

Author Response to Anonymous Referee RC1 (egusphere-2026-3355-RC1)

We thank the referee for the careful reading of our manuscript and for the constructive and detailed comments. We have addressed all comments below. Changes to the manuscript are marked using track changes in the revised version

Major Comment 1 - Reinterpretation of the WRF Resolution and Two-Way Nesting Feedback

Referee. In Sect. 3.5 and 4.2, the authors note a counterintuitive result where WRF-d01-2way (12 km) achieves lower bias (ME) and RMSE than WRF-d02-2way (4 km), attributing this primarily to grey-zone boundary layer limitations and lateral boundary constraints. However, a more plausible numerical explanation stems from the two-way nesting feedback mechanism itself. In WRF two-way nesting, prognostic fields (and certain surface/accumulated variables) calculated on the child domain (d02) are transferred back to the parent domain (d01) via spatial area-averaging (3x3 grid-cell filtering because of domain ratio).

This spatial aggregation naturally filters out high-frequency noise and local sub-grid variability. When verified against point station observations, the smoothed d01 field inherently benefits from reduced variance and a mitigated "double penalty" effect (where small spatial/temporal displacement errors in fine-scale circulations double the statistical error penalty). The fact that d02 achieves higher correlation coefficients (CC) in June despite higher ME strongly supports this spatial filtering hypothesis. The authors should revise the discussion in Sect. 4.2 to address the spatial area-averaging and variance-reduction effects of two-way feedback rather than relying on speculative grey-zone arguments.

We thank the reviewer for this important clarification. We agree that the spatial averaging inherent in the WRF two-way feedback mechanism provides a more complete explanation for the superior point-verification performance of WRF-d01-2way relative to WRF-d02-2way. In WRF two-way nesting, fine-grid prognostic fields are averaged over the nine child grid cells contained within each parent grid cell before being fed back to the parent domain (Skamarock et al., 2019). This averaging reduces spatial variance in the parent domain solution, which mitigates the penalty associated with small spatial displacement errors in fine-scale circulations that disproportionately inflate point-scale error metrics at higher resolution (Mass et al., 2002). The higher correlation coefficient of WRF-d02-2way in June supports this interpretation, as the finer domain captures temporal variability more accurately despite its higher ME. We have revised Sect. 4.2 to reflect this mechanism as the primary explanation, while retaining grey-zone boundary layer behavior and lateral boundary constraints as contributing factors that would require dedicated sensitivity experiments to disentangle.

The revised Sect. 4.2 reads as follows:

A counterintuitive finding of this study is that increasing horizontal resolution from 12 km to 4 km within the WRF framework does not consistently improve near-surface wind forecasts. WRF-d01-2way outperforms WRF-d02-2way in terms of ME across both months and all TRI classes, despite the latter resolving topographic gradients at three times the spatial detail.

Interpreting this result requires some care, since the two configurations differ not only in resolution but also in how the two-way feedback mechanism operates at each resolution level. In WRF two-way nesting, fine-grid prognostic fields are spatially averaged over the nine child grid cells contained within each parent grid cell before being fed back to the parent domain (Skamarock et al., 2019). This averaging filters high-frequency spatial variability from the child domain solution before it reaches the parent. When verified against point station observations, the smoothed parent domain field therefore presents lower spatial variance than the fine-grid solution itself. This variance reduction may mitigate the penalty associated with small spatial displacement errors in fine-scale circulations, which disproportionately inflate point-scale error metrics at higher resolution (Mass et al., 2002). The fact that WRF-d02-2way outperforms WRF-d01-2way in correlation coefficient in June (CC: 0.545 vs. 0.534) is consistent with this interpretation: the finer domain captures temporal variability more accurately, while its higher spatial detail introduces displacement errors that inflate ME and RMSE in point verification.

Grey-zone boundary layer behavior and lateral boundary constraints may also contribute to the observed pattern. At 4 km, WRF's boundary layer scheme operates near the transition between the fully parameterized and partially resolved turbulence regimes, where conventional PBL assumptions may no longer hold (Wyngaard, 2004; Shin and Hong, 2015). Furthermore, lateral boundary conditions from the 12 km parent constrain the mesoscale solution of the inner domain and may limit the independent development of fine-scale circulation features that higher resolution would otherwise produce (Warner et al., 1997; Harris and Durran, 2010). Disentangling the relative contributions of these factors would require dedicated sensitivity experiments beyond the scope of this study.

Major Comment 2 - Need for Process-Level and Vertical Profile Diagnostics

Referee. To fulfill the model development and evaluation objectives of GMD, the analysis should extend beyond single-level (10 m) scalar statistics. The authors identify a compensating bias structure in MPAS. To physically explain this behavior, representative case studies or composite analyses over high-TRI areas should include (e.g. Vertical profiles of horizontal wind speed, and turbulent kinetic energy (TKE) within the boundary layer / the implications of differing vertical coordinates on horizontal pressure gradient force (PGF) errors over steep topography ...)

We agree that process-level diagnostics would provide valuable insight into the physical mechanisms underlying the observed performance differences. However, the study domain lacks upper-air observational coverage sufficient to validate vertical profiles of wind speed, turbulent kinetic energy, or boundary layer depth over the high-TRI stations where the performance gap is most pronounced. Model-to-model comparison of vertical profiles without observational reference would not constitute meaningful validation in the context of this evaluation study. We have therefore addressed this limitation explicitly in the revised Sect. 4.4, noting that the absence of upper-air observations limits the ability to attribute surface-level differences to specific boundary layer processes, and identifying this as an important direction for future work. We have also strengthened the physical discussion in Sect. 4.1 by addressing the potential role of differing vertical coordinate formulations between MPAS and WRF over steep orography.

The following text has been added to Sect. 4.1:

...

This structural advantage appears most clearly at high-TRI stations where boundary layer and orography interactions are strongest.

Beyond mesh geometry, differences in the vertical coordinate formulation between the two models may also play a role. MPAS employs a height-based vertical coordinate with smoothed terrain-following surfaces, while WRF uses a hybrid sigma-pressure coordinate. Over steep orography, terrain-following coordinates can introduce errors in the horizontal pressure gradient force that affect the representation of near-surface flow (Klemp, 2011). While Shin et al. (2025) found that these numerical differences did not produce substantial changes in surface wind speed in their sensitivity experiments over the Korean Peninsula, the degree to which this finding generalizes to the more diverse topographic setting of Türkiye remains an open question.

The following text has been added to Sect. 4.4:

...

Disentangling these contributions would require dedicated numerical sensitivity experiments beyond the scope of this study.

A further limitation concerns the absence of upper-air observations over the study domain. The compensating bias structure identified in Sect. 3.6, and the terrain-dependent performance differences discussed in Sect. 4.1, likely reflect differences in how MPAS and WRF represent the planetary boundary layer structure over complex terrain. However, without vertical profile observations such as radiosonde measurements, it is not possible to directly evaluate the boundary layer depth, turbulent kinetic energy, or wind shear profiles simulated by the two models. This limits the ability to attribute the observed surface-level differences to specific physical processes within the boundary layer and represents an important direction for future work.

...

Major Comment 3 - Limitations of Nearest-Neighbor Point Extraction

***Referee.** The model-to-observation pairing extracts the value at the nearest grid center without interpolation. Because a 12 km grid cell represents a footprint of $\sim 144 \text{ km}^2$ while a 4 km grid cell represents $\sim 16 \text{ km}^2$, comparing single point measurements directly against grid-scale values introduces representative scale discrepancies. The manuscript should explicitly acknowledge this representativeness disparity or consider spatial verification perspectives (e.g., neighborhood verification or upscale spatial smoothing) to evaluate the true added value of high-resolution models.*

We acknowledge this representativeness limitation. In this study, modelled 10 m wind speed is extracted at the grid cell center nearest to each station location without horizontal interpolation. A single 12 km grid cell represents a footprint of approximately 144 km^2 , while a 4 km grid cell represents approximately 16 km^2 . Comparing point observations directly against grid-scale values introduces a representativeness mismatch that is inherently resolution-dependent and may systematically disadvantage higher-resolution configurations in point verification (Mass et al., 2002). We have addressed this explicitly in the revised Sect. 4.4. Neighborhood-based verification methods could partially address this issue, but are beyond the scope of the current study and are identified as a direction for future work.

The following text has been added to the end of Sect. 4.4:

...

A further methodological consideration concerns the representativeness of the model and observation pairings. In this study, modelled 10 m wind speed is extracted at the grid cell center nearest to each station location without horizontal interpolation. A single 12 km grid cell represents a footprint of approximately 144 km², while a 4 km grid cell represents approximately 16 km². Comparing point observations directly against grid-scale values introduces a representativeness mismatch that is inherently resolution-dependent and may systematically disadvantage higher-resolution configurations in point verification (Mass et al., 2002). Neighborhood-based verification methods could partially address this issue but are beyond the scope of the current study and are identified as a direction for future work.

Minor Comment 1

Referee. MPAS Lateral Boundary Conditions: Please provide a brief explanation of how lateral boundary relaxation zones are configured in MPAS-Limited-Area relative to the boundary buffer zones in WRF d01/d02 to ensure consistent comparison of boundary forcing effects.

A clarifying paragraph describing the MPAS-Limited-Area lateral boundary configuration has been added to Sect. 2.4. The WRF lateral boundary buffer distances were already documented in Sect. 2.3.

The following text has been added to Sect. 2.4:

...using the `create_region` tool from MPAS-Limited-Area (<https://github.com/MPAS-Dev/MPAS-Limited-Area>). In the MPAS-Limited-Area configuration, lateral boundary conditions are applied through a relaxation zone along the outer boundary of the clipped domain. The 4 km quasi-uniform inner core entirely encompasses the 509 TSMS evaluation stations, and is separated from the domain edge by approximately 1020 km zonally and 720 km meridionally, a buffer that encompasses both the 4–12 km transition zone and the quasi-uniform 12 km outer region.

Minor Comment 2

Referee. Lines 498-507: The authors state that numerical differences (advection order, vertical coordinate formulation) were not isolated through sensitivity runs based on findings from previous studies. While understandable, the text should clarify that topography-induced flow separation and wave breaking over Türkiye's terrain might respond differently to vertical coordinate formulation and horizontal diffusion than over the previous studies.

We agree that the generalization of Shin et al. (2025) findings to Türkiye's more diverse topographic setting warrants explicit qualification. The following sentence has been added to Sect. 4.4.

The following text has been added to Sect. 4.4:

...While Shin et al. (2025) found that these numerical differences did not produce substantial changes in surface wind speed, the degree to which this finding generalizes to Türkiye's more diverse topographic setting remains an open question. In particular, topography-induced flow separation and wave breaking over Türkiye's complex terrain may respond differently to vertical coordinate formulation and horizontal diffusion schemes than over the relatively smoother terrain of the Korean Peninsula, and this warrants dedicated investigation in future work.

Minor Comment 3

Referee. Figure 10: The zero-crossing analysis across observed wind speed bins provides valuable insight. It would be helpful to include sample sizes (or frequency distributions) for each wind speed bin to highlight how representative the high-wind bins are within each TRI class. separation and wave breaking over Türkiye's terrain might respond differently to vertical coordinate formulation and horizontal diffusion than over the previous studies.

We agree that sample sizes per wind speed bin would aid interpretation of the zero-crossing analysis, particularly for the high-wind bins where sample counts are lower. Sample sizes for each wind speed bin and TRI class will be added to Fig. 10 in the revised manuscript.

Minor Comment 4

Referee. The unit for wind speed is inconsistently formatted across the manuscript (e.g., "m s⁻¹", "m s-1"). Please unify all instances into standard SI format ($m s^{-1}$).

All instances of wind speed units have been standardized to $m s^{-1}$ throughout the manuscript, including in the text, table headers, figure captions, and axis labels.

The reference to be added:

Mass, C. F., Owens, D., Westrick, K., and Colle, B. A.: Does increasing horizontal resolution produce more skillful forecasts? The results of two years of real-time numerical weather prediction over the Pacific Northwest, B. Am. Meteorol. Soc., 83, 407–430, [https://doi.org/10.1175/1520-0477\(2002\)083<0407:DIHRPM>2.3.CO;2](https://doi.org/10.1175/1520-0477(2002)083<0407:DIHRPM>2.3.CO;2), 2002.