artifacts. However, we have also observed numerical artifacts when using closer to recommended nesting ratios of 1:5 (Peña and Mirocha, 2024). We use the 1:11 ratio here for two reasons. First, it was a necessary compromise to reach LES permitting resolution while keeping the multiscale simulation computationally feasible-. Second, we avoid the grey zone (300–1000 m), where neither a PBL scheme nor LES is appropriate (Arthur 2025 and Juliano 2022). With the fetch in D04 for the directions of the event (Section 2.1) and the use of the cell perturbation method, turbulence can sufficiently spin-up before reaching RUNE's Location (at the first hour of simulation). We now discuss these aspects in Section 3.4 and Section 4.	Authors
5	The turbulence statistics in Figure 6 suggest that the D04 (100 m) domain may not fully function as a true Large-Eddy Simulation. The resolved vertical variance (<w'w'>) in Figure 6c is very small for D04, indicating that much of the momentum transport is handled by the SGS. This is consistent with the authors' observation that the agreement with lidar at 40 m could partly result from SGS-induced shear rather than fully resolved turbulence. Additionally, the peak of horizontal variance occurs at 70 m in D04, compared to 30 m in D06, suggesting that the 100 m grid may be too coarse to fully capture the near-surface turbulence structure. It would be helpful for the authors to clarify whether the D04 results should be considered LES or are effectively a high-resolution mesoscale simulation with significant SGS contribution.	RC2
	As the reviewer points out, a large portion of the momentum fluxes rely on the SGS model when using a 100-m coarse LES. This is why 100-m LES are known as "coarse LES" or "poor-man's LES"(not high-resolution mesoscale). But note that the choice of 100 m is also intentional: this resolution is widely used in industry and operational multiscale modeling because it is computationally affordable (Vortex2018 and Kshetri 2025). We have clarified in the Simulations Subsection.	Authors
6	The authors provide a detailed description of the SWAN configuration, including whitecapping and depthinduced breaking. However, there is no discussion of -nonbreaking -waveinduced mixing- or sea spray effects, both of which can influence wind speeds and sea states. In addition, initializing the open boundaries at zero may neglect preexisting swell energy that could influence the wave field and associated -waveinduced- stress. Could the authors clarify whether these processes are accounted for in the WBLM or WRF physics? If not, a discussion of how their omission might affect the simulated wind profiles and surface drag would help contextualize the results. At minimum, these limitations should be acknowledged in the manuscript.	RC2
	Thank you for the clarifying comments. Regarding the wave boundary conditions, we used a hot start for SWAN, so the wave field was already developed at the beginning of the simulation, and the outermost domain is large enough for wave growth to occur before	Authors

	reaching the RUNE area. In addition a storm situation is likely more reflected in wind-driven waves rather than swell which is another reason, why for this particular case we don't need wave spectra at the open boundaries. Concerning the physical processes, SWAN considers processes related to sea spray or nonbreaking -wave induced mixing, specifically by modeling white capping and wave-induced breaking, which are the primary sources of sea spray. While SWAN does not directly simulate individual spray droplets, it incorporates advanced wave boundary layer models and whitecapping dissipation parameters that are essential for accurate sea-spray, wind-input, and heat flux parameterizations, particularly in coupled atmosphere-wave models. We added text to explain these in the revised version, Section 2.3 (Simulations)	
7	Lines 169-173: The validation approach using D03 (finest mesoscale) and D04 (coarsest LES) against RUNE observations is reasonable, as these domains have sufficient runtime to provide meaningful comparisons. However, the analysis of turbulence fluxes in D05 and D06 is limited by the very short runtime of these fine LES domains. Early spin-up and underdeveloped turbulence, especially given the extreme nesting ratios, may affect the representativeness of the results. The authors should clarify this limitation and discuss how it might influence the conclusions drawn from D05 and D06.	RC2
	All domains share the same spinup- period, despite the shorter total runtime of D05 and D06.In the comparison of the fluxes from D06, D5, and D04 (section 3.4 of the revised manuscript) the results from all domains are taken from the 2 hours after a spin-up of 8 hrs from the same initialization time. The purpose of D05 and D06 in this study is to illustrate how horizontal resolution affects the ratio of resolved to SGS turbulence. We now have clarified this point in the Simulations and turbulent fluxes Sections (2.3 and 3.4)	Authors
8	The authors discuss ocean wave–induced surface roughness and its associated processes multiple times, but they do not show how surface roughness differs between the coupled and uncoupled simulations. It would be helpful to include the distribution of surface roughness length for D04 and D03, either in the main text or in the supplementary material.	RC2
	Following the reviewer comment, a Figure is now added (new Figure 3) in the Results to show the surface roughness length distribution from D04 and D03 for both WRF only and coupled simulations.	Authors
9	The authors note in Figure 6 that turbulence levels (velocity variances and covariances) increase when wind–wave interactions are included. While previous work is cited to suggest this behavior is expected, the manuscript does not clearly explain the physical mechanism behind this enhancement for the specific North Sea case. Could the authors clarify whether the increased turbulence is primarily driven by higher effective surface roughness from the WBLM, or whether it results from dynamic feedback between the resolved wave-induced stress and SGS? Providing this mechanistic context would strengthen the paper by moving beyond observation toward a more robust physical interpretation.	RC2
	Thank you for the comment. Two different mechanisms contribute to the turbulence differences shown in Fig. 6. The first is a resolutionrelated effect associated with the transition from the mesoscale domain (D03) to the LES domain (D04): switching from a PBL scheme to the Deardorff SGS model changes the partition between resolved and SGS turbulence, and the 100 m grid spacing limits how much of the turbulent spectrum can be explicitly resolved. This explains the differences between D03 and D04, independent of coupling. The second mechanism is the effect of wind–wave coupling itself, seen when comparing the coupled and uncoupled simulations within the same domain. Here, the WBLM modifies the surface roughness and -nearsurface stress, enhancing shear production and increasing resolved turbulence.- The finer domains (D05 and D06) further illustrate how increasing horizontal resolution shifts a larger fraction of turbulence from the SGS to the resolved scales. We have added clarifications on these aspects in the Discussion Section 4.	Authors
Minor specific points RC2		
1	I suggest swapping Figures 1 and 2 and referring to the map in Section 2.1. The manuscript currently discusses the observed wind and wave characteristics before clearly identifying the	RC2

	observation location. Since these characteristics are highly location-dependent, presenting the map first would improve clarity for the reader.	
	We agree that introducing the observation location earlier improves the clarity of Section 2.1. However, Figure 2 contains both the RUNE location and the simulation domain configuration, which belongs naturally in the modeling subsection. We do that to limit the number of figures within the manuscript. To address your point without disrupting the structure, we now exchange Figure 2 and Figure 1 locations in the Section 2.1, so the reader can immediately identify the measurement location before the observed wind and wave characteristics are discussed.	Authors
2	Line 123: How frequently is the SST (OSTIA) is updated?	RC2
	OSTIA is updated every six hours, which is also added to the manuscript (Simulation Section 2.3).	Authors
3	Line 193: Please revisit this to ensure the description is a complete sentence	RC2
	The sentences are revised as suggested.	Authors
4	Figure 4: It took some time to fully understand this figure. The authors may consider revising the figure caption and possibly renumbering the panels to make it more intuitive and easier for readers to interpret.	RC2
	The caption is revised as suggested.	Authors
5	Figure 5: Please correct the typo in the time label from 7:00 UTC -> 17:00 UTC.	RC2
	The typo is now corrected.	Authors
6	figures: To avoid confusion, please consider adding the domain number next to each experiment in the figure legends	RC2
	The legends are now also revised.	Authors